Spectral Analysis of Sound Errors in Persons With Apraxia of Speech and Aphasia

Julie L. Wambaugh
Patrick J. Doyle
Joan E. West
Michelene M. Kalinyak
Aphasia Rehabilitation Research Laboratory and Clinic
Highland Drive VA Medical Center, Pittsburgh

Incorrect productions of [ʃ] were compared to correct productions of [s] and [ʃ] in two subjects with apraxia of speech and aphasia. Frequencies of the peak spectral energy (FPSE) were extracted from power spectrums of monosyllabic words. In addition, in order to verify that [s] and [ʃ] could be differentiated in terms of FPSE, we obtained speech samples of correct productions of [s] and [ʃ] from two normal male speakers. A perceptual task was also completed by requiring two listeners to categorize recorded productions. Results revealed significant differences among sound categories for Subject 1, but not for Subject 2. That is, Subject 1's incorrect [ʃ] productions were acoustically different from his homonymous [s] productions, whereas the productions were not acoustically different for Subject 2. These findings suggest that Subject 1's /ʃ/ errors were not errors of substitution, but would be more appropriately characterized as distortions. Conversely, this spectral analysis did not clarify the nature of Subject 2's incorrect /ʃ/ productions; both substitution and distortion remained possible appropriate descriptors of his sound errors.

Numerous investigations have been conducted in which the speech of persons with apraxia of speech and/or aphasia has been examined through the use of acoustic analysis (for reviews of this literature, see McNeil & Kent, 1990, and Wambaugh, Doyle, Kalinyak, & West, in press-a). These acoustic analyses have allowed us to see phenomena unavailable through more traditional, perceptual means and, thus, have increased our knowledge of sound production in apraxic/aphasic speakers. Such knowledge may broaden our understanding of the nature of these disorders and assist us in developing appropriate treatments.

In most acoustic analyses of aphasic and/or apraxic speech, the disordered speech has been compared to normal speech (Harmes et al., 1984; Shinn & Blumstein, 1983; Ziegler & von Cramon, 1986). This approach allows a description of the deviant speech and can provide a determination of whether or not an individual's sound productions differ from normal, but is limited in determining the nature of the deviation. There has been little use of acoustic analyses to make comparisons of correct and incorrect sound productions by individual apraxic/aphasic speakers. Comparing sound productions acoustically within an individual's sound repertoire has been fruitful when used in determining if sound errors of misarticulating children have probable phonologic or phonetic origins (Weismer, 1984). Application of this avenue of research to apraxic/aphasic sound errors could provide insights into the nature of such errors. That is, a determination of how an apraxic/aphasic individual varies sounds within his or her sound repertoire, particularly in the form of subphonemic contrasts among sounds, could clarify the origins of some errors.

The purpose of this investigation was to compare incorrect productions of /ʃ/ (the error—often perceived as [s]) to correct productions of /s/ (the error sound), as produced by two speakers with apraxia of speech and aphasia. For example, we were interested in examining the [s] in *shoe* → *Sue* or *shoot* → *suit* in relation to correctly produced homonyms of those error sounds, such as the [s] in *Sue* → *Sue* or *suit* → *suit*. We hypothesized that if these sound errors were substitution errors, then substituted sounds would not appear acoustically different from their correctly produced homonyms (e.g., the [s] in *shoe* → *Sue* should be identical to the [s] in *Sue* → *Sue*). Conversely, if these sound errors did not stem from a sound selection disorder, it would appear likely that a phonologically intact system would attempt to preserve some differentiation of sound categories, particularly with potentially homonymous productions. In addition to comparing incorrect [ʃ] and correct [s] productions, we were interested in comparing post-treatment, correct [ʃ] productions to the pretreatment productions.

The study of obstruent spectra has been the focus of numerous investigations, so that we may predict the spectra for /s/ to have a major prominence near 6 kHz and /ʃ/ to have a peak near 2.5 kHz, when produced by male speakers (Heinz & Stevens, 1961; Hughes & Halle, 1956). The frequency of peak spectral energy (FPSE) is lower for /ʃ/ because of the place and degree of constriction and the occurrence of lip rounding (Fant, 1960; Stevens, 1971). Whereas the FPSE is thought to be an important perceptual cue for the sibilant fricatives (Soli, 1981), more recent investigations have indicated that additional aspects of the obstruent spectra may also have important roles in correct classification of fricatives (Forrest, Weismer, Milenkovic, & Dougall, 1988; Hoole, Ziegler, Hartmann, & Hardcastle, 1989). The analysis of obstruent spectra has rarely been applied to the speech of apraxic/aphasic persons (Harmes et al., 1984; Shinn & Blumstein, 1983; Ziegler & von Cramon, 1986) and has not been used to compare correct and incorrect productions by individual apraxic/aphasic speakers.

