Acoustic Analysis of Stress Contrasts Produced by Hearing-Impaired Children

772

Angela La Bruna Murphy, Nancy S. McGarr, and Fredericka Bell-Berti

Acoustic measures (vowel duration, peak F_0, and peak amplitude) and stress judgments were made of spondaic words with alternating syllabic stress produced by three hearing-impaired subjects and a hearing control. When these subjects correctly produced stressed syllables, they did so with increased amplitude, duration, and F_0. When they did not, no acoustic feature (or combination of features) accounted for their errors. The data suggest that there are important between-speaker differences in the overall hierarchy of cues adopted to convey lexical stress.

In a normal-hearing person's speech, the speech signal is structured at the suprasegmental level by features that serve to organize sequences of syllables into words and phrases. One suprasegmental characteristic, lexical stress, seems to be an important cue to word identification (Gaitenby & Mermelstein, 1977). Studies of hearing speakers have established the importance of three acoustic cues (vowel duration, amplitude, and fundamental frequency [F_0]) in conveying lexical stress. For example, Fry (1955), measuring stress changes in words such as *object* and *subject*, concluded while both intensity and duration serve to cue stress, duration was the more salient cue. However, Bolinger (1958) proposed that pitch prominence was primary. Lehiste (1976) noted that stress production is usually associated with an increase in amplitude, F_0, and vowel duration. Gaitenby and Mermelstein (1977) found F_0 to be a weak cue to lexical stress, proposing that intensity was the most important cue, followed by duration. Gelfer (1980), measuring the acoustics of contrastively stressed and unstressed syllables, found that although all utterances were perceived as having stress on the intended syllable, there were significant differences among the speakers in the acoustic manifestations of lexical stress.

With respect to hearing-impaired persons, there have been numerous studies of segmental and suprasegmental production (cf. Osberger & McGarr, 1982). These studies, often descriptive in nature, note that *speech* produced by persons with severe to profound hearing loss is "staccato-like," suggesting a failure to differentiate stressed and unstressed syllables. Another common descriptor is "flat and devoid of melody" suggesting a failure to vary F_0. Ando and Canter (1969) reported that hearing-impaired subjects produced only stressed syllables. Similarly, Osberger and Levitt (1979) found that hearing-impaired speakers did not distinguish

---

Ms. Murphy is Speech Pathologist at the School for Language and Development in Bellmore, NY. Dr. McGarr is Senior Research Associate at Haskins Laboratories in New Haven, CT, and Dr. Bell-Berti is Professor in the Speech, Communication Science, and Theater departments of St. John's University in Jamaica, NY.

between stressed and unstressed syllables. Nickerson (1975), on the other hand, reported that stressed syllables produced by hearing-impaired speakers were usually accompanied by an F_0 change. McGarr and Harris (1983) studied a hearing-impaired speaker whose acoustic measures of stress contrasts did not reveal one "wrong" stereotypic acoustic pattern (e.g., failure to use F_0). Rather, the acoustics of the productions were variable from token to token. Perceptual errors could not be correlated with a consistently incorrect pattern of F_0, amplitude, or duration. Moreover, even when acoustic cues were correct, they were not always transparent to the listener. Interestingly, the speech acoustic analysis of an adventitiously deafened subject prior to receiving a cochlear implant (Leder, Spitzer, Milner, Flevaris-Phillips, Richardson, & Kirschner, 1986) showed the same results by McGarr and Harris (1983). Variability decreased post-implant.

Since either the absence or the correct use of lexical stress is expected to affect the speech intelligibility, and since there is little evidence as to the various acoustic correlates of stress produced by hearing-impaired persons, this study was undertaken. The purpose was threefold: First, to obtain perceptual judgments from normal-hearing listeners of lexical stress produced in utterances by severely-to-profoundly hearing-impaired talkers; second, to make acoustical measures (amplitude, F_0, and duration) of the talker's production; third; to compare the perceptual and acoustical measures.

Methods

Subjects. Two groups of subjects were used in this study: Talkers and listeners.

Talkers: The speakers were 7 females and 6 males ranging in age from 9 to 19 years (mean age of 13.9 years) with congenital profound hearing loss. The pure tone averages of the subjects ranged from 92 dB to 118 dB HL in the right ear and 85 dB to 120 dB HL in the left ear (ANSI, 1969). The speech of three of these subjects, S1 (15 years, PTA=60 dB HL), S2 (18 years, PTA=92 dB HL), and S3 (12 years, PTA=58 dB HL), was subjected to acoustic analysis. All subjects had been attending the Lexington School for the Deaf for a minimum of two years before participating in the study. In addition, one normal-hearing adolescent (a 16-year-old female) was recorded as a control.

Listeners: The listeners were 45 persons who ranged in ages from 16 to 60 years. All reported having normal hearing, were native speakers of English, and were inexperienced with respect to the speech of hearing-impaired persons. An inexperienced listener was defined as having no formal training in, or extensive exposure to, the speech of hearing-impaired persons (McGarr, 1983). Listeners also judged the recordings of the hearing control to ensure that her productions were perceptually correct.