Spectral analysis, specifically the determination of FPSE, was selected for comparing the /ʃ/ and /s/ productions of our subjects. This method was chosen because of its probable importance as a perceptual cue and its potential for clinical application.

Methods

Subjects

The subjects were two Caucasian men with moderate to severe apraxia of speech and Broca's aphasia. Both were native English speakers, had passed audiometric screening tests, exhibited right hemiparesis, and displayed no significant dysarthria. Subject 1 was 63 years old, premorbidly right-handed, had completed 12 years of formal education, and was a retired contractor. Subject 2 was 52 years old, premorbidly left-handed, had completed over 18 years of formal education, and was a retired dentist. Subjects 1 and 2 were 76 months and 33 months postonset of left-hemisphere stroke, respectively. Assessment results for both subjects are presented in Table 1.

Each subject's speech was characterized by difficulty initiating speech, trial and error groping, numerous sound substitutions, occasional sound distortions, and variability in productions of multisyllabic words over repeated trials. Sound errors in monosyllabic words tended to be more consistent, in that certain sounds were usually produced incorrectly, and the errors tended to be similar across repetitions. Additionally, speech elicited through repetition was judged to be equal or superior to spontaneous speech. The subjects were very similar in the quality and quantity of their verbal productions. Subject 2, however, was much more aware and responsive to his sound errors than was Subject 1. That is, he usually spontaneously indicated when productions were incorrect, and he attempted self-corrections much more frequently than did Subject 1. Both subjects displayed language skills consistent with the diagnosis of Broca's aphasia: reduced phrase length; agrammatism, with predominant use of content words; diminished naming and repetition skills; and relatively spared auditory comprehension.

Speech Samples

The speech samples for Subject 1 were 70 monosyllabic words: 30 in which the word final /ʃ/ was produced incorrectly; 30 in which the word final /s/ was produced correctly; and 10 in which the word final /ʃ/ was produced correctly (see Appendix A). The speech samples for Subject 2 were 81 monosyllabic words: 27 in which the word initial /ʃ/ was produced incorrectly; 27 in which the word initial /s/ was produced correctly; and 27 in which the word initial /ʃ/ was produced correctly (see Appendix B). For Subject 2, the three sets of words were complete sets of minimal contrasts. That is, for each word in each list, there was a corresponding word, with only the initial consonant varied, in both of the other lists. Subject 1 had fewer correct productions of /ʃ/, so that only the correct /s/ and incorrect /ʃ/ word lists were completely matched.

Procedure

All recordings were made with each subject seated facing an experimenter in a quiet room. A Sony EMC 155 lapel microphone was used with a Sony TC-D5M cassette recorder. The speech samples were collected during subjects' participation in a treatment research project. For both subjects, the incorrect [ʃ] productions were obtained during baseline phases in which /ʃ/ words were elicited through repetition. The correct [s] words were collected in initial treatment sessions in which the subjects were required to produce these words, through repetition, to contrast with the /ʃ/ words that they had produced incorrectly. The correct [ʃ] words were obtained in probes, identical to those conducted in baseline, that occurred during the treatment phase.

Normal Speech Samples

Although the literature indicates that /ʃ/ and /s/ can be differentiated in terms of FPSEs (Edwards, 1992), we wanted to verify this observation using normal speakers and our apraxic/aphasic subjects' specific stimuli. For this reason, we obtained samples of correct productions of /ʃ/ and /s/ from two normal, male speakers, who were matched to the apraxic subjects by age, education, and regional dialect.

Normal Subject 1 produced the same 30 final /ʃ/ and 30 final /s/ words as Apraxic Subject 1, and Normal Subject 2 produced the same 27 initial /ʃ/ and initial /s/ words as Apraxic Subject 2. The normal subjects' productions were elicited in random order through repetition. Two project staff members independently verified that all of the normal subjects' productions were perceptually correct.