Speech Samples. The hearing-impaired subjects recorded the speech samples in an anechoic chamber. An experienced teacher served as recorder and conveyed instructions to the subjects both orally and in sign language. Each subject produced three nominally spondaic words, *cupcake*, *bathtub*, and *hotdog*, 10 times each, systematically alternating primary lexical stress between the first and second syllables. This resulted in a total of 60 productions for each of the 14 speakers (3 words x 2 stress patterns x 10 repetitions = 60 productions).

These nominally spondaic words were used for several reasons. First, they were highly familiar to hearing-impaired individuals. Second, such stimuli did not require that the subjects have the semantic sophistication needed to distinguish between such traditionally used noun-verb pairs such as *OBject* and *obJECT* (e.g., Fry, 1955), when they are produced in isolation. The words were not produced in a carrier phrase because we wanted to have them produced in one breath group (Whitehead, 1983). Finally, the speakers for this study were also participants in a parallel study examining the relationship between speech perception and production in hearing-impaired individuals; thus, it was necessary to use parallel utterances in both phases of the research (Rubin-Spitz, McGarr, & Youdelman, 1986). For similar reasons, such spondaic stimuli have also been used in an extensive study of normally-hearing children's acquisition of phonological stress contrasts (Atkinson-King, 1973).

Perceptual Test. Ten listeners judged the speech of each talker; however, each listener judged the recordings of no more than three talkers. Furthermore, the talker's position in a listening sequence varied so that no talker was heard in any sequence position more than four times. This design was used in order to minimize learning effects. The subjects were asked to attend to each utterance and to indicate on an answer sheet if stress occurred on the first or second syllable of the utterance. The listeners were given practice words to ensure that they understood the instructions and were able to perform the task. The results of the listeners' perceptual evaluations of the talkers were used to select three hearing-impaired talkers for detailed acoustic analysis. The results were also used to identify the individual tokens as perceptually correct, perceptually incorrect, and perceptually equivocal stimuli. If 8 to 10 listeners judged an utterance as having stress on the intended syllable, it was coded as being correct; if 3 to 7 listeners judged an utterance as having stress on the intended syllable, it was coded as equivocal; if 0 to 2 listeners judged an utterance as having stress on the intended syllable, it was coded as incorrect. These data thus permitted an examination of the acoustic correlates of stress patterns perceived as correct and incorrect stress patterns. (Note that when judged incorrect, the token was identified as the minimally contrastive target word; for example, *CUPcake* was identified as *cupCAKE*.)

Acoustic Analysis. The hearing-impaired talkers' productions and also the hearing controls were analyzed acoustically using the Interactive Laboratory System (ILS) software and also waveform analysis software developed at the Center for Research in Speech and Hearing Services, Graduate Center, CUNY, for use on an IBM-XT microcomputer. ILS was used for A/D-D/A data conversion (at a sampling of 10K samples per second) and spectral analysis. The CUNY software provided token location, automatic measurement of amplitude and F_0 over 100 msec intervals, and a visual display of the waveform either alone or with F_0 and amplitude. Peak vowel amplitude measurements were taken directly from the CUNY automatic measures, at a visually identified 10 msec interval within a stable amplitude region, generally near the vowel midpoint. For F_0, the automatic measurements, obtained by an algorithm that identified the pitch period, were used whenever possible. In difficult cases, usually occurring with pitch breaks or extremely high fundamentals, either the pitch period was identified visually or ILS was used to obtain a graphical display and the spectral measurements for the frequency ranges of 0 to 1 KHz. Vowel duration measurements were made using ILS with 1 msec resolution.

Duration measures were made in *cupcake*, *bathtub*, and *hotdog* by placing cursors at the beginning and end of each vowel and determining the distance (in msec) between them. For *cup*, *cake*, *bath*, *tub*, and *dog*, the beginning of the vowel was