Acoustic Analysis

Recorded speech data were sampled at 20 kHz and power spectrums were made by fast Fourier transform (FFT) on the Computerized Speech Lab (CSL; 1994). Oscillographic displays on the CSL were used to extract the fricated segments to be transformed. The onset of frication was determined by consistent noise energy, and the offset of frication was determined by the periodicity of the following vowel transition. The center point of the fricative segment was located, and 80-ms time windows were selected by marking 40 ms from the left and right of the center point. From this 80-ms window, the average power spectrum was produced by the CSL (see Figure 1 for an example of a power spectrum). The CSL recorded the 256 coordinates for each curve, and, through computerized ranking of these values, we determined the frequency of the largest spectral peak for each segment. We also verified each of the FPSEs by marking the largest peak of each displayed power spectrum, using the cursor function of the CSL.

Perceptual Analysis

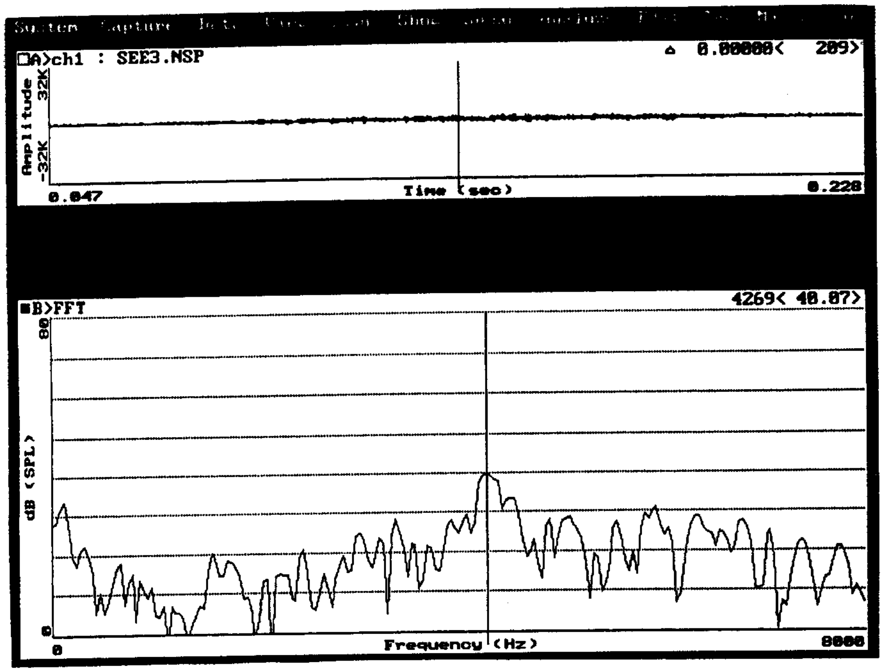
The speech data analyzed in the preceding acoustic analyses were used to prepare a listening tape with the total group of 151 words randomized with 5-second intervals between each word. Two speech-language pathologists, not involved with this research, served as listeners. For each stimulus, they were asked to choose between minimal /ʃ/–/s/ word pairs or to indicate that they could not make a determination, based on what they heard.

The listeners performed this task independently, in a quiet room. A Sony TC-D5M cassette recorder and Yamaha MS20S monitor speaker were used. The listeners were allowed to set the speaker to a comfortable listening level and could replay stimulus items as often as desired.

TABLE 1. Assessment results.

Measure	Subject 1	Subject 2
Western Aphasia Battery (AQ)	31.2	29.3
Western Aphasia Battery Classification of Aphasia	Broca's	Broca's
Test of Adolescent/Adult Word Finding (total raw score)	0	1
Revised Token Test (overall score)	11.74	8.92
Apraxia Battery for Adults	2 ratings "severe" 2 ratings "moderate" 1 rating "mild"	4 of 5 ratings "severe"
Mean Length of Utterance (in words) (from 15-min. conversational sample)	1.28	2.89

FIGURE 1. An example of an oscillographic display of a fricated segment and an average power spectrum produced by the CSL.



Reliability

Consensus agreement, through independent transcription by two investigators, was used to determine that the speech samples in each category were accurate representations of that category.

To evaluate the reliability of our analysis procedure, a second investigator extracted fricated segments and derived the peak frequencies for 30% of the speech samples, selected at random. The Pearson product-moment correlation was .83 ($p < .05$).

Results

Figure 2 displays the distributions of peak frequencies for incorrect [ʃ], correct [s], and correct [ʃ] productions for Apraxic Subjects 1 and 2. Actual values were rounded to the closest multiple of 500 for graphing purposes, but were not rounded for statistical operations. Table 2 displays descriptive statistics for peak frequencies for all subjects. As seen in Figure 2 and Table 2, the majority of Subject 1's incorrect /ʃ/ peak frequencies fell within the 2,000- to 4,500-Hz range, and had a median value of 3,555 Hz. His correct /ʃ/ peak frequencies also fell within this range and had a median of 3,320 Hz. The majority of Subject 1's correct /s/ peak frequencies were between 4,000 and 6,000 Hz, with a median value of 4,258 Hz.

As shown in Figure 2, there was a wide spread of peak frequency values for Subject 2's incorrect [ʃ] productions, with a median value of 4,414 Hz. His correct [ʃ] productions had a slightly higher median of 4,961 Hz, but similar wide spread of values. Almost all of Subject 2's correct [s] peaks fell within the 4,000- to 5,000-Hz range, with a median of 4,296 Hz.

The spread of peak frequency values for the normal subjects is shown in Figure 3 and was much more restricted than the spread for the apraxic subjects. For both normal subjects, there was a clearer separation of values between sounds. As shown in Table 2, Normal Subject 1 had a median value of 2,407 Hz for /ʃ/ and 7,029 Hz for /s/, and Normal Subject 2 had a median value of 2,877 Hz for /ʃ/ and 5,202 Hz for /s/.

Friedman repeated measures tests were conducted for each subject separately to evaluate the differences in the median values among the different sound categories. These analyses examined how each subject varied sounds within his own sound system and not in comparison to the other subjects.

For the normal subjects, statistically significant differences among sounds were found. The peak frequencies for /ʃ/ were significantly lower than those for /s/ at $p < .05$ for both subjects.

For the apraxic subjects, the effect of sound category was significant at $p < .05$ for Subject 1, but it was not significant for Subject 2. Follow-up tests were conducted to determine the nature of the difference among sound categories for Subject 1. Values for correct /s/ were significantly higher than values for incorrect /ʃ/. The remaining pairwise comparisons were not statistically significant.

Results of the perceptual analysis are presented in Table 3. For each prespecified sound category, the percentage of items that the listeners identified as belonging in that category is shown. As seen in Table 3, the unfamiliar listeners' categorizations of Subject 2's productions coincided with our original transcriptions. However, for Subject 1, the listeners identified incorrect [ʃ] productions more often as being [ʃ] rather than [s].

Conclusions

On the basis of this single spectral analysis, it appears that Apraxic Subject 1's /ʃ/ errors were not errors of substitution, but would be more appropriately characterized as distortions. That is, his incorrect [ʃ] productions were acoustically different from his homonymous [s] productions. In contrast, this analysis indicated that Apraxic Subject 2's [ʃ] errors may stem from a sound selection disorder, in that his erroneous productions were not statistically significantly different from his homonymous [s] productions. This finding is in keeping with our perceptual observations of Subject 2. That is, he appeared to treat /s/ and /ʃ/ as a single sound class in that he substituted /ʃ/ for /s/ in clusters, while substituting /s/ for /ʃ/ in singletons. However, given the wide variability seen in peak frequencies of Apraxic Subject 2's incorrect [ʃ] productions, as compared to his [s] productions, we cannot rule out problems in precision of production as a source of his errors.

Whereas this measure of FPSE did differentiate correct /ʃ/ and /s/ productions for our normal subjects, it did not for our apraxic speakers. There could be several reasons for this. We cannot assume that our apraxic subjects' post-treatment correct /ʃ/ productions were produced with *normal* articulatory configurations, so we cannot assume that their productions should appear normal acoustically. Or perhaps examination of spectral peaks was not the best or only analysis we should have used in examining these disordered speakers' productions. We did note that the apraxic subjects' power spectrums appeared to be less peaked, overall, than those of the normal subjects. It is possible that a more general measure of energy shift, such as centroid determination, might be a better reflection of these

FIGURE 2. Distributions of FPSEs for incorrect [ʃ], correct [s], and correct [ʃ] productions for Apraxic Subjects 1 and 2.