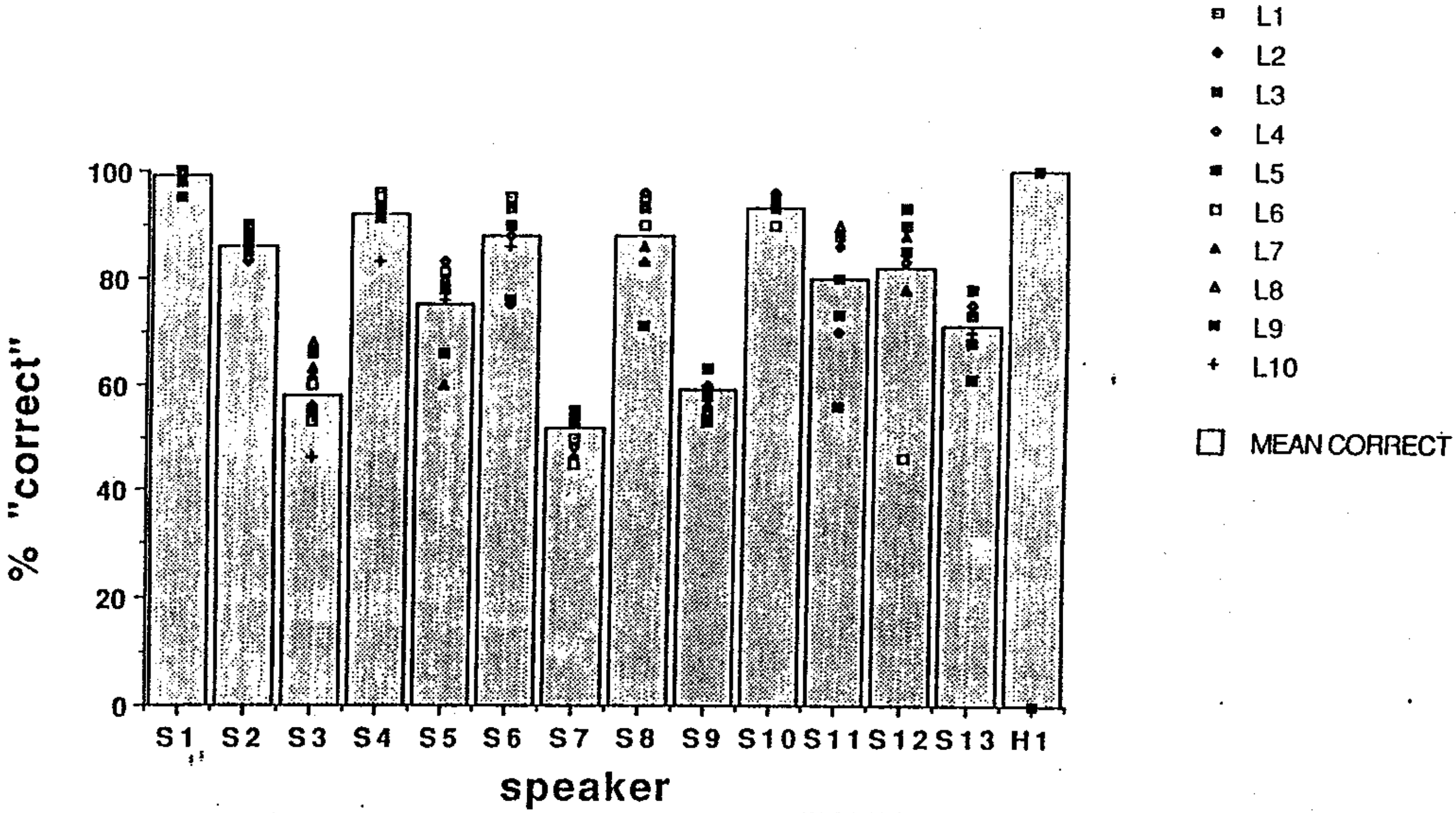

Figure 1. Individual listener judgments (for the 10 listeners, indicated with the small symbols) and pooled listener judgments (stippled bars) for 13 hearing-impaired speakers (S1-S13) and one normal-hearing speaker (H1).