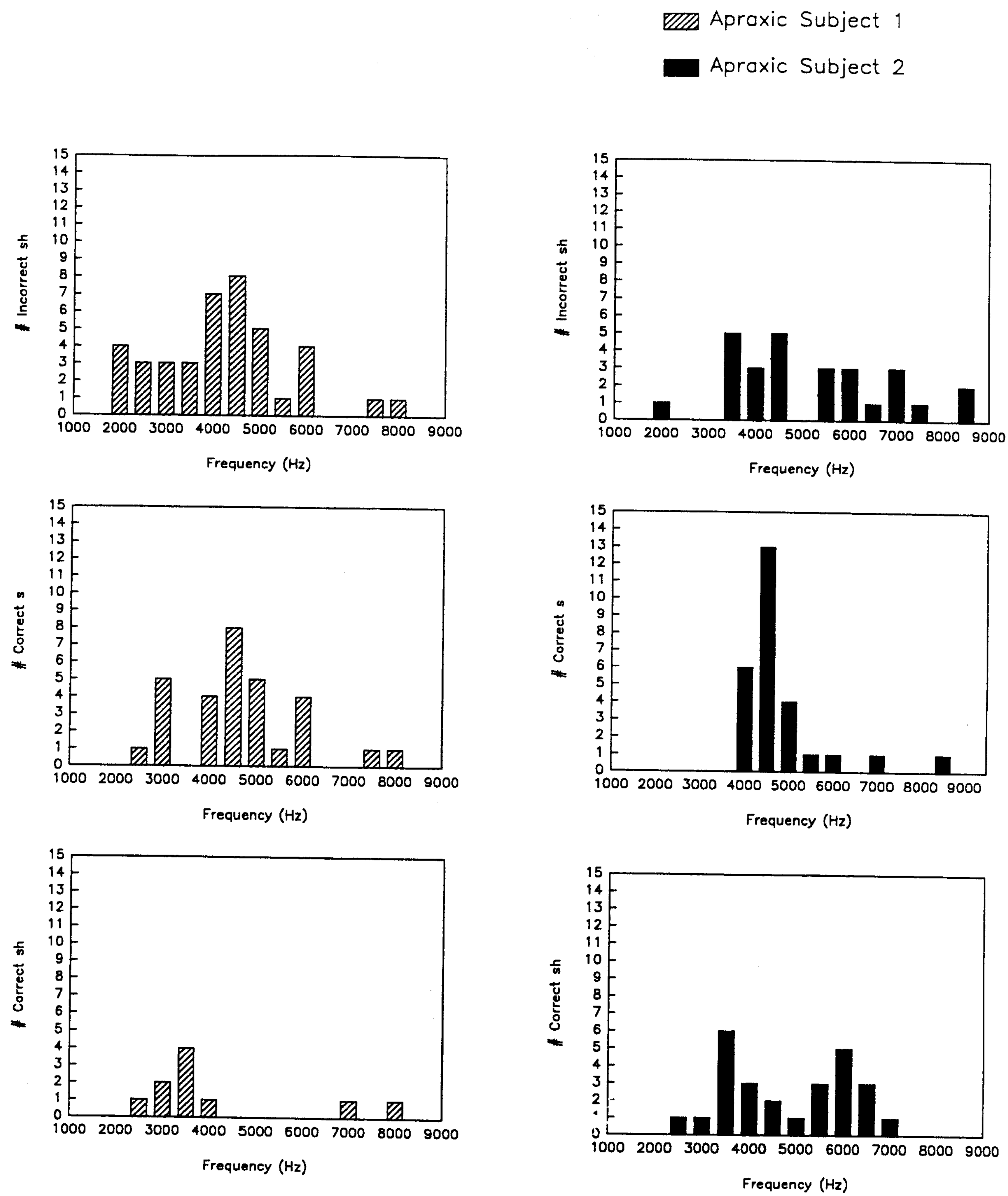
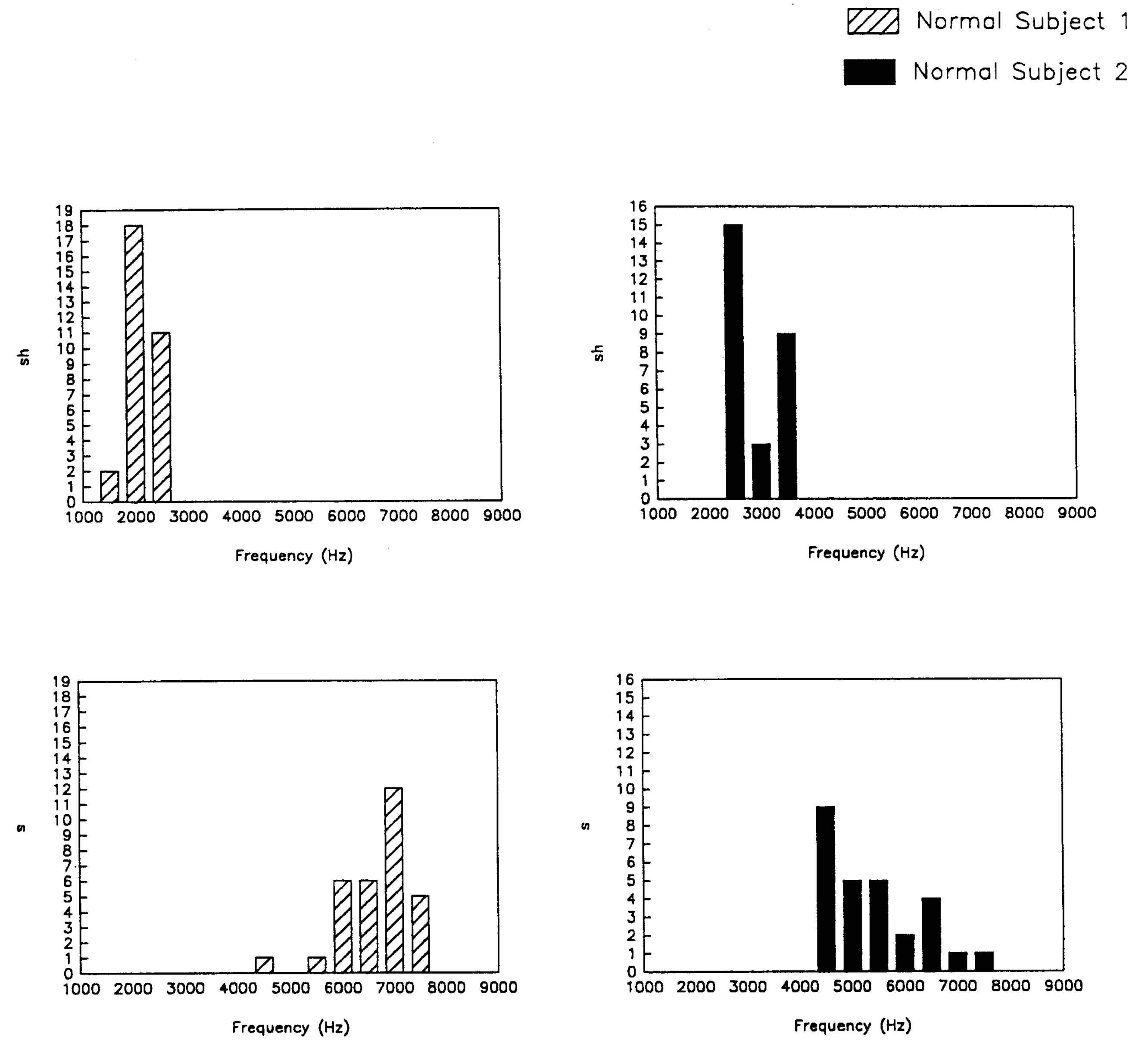