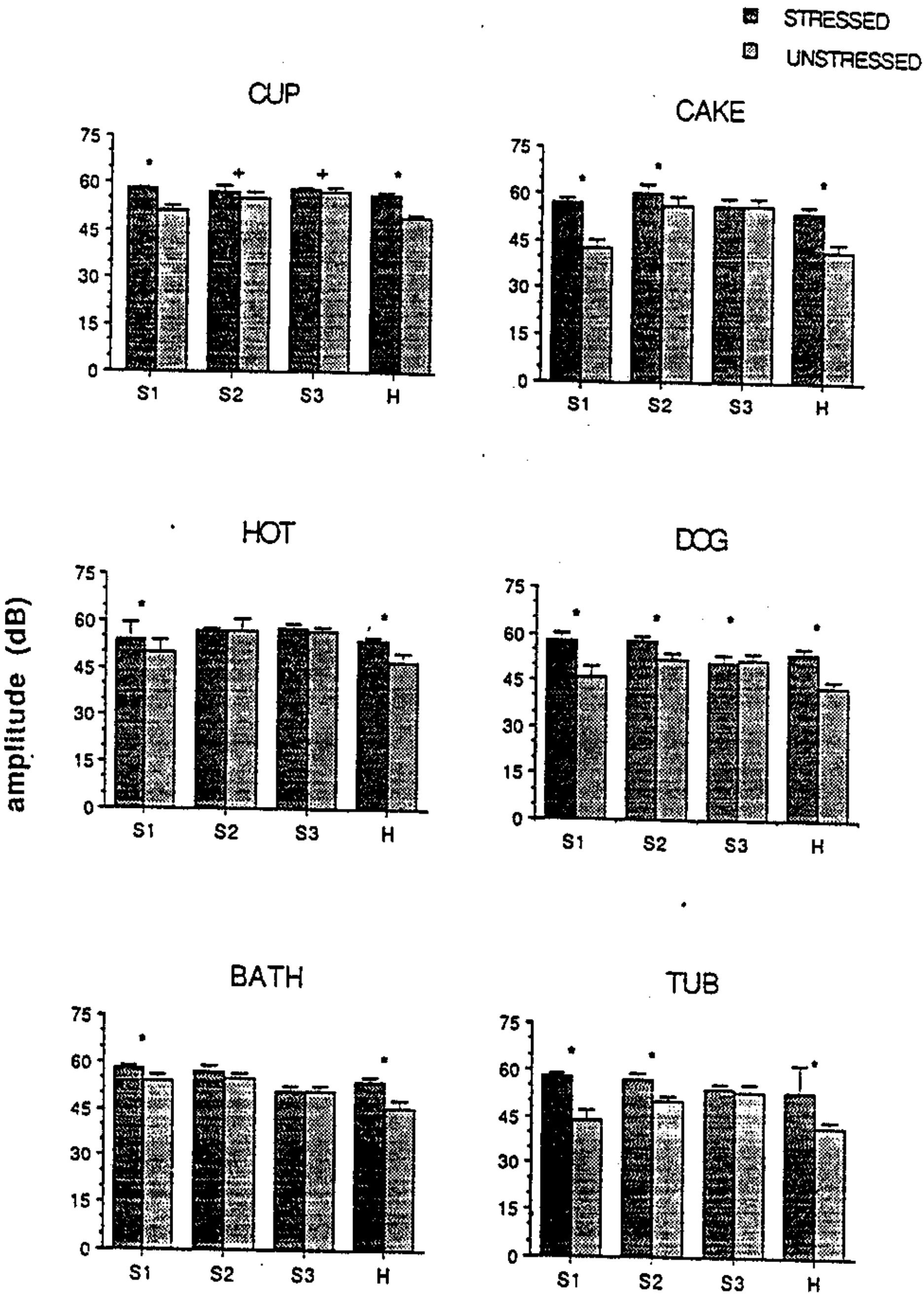

Figure 2. Mean amplitudes and standard deviations for each syllable for each of three hearing-impaired speakers and one normal-hearing speaker.
*indicates differences significant at p<0.001.