FIGURE 3. Distributions of FPSEs for correct [ʃ] and [s] for Normal Subjects 1 and 2.



productions. Additionally, discriminant analysis of the spectral moments (i.e., mean, variance, skewness, kurtosis) would appear to be warranted, in light of the findings of Forrest and colleagues (1988) with normal speakers. Forrest and colleagues calculated FFTs every 10 ms throughout the frication interval and used the FFTs as random probability distributions to perform discriminant function analyses. They found very high rates of correct classification of /s/ and /ʃ/ (i.e., 93%–100%) using Bark transformed moments obtained from the first 20 ms of the sibilants, with skewness being the most heavily weighted variable. Such analyses might provide more accurate characterizations of the disordered sibilant productions of apraxic/aphasic speakers than a single calculation of FPSE.

A need for additional analyses is particularly evident with Subject 2, in that his correct [ʃ] and [s] productions were clearly differentiated perceptually but not in terms of FPSE. Conversely, the perceptual analysis of Subject 1's productions supported the acoustic findings that incorrect [ʃ] productions appeared to be more like distortions than substitutions.

Our analyses are preliminary and were conducted to determine if this line of research could have clinical utility. Since this investigation was conducted, we have completed a more thorough examination of the productions of an apraxic/aphasic subject who consistently produced all voiced stops and affricates as their voiceless cognates. We examined over 800 productions (i.e., 400 incorrect productions and 400 correct homonym productions) and found very consistent, statistically significant, subphonemic differences between these groups of sounds that would have been impossible to detect perceptually (Wambaugh, Doyle, West, & Kalinyak, 1995). Obviously, these findings should be replicated across many more subjects and many more sounds

TABLE 2. Frequency of the peak spectral energy (Hz): descriptive statistics.

Subject	Sound Category	Mean	Standard Deviation	Median	25–75% Spread
Apraxic Subject 1	incorrect /ʃ/	3427	1147	3555	2451–4301
	correct /s/	4403	1332	4258	3711–5195
	correct /ʃ/	3902	1837	3320	2656–3984
Apraxic Subject 2	incorrect/ʃ/	4965	1682	4414	3692–6269
	correct /s/	4557	1072	4296	4083–4531
	correct /ʃ/	4684	1340	4961	3398–5859
Normal Subject 1	correct /s/	6900	707	7029	6475–7466
	correct /ʃ/	2398	214	2407	2292–2538
Normal Subject 2	correct /s/	5633	919	5202	4843–6254
	correct /ʃ/	3103	410	2877	2758–3541

TABLE 3. Percentage of items identified as belonging to the specified category.

Subject	Incorrect /ʃ/		Correct /s/	Correct /ʃ/
	Identified as /ʃ/	Identified as /s/		
Apraxic Subject 1	69	31	63	87
Apraxic Subject 2	9	91	98	85

Note. Values reflect a mean percentage for both listeners for an individual apraxic subject's productions.

before any firm conclusions regarding the nature of these errors can be drawn.

As noted previously, these apraxic subjects were participants in a treatment research project, specifically an examination of the effectiveness of a minimal contrast treatment (Wambaugh, Doyle, Kalinyak, & West, in press-b). These acoustic analyses were conducted after treatment concluded, so treatment was not designed or modified based on these findings. We can only speculate about the relationship between the subjects' response to treatment and the findings from this investigation. Subject 1 received treatment first for /tʃ/, next for /ʃ/, and finally for /z/, applied in a multiple-baseline design. Minimal contrast treatment was not effective in promoting improved productions of /ʃ/, but was effective for the other two sounds. Subject 2 received treatment first for /ʃ/, then for /r/, and finally for /sw/. For Subject 2, minimal contrast treatment was successful in facilitating increased correct productions of /ʃ/, as well as correct productions of the other sounds. We may speculate that a subject with consistent subphonemic differences between his or her error sound and its homonym, such as Subject 1, may require treatment designed to refine articulatory placement. Our treatment, which focused on sound level contrasts, probably did not provide enough placement instruction. Conversely, a subject who does not differentiate his or her error sound from its homonym may optimally benefit from a treatment program designed to facilitate sound selection, such as minimal contrast.

As indicated earlier, these data are preliminary and any connections between these findings and treatment results are speculative. However, this line of research may have important implications for understanding these errors and subjects' response to treatment.

Acknowledgments

This research was supported by the Department of Veterans Affairs Rehabilitation Research and Development Project #C692-2RA. The authors wish to thank Christopher Moore and Barry Slansky for their assistance with this project.

References

Computerized speech lab [Computer software and hardware]. (1994). Pine Brook, NJ: Kay Elemetrics Corp.

Edwards, H. T. (1992). *Applied phonetics: The sounds of American English.* San Diego: Singular Publishing Group.

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

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84,* 115–123.

Harmes, S., Daniloff, R. G., Hoffman, P. R., Lewis, J., Kramer, M. B., & Absher, R. (1984). Temporal and articulatory control of fricative articulation by speakers with Broca's aphasia. *Journal of Phonetics, 12,* 367–385.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricatives. *Journal of the Acoustical Society of America, 33,* 589–596.