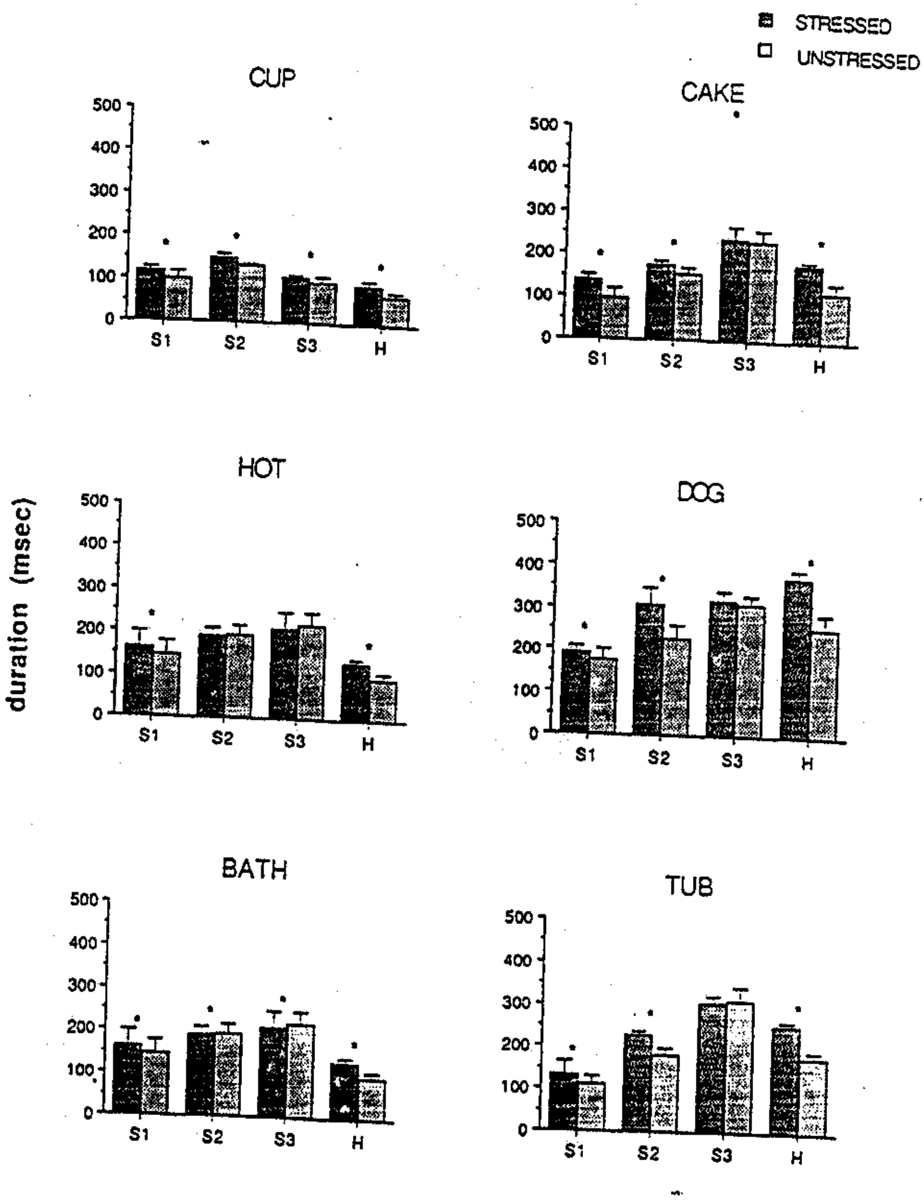

Figure 3. Mean vowel durations and standard deviations for each syllable of three hearing-impaired speakers and one normal-hearing speaker.
*indicates differences significant at p.<0.001

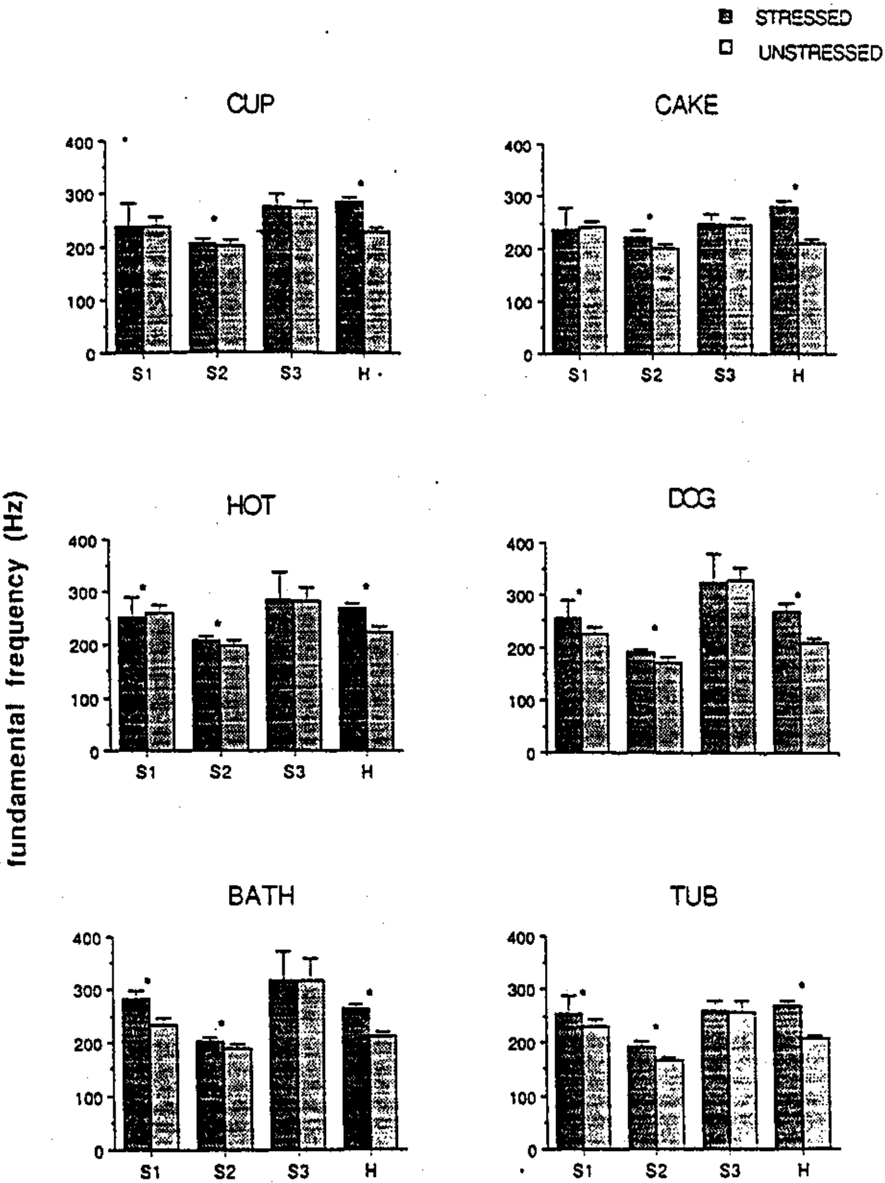

Figure 4. Mean F_0 and standard deviations for each syllable for each of three hearing-impaired speakers and one normal-hearing speaker.
*indicates differences significant at $p<0.001$.

identified as the onset of glottal pulsing (periodicity in the waveform) following the initial stop release. For *hot*, vowel onset was identified as the point at which fricative noise was no longer evident in the waveform, since in English, /h/ is often produced with vocal fold edge vibrations, leading to the presence of periodicity in the waveform during /h/. The end of the vowel was identified as the point at which higher frequency format energy was no longer evident in the waveform (suggesting vocal tract constriction for the consonant). Cursor position was specified as time (in msec) from the beginning of the sample; vowel duration was determined by calculating the difference between the points.

Results

Perception. The three hearing-impaired speakers whose tokens were acoustically analyzed were selected based on the results of the perception tests (Fig. 1). Individual listeners' scores for each speaker were generally quite similar. The lowest individual score for any speaker was 45%; the highest score was 100%. Hearing-impaired speakers, S1 and S2, were chosen because they achieved two of the highest scores with close listener agreement. That is, syllables that the subjects intended to stress were perceived by the listeners as stressed. The third hearing-impaired subject, S3, was chosen because his scores were among the most variable.

Production. Because the vowels in the first and second syllables of the test words have intrinsically different amplitude, duration, and F_0, primary comparisons of acoustic measures were made between stressed and unstressed first syllables and between stressed and unstressed second syllables. That is, we compared the mean amplitude, duration, and F_0 of *CUP* and *cup*, *CAKE* and *cake*, *HOT* and *hot*, *DOG* and *dog*, *BATH* and *bath*, *TUB* and *tub*, (Figs. 2, 3, and 4 respectively). T-tests were used to determine the significance of the differences between the means for each subject for each syllable pair.

For the Hearing Control, (H1), the three figures show that amplitude, duration, and F_0 were significantly greater for the stressed syllable than the unstressed syllable ($p<0.001$).

Hearing-impaired Subject 1, (S1), generally followed the same pattern as the hearing subject: T-tests revealed statistically significant differences ($p<0.001$) between means in 15 of 18 comparisons. However, F_0 comparisons for *cake* and *hot* show a reversal of this pattern, and *cup* has the same F_0 in both stress conditions (Fig. 4). Note however, that listeners perceived her productions as intended, undoubtedly because amplitude and duration cues were appropriate for the stressed syllable.

Hearing-Impaired Subject 2, (S2), also followed the same general pattern as the hearing control: T-tests revealed statistically significant differences ($p<0.001$) between means in 14 of 18 comparisons. This subject differed from the hearing control in his production of stressed versus unstressed *hot*: The mean vowel duration was slightly greater (not significantly) for the unstressed syllable (Fig. 3), and the mean amplitude was the same for the unstressed and stressed syllables (Fig. 2). Not surprisingly, some of the utterances were perceived as incorrect when he attempted to produce *hotDOG*.

Hearing-impaired Subject 3, (S3), frequently failed to differentiate between stressed and unstressed syllables in amplitude, duration, or F_0 (Figs. 2, 3, & 4). T-tests revealed ($p<0.001$) between the means in only 4 of 18 comparisons. For this subject, 8 of the 10 tokens of *cupCAKE* (i.e., stress intended on the second syllable) were identified as *CUPcake* (as if stress were intended on the first syllable). To try