Hoole, P., Ziegler, W., Hartmann, E., & Hardcastle, W. (1989). Parallel electropalatographic and acoustic measures of fricatives. *Clinical Linguistics and Phonetics, 3,* 59–69.

Hughes, G. M., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America, 28,* 303–310.

McNeil, M. R., & Kent, R. D. (1990). Motoric characteristics of adult aphasic and apraxic speakers. In G. R. Hammond (Ed.), *Advances in psychology: Cerebral control of speech and limb movements* (pp. 349–386). New York: Elsevier/North Holland.

Shinn, P., & Blumstein, S. E. (1983). Phonetic disintegration in aphasia: Acoustic analysis of spectral characteristics for place of articulation. *Brain and Language, 20,* 90–114.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America, 70,* 976–984.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America, 50,* 1180–1192.

Wambaugh, J. L., Doyle, P. J., Kalinyak, M. M., & West, J. E. (in press-a). A critical review of acoustic analyses of aphasic and/or apraxic speech. *Clinical Aphasiology.*

Wambaugh, J. L., Doyle, P. J., Kalinyak, M. M., & West, J. E. (in press-b). A minimal contrast treatment for apraxia of speech. *Clinical Aphasiology.*

Wambaugh, J. L., Doyle, P. J., West, J. E., & Kalinyak, M. M. (1995). Voicing errors in aphasic/apraxic speech: An in-depth VOT analysis. Manuscript in preparation.

Weismer, G. (1984). Acoustic analysis strategies for the refinement of phonological analysis. *ASHA Monographs, 22,* 30–52.

Ziegler, W., & von Cramon, D. (1986). Disturbed coarticulation in apraxia of speech: Acoustic evidence. *Brain and Language, 29,* 34–47.

Contact author: Julie L. Wambaugh, PhD, Aphasia Rehabilitation Research Laboratory (126), 7180 Highland Drive, VA Medical Center, Pittsburgh, PA 15206

Key Words: apraxia, acoustics, aphasia, speech, spectral analysis

Appendix A

Words Used in Analyses: Subject 1

Correctly Produced Word Final /ʃ/	Incorrectly Produced Word Final /ʃ/		Correctly Produced Word Final /s/	
ash 1	ash 1	mash 2	ass 1	mass 2
leash 1	ash 2	mash 3	ass 2	mass 3
leash 2	ash 3	mash 4	ass 3	mass 4
leash 3	ash 4	mash 5	ass 4	mass 5
leash 4	ash 5	mash 6	ass 5	mass 6
leash 5	ash 6	mash 7	ass 6	mass 7
leash 6	ash 7	mash 8	ass 7	mass 8
push 1	ash 8	push 1	ass 8	puss 1
push 2	ash 9	push 2	ass 9	puss 2
rush 1	leash 1	rush 1	lease 1	Russ 1
rush 2	leash 2	rush 2	lease 2	Russ 2
	leash 3	rush 3	lease 3	Russ 3
	leash 4	rush 4	lease 4	Russ 4
	leash 5	rush 5	lease 5	Russ 5
	mash 1	rush 6	mass 1	Russ 6

Appendix B

Words Used in Analyses: Subject 2

Correctly Produced Word Initial /ʃ/		Incorrectly Produced Word Initial /ʃ/		Correctly Produced Word Initial /s/	
shack 1	sheep 1	shack 1	sheep 1	sack 1	seep 1
shack 2	sheep 2	shack 2	sheep 2	sack 2	seep 2
shack 3	sheep 3	shack 3	sheep 3	sack 3	seep 3
shed 1	shy 1	shed 1	shy 1	said 1	sigh 1
shed 2	shy 2	shed 2	shy 2	said 2	sigh 2
shed 3	shine 1	shed 3	shine 1	said 3	sign 1
shave 1	shock 1	shave 1	shock 1	save 1	sock 1
shave 2	shock 2	shave 2	shock 2	save 2	sock 2
shave 3	shock 3	shave 3	shock 3	save 3	sock 3
shave 4	shock 4	shave 4	shock 4	save 4	sock 4
she 1	shoot 1	she 1	shoot 1	see 1	suit 1
she 2	shoot 2	she 2	shoot 2	see 2	suit 2
she 3		she 3		see 3	
she 4		she 4		see 4	
she 5		she 5		see 5	