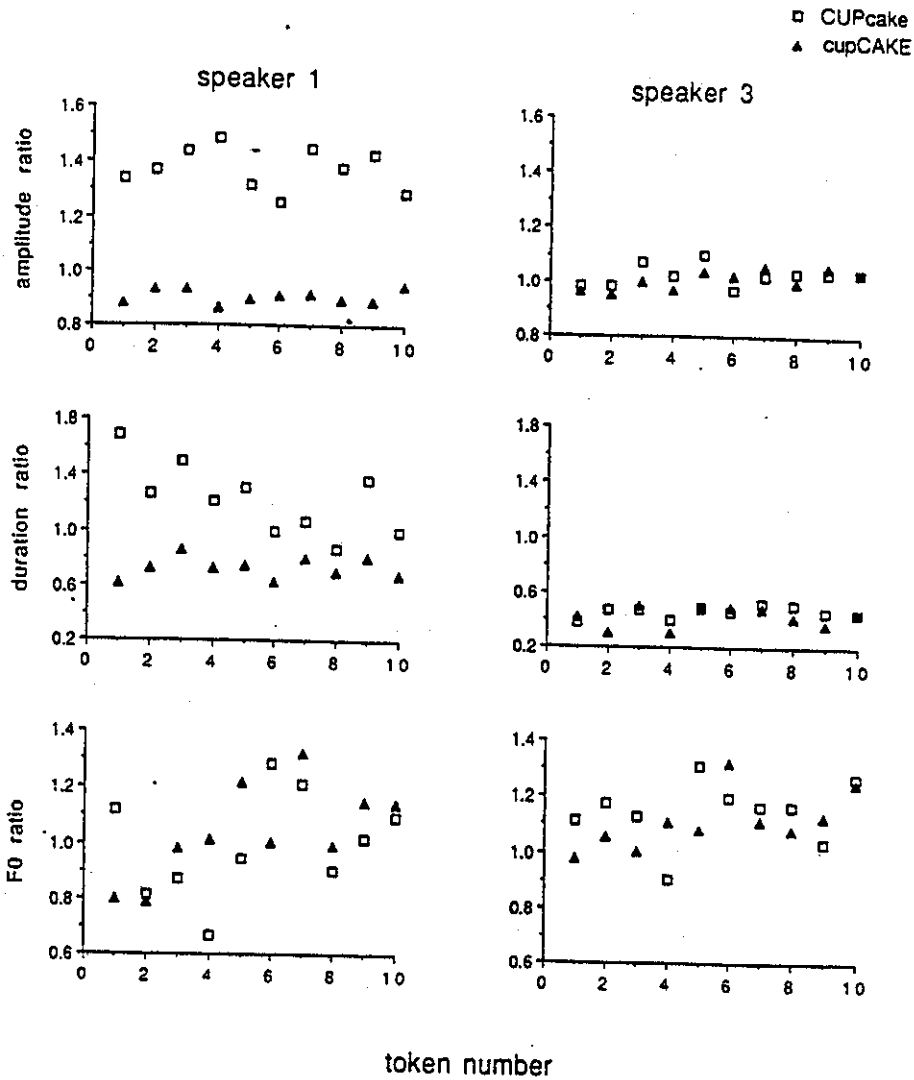

Figure 5. Amplitude, duration, and F_o ratios (syllable 1/syllable 2) for *CUPcake* and *cupCAKE* for S1 and S3.

to understand this result, we compared amplitude, duration, and F_0 ratios (syllable 1 divided by syllable 2) for this subject and for S1, all of whose tokens of this pair of utterances were identified as intended (Fig. 5). Perhaps listeners were unable to identify the intended stress of *cupCAKE* because the relative amplitude and duration of these syllables of these tokens were essentially the same as those of *CUPcake*. Since the amplitude and duration ratios for the tokens of *CUPcake* are low in relation to those for S1, we might speculate that S3's productions were identified as having primary stress on the first syllable because of an expectation of stress in that position of this lexical item.

Discussion

Normal talkers seem to differ in the primary acoustic cues they use to convey lexical stress. For example, Fry (1955) and Nakantani and Shaffer (1978) suggested that changes in vowel duration are used as the primary cue to lexical stress. Lehiste (1976) suggests that F_0 changes are equally important as changes in vowel duration. Alternatively, Gaitenby and Mermelstein (1977) concluded that amplitude changes are the primary cue to lexical stress. Gelfer (1980) noted that subjects may vary in the primary acoustic cue they use. These differences among subjects do not, however, effect perception—utterances are perceived as correct. In this study, differences among speakers were evident in the productions of hearing-impaired subjects. While the subjects tended to produce stressed syllables with increased amplitude, longer duration, and increased F_0 than the corresponding unstressed syllables, when a failure of this acoustic pattern occurred, it was variable in nature. Moreover, even when utterances differed from the expected acoustic patterns, a subject's utterances were not always perceived as incorrect. Thus, the relationship between listener judgment and acoustics was not always straightforward. This has been previously noted in the speech of hearing-impaired persons (e.g., McGarr & Harris, 1983).

It is true, however, for these three hearing-impaired speakers, that when they did not increase amplitude, duration, and F_0 for the lexically stressed versus unstressed syllable, the listener responses were more variable. For example, 7 to 10 listeners perceived S2 as incorrect for the tokens in which vowel duration was shorter for the stressed than the unstressed syllable. S3, who had great difficulty in conveying intended stress patterns, showed variable patterns of amplitude, duration, and F_0, although the tokens were often not perceived as intended.

In general, then, the hearing-impaired subjects often used a similar pattern of increased amplitude, duration, and F_0 for stressed syllables as the hearing subject. But when this acoustic pattern was not followed, the breakdown pattern was speaker specific. For example, S3 did not increase duration consistently for all of the stressed syllables, and S1 and S2 failed to increase F_0 consistently. For S1 and S3, however, it is interesting to note that this effected perception in only some instances, perhaps indicating that F_0 plays only a secondary role in conveying lexical stress, at least in this sample.

Clinically, these data would seem to bear on issues of diagnosis and treatment. For example, differences between subjects may be important when interpreting the literature on the speech production abilities of hearing-impaired persons. Furthermore, it is common to take one representative token as a sample of a subject's production. Based on such a restricted sample, one acoustic correlate may be afforded greater significance than another, or a particular speaker may be faulted for failure to use a standard cue for stress (e.g., failure to vary F_0). It is possible that, just as hearing speakers, hearing-impaired talkers may marshall a different hierarchy of acoustic cues to signal stress.

It is also important to note that the number of studies of the acoustic correlates of stress in the speech of hearing-impaired persons are limited. While it is not possible to make general statements as to the homogeneity or heterogeneity of stress-contrast production in hearing-impaired persons, our data suggest that there are the same type of individual differences as in the normal-hearing population. These differences among speakers may be of critical importance for teachers of hearing-impaired persons to consider when working on suprasegmental characteristics of speech. For example, before implementing a specific treatment program, one ought to try to determine if a stress contrast might be conveyed by a set of acoustic cues best for a particular child. An aggregate of acoustic cues (amplitude, duration, and F_0) ultimately determine perceptual judgments. However, it may help to obtain listeners' judgments of multiple productions of the same tokens. If listeners agree that the tokens are perceptually correct, then the issue of which acoustic cue predominates is moot. If the listeners perceive the utterance as incorrect or are equivocal in their judgments, then the speech teacher must decide which acoustic parameter is most likely to lead to successful perception. For example, one child may receive the greatest benefit from working on increasing amplitude contrasts between stressed and unstressed syllables, whereas another child may be more intelligible if he increases duration contrasts between stressed and unstressed syllables.

In this study, when listeners judged hearing-impaired speakers to be incorrect or equivocal, the speakers tended to have variable productions (e.g., S3 in this study, and also the hearing-impaired speaker in McGarr & Harris, 1983). The relationship between production variability and listener judgment is a complex one. However, stabilizing productions in any acoustic domain (i.e., achieving consistent differences in duration, amplitude, or F_0 for stressed versus unstressed syllables) would be helpful. Furthermore, within-subject variability suggests that speech training should provide opportunity for multiple repetitions or practices of the stimuli.

Our data suggest that it is important to appreciate the significance of the individual differences among hearing-impaired students. Although all three subjects shared common degrees of hearing impairment, it would seem inappropriate to follow the same speech remediation program with all three. Instead, individual assessment procedures, based on multiple productions of stimuli, rather than traditional assessment procedures, should be used to determine the course of remediation that will best help each student to achieve more intelligible speech.

ACKNOWLEDGMENTS

This study was completed in partial fulfillment for a Master of Arts Degree at St. John's University, and we thank Drs. Mary Boyle and Donna Geffner for their help. We also thank Marion MacEachron and Mark Weiss. Ms. MacEachron provided valuable comments on the manuscript as well as programming support. We also wish to thank the staff and students of the Lexington School for the Deaf who so graciously gave of their time. This research was funded by Grant NS 17764 to The City University of New York and by NS 13617 to Haskins Laboratories.

REFERENCES

Ando, K., & Canter, G. (1969). A study of syllabic stress in some English words as produced by deaf and normally hearing speakers. *Language and Speech, 12*, 247-255.

ANSI (1969). American National Standards Institute Specifications for Audiometers (ANSI S3.6). New York, NY: ANSI.

Atkinson-King, K. (1973). Children's acquisition of phonological stress contrast. *UCLA Working Papers in Phonetics, No. 25.*

Bolinger, D.L. (1958). A theory of pitch accent in English. *Word, 14*, 109-149.

Fry, D.B. (1955) Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 17*, 765-768.

Gaitenby, J.H., & Mermelstein, P. (1977). Acoustic correlates of perceived prominence in unknown utterances. *Haskins Laboratories Status Report of Speech Research, 49*, 210-216.

Gelfer, C.E. (1980). The ability to vary stress in a subject with apraxia of speech. *Journal of the Acoustical Society of American, 67*, S79 (A).

Leder, S.B., Spitzer, J.B., Milner, P., Flevaris-Phillips, C., Richardson, F., & Kirshner, J.C. (1986). Reacquisition of contrastive stress in an adventiously deaf speaker using a single-channel cochlear implant. *Journal of the Acoustical Society of America, 79*, 1967-1973.

Lehiste, I. (1976). Suprasegmental features of speech. In N.J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*, *225–239*. New York: Academic Press.

McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26*, 451–458.

McGarr, N.S., & Harris, K.S. (1983). Articulatory control in a deaf speaker. In I. Hochberg, H. Levitt, & M.J. Osberger (Eds.), *Speech of the Hearing Impaired: Research, Training and Personnel Preparation, 75–97.* Baltimore, MD: University Park Press.

Nakatani, L.H., & Shaffer, J.A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234-245.

Nickerson, R. (1975). Characteristics of the speech of deaf persons. *Volta Review, 77*, 342-362.

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

Osberger, M.J., & McGarr, N.S. (1982). Speech production characteristics of the hearing impaired. In N. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice, Vol. 8*, 221-283. New York: Academic Press.

Rubin-Spitz, J., McGarr, N.S., & Youdelman, K. (1986). Perception and stress contrasts by the hearing and the hearing impaired. *Journal of the Acoustical Society of America, 79*, (S1) (A).

Whitehead, R. (1983). Some respiratory and aerodynamic patterns in the speech of hearing-impaired speakers. In I. Hochberg, H. Levitt, & M.J. Osberger (Eds.), *Speech of the Hearing Impaired: Research, Training and Personnel Preparation.* Baltimore, MD: University Park Press.