Use of Prosody by Children with Severe Dysarthria: A Cantonese Extension Study

Tara L. Whitehill, Ph.D.
University of Hong Kong

Rupal Patel, Ph.D.
Northeastern University, Boston

Jenny F-C. Lai, B.Sc.
University of Hong Kong

Several recent studies have shown that speakers with severe dysarthria can make use of prosody to communicate intentions. Patel and Salata (2006) investigated the abilities of five children with severe dysarthria to convey three pitch levels and three duration lengths to their caregivers, using an interactive computer game. In this extension study, four Cantonese-speaking children with severe dysarthria were asked to produce five durations and five pitch levels. Rising and falling pitch levels were included to simulate the rising and falling lexical tones of Cantonese. Acoustic analysis was performed for all productions to determine the relationship between acoustic cues and listener perception. There was substantial interparticipant and intraparticipant variability in the children's productions and in caregiver accuracy. However, accuracy was above chance level for a number of prosodic targets for each speaker. Some of the perceptual responses could be easily explained by the acoustic data, where clear distinctions were apparent. However, in other cases there was much overlap in the acoustic signal, suggesting that listeners were employing additional cues. The results of this study strengthen previous findings that prosody can be used reliably by speakers with severe dysarthria to communicate with familiar partners. Further cross-linguistic studies would be informative.

The dimension of prosody has been relatively neglected in studies of dysarthric speech (Leuschel & Docherty, 1996). Investigations of prosody in this population are of both theoretical and practical interest. Previous studies have shown that speakers with severe dysarthria characterized by reduced segmental intelligibility may be able to make use of prosodic cues to convey their intentions (cf., Le Dorze, Ouellet, & Ryalls, 1994; Patel 2002a, 2003, 2004; Yorkston, Beukelman, Minifie, & Sapir, 1984). Patel (2002a) showed that adults with severe dysarthria were able to consistently produce sustained vowels at three distinctive durations and two distinctive pitches. These same speakers could also signal the question-statement contrast using prosodic cues, and listeners were highly accurate at classifying their productions (Patel 2002b, 2003). This preserved ability to signal prosodic contrasts has also been noted in contrastive stress tasks in which speakers with dysarthria utilize combinations of fundamental frequency (F0), intensity, and duration to signal stress (Patel, 2004; Patel & Watkins, 2007; Yorkston et al., 1984).

More recently, Patel and Salata (2006) used an interactive computer game to investigate communicative interactions between five children with

Journal of Medical Speech-Language Pathology
Volume 16, Number 4, pp. 293–299

severe dysarthria and their caregivers. The children were asked to produce the sustained vowel /a/ at three distinct pitch levels (high, medium, low) and three distinct durations (long, medium, short), as well as nine combinations of these targets, which their caregivers were asked to identify. Caregiver accuracy was variable across the five dyads, but accuracy was above chance level for several targets produced by each speaker. Accuracy for pitch and duration targets in isolation was generally higher than for combined conditions. The current study was designed as an extension study of Patel and Salata's work. Since caregiver accuracy was relatively high for the three duration targets in the Patel and Salata study, we added two further duration parameters. Also, we added two pitch categories, rising and falling, to simulate the lexical tone patterns of Cantonese.

Cantonese has six contrastive lexical tones: high level, mid level, low level, high rising, low rising, and low falling. We were interested in exploring the issue of whether tonal-language listeners are better at perceiving pitch differences than nontonal language listeners. Francis and Ciocca (2003) found that monolingual English-speaking listeners and native Cantonese-speaking listeners performed similarly in tasks that involved discrimination of speech sounds differing in F0; the Cantonese listeners were less sensitive than the English listeners to pitch differences for nonspeech sounds. Although the current study did not systematically compare Cantonese- and English-speaking dyads, it allowed preliminary investigation of this issue in a dysarthric population. Specifically, the present study sought to determine whether Cantonese-speaking children with severe dysarthria could produce five distinct pitch categories and five distinct duration lengths and whether their caregivers could accurately identify their intentions.

METHOD

Participants

Four boys with severe dysarthria secondary to cerebral palsy and their mothers served as participants. The boys ranged in age from 7 to 13 years. They all used augmentative and alternative communication (AAC) as their primary means of communication and had adequate hearing, vision, and cognitive skills to complete the tasks. Cantonese (Chinese) was the primary language of the boys and their mothers.

Materials and Procedures

The materials and procedures were based on those developed and reported by Patel and Salata (2006). Briefly, a modified "Wizard of Oz" methodology was used in which children were led to believe that their vocalizations could control the movements of characters in a simple computer game. In actuality, the mother's interpretation of her child's intended vocalization resulted in character movements. In the pitch game, children were asked to produce sustained /a/ at five pitch levels (high, mid, low, rising, falling), corresponding to each fish in the game (Figure 1, A and B). Similarly, in the duration game, five distinct sustained /a/ lengths were requested (shortest, short, mid, long, longest) (see Figure 1, C and D). Five randomly ordered repetitions were requested for each distinct pitch and duration target. During each game, the mothers were seated in front of a separate computer but in the same room as their child. They were not privy to information about which target their child was asked to produce. Upon hearing each vocalization, they were asked to classify the production as one of the five distinct targets. This differed from the procedures in the Patel and Salata (2006) study, where mothers were in a separate room and heard the productions via headphones, and thus had access to only acoustic information. Practice was provided before each game, a rest was given between games, and the order of the games was counterbalanced across dyads. Audio recordings were made using a Toplux M-750 digital audio recorder and a SONY LAV 05 lapel microphone secured 10–15 cm from the speaker's mouth. Data collection took place in a quiet room at the boys' homes or schools.

Acoustic analysis was performed using Praat, Version 4.0.3 (Boersma & Weenink, 2000). For the pitch targets, mean F0 was measured at five equally spaced time points of the vocalic segment. For the duration targets, the beginning and end of waveforms were manually marked and measured. Interexaminer and intraexaminer reliability for the acoustic analyses (based on 10% of the samples) were 0.99 or higher (Pearson's r). Caregiver accuracy was determined for each target for each speaker. Error analyses were conducted using confusion matrixes.

RESULTS AND DISCUSSION

Caregiver accuracy scores for the pitch targets are shown in Figure 2. As can be seen, accuracy was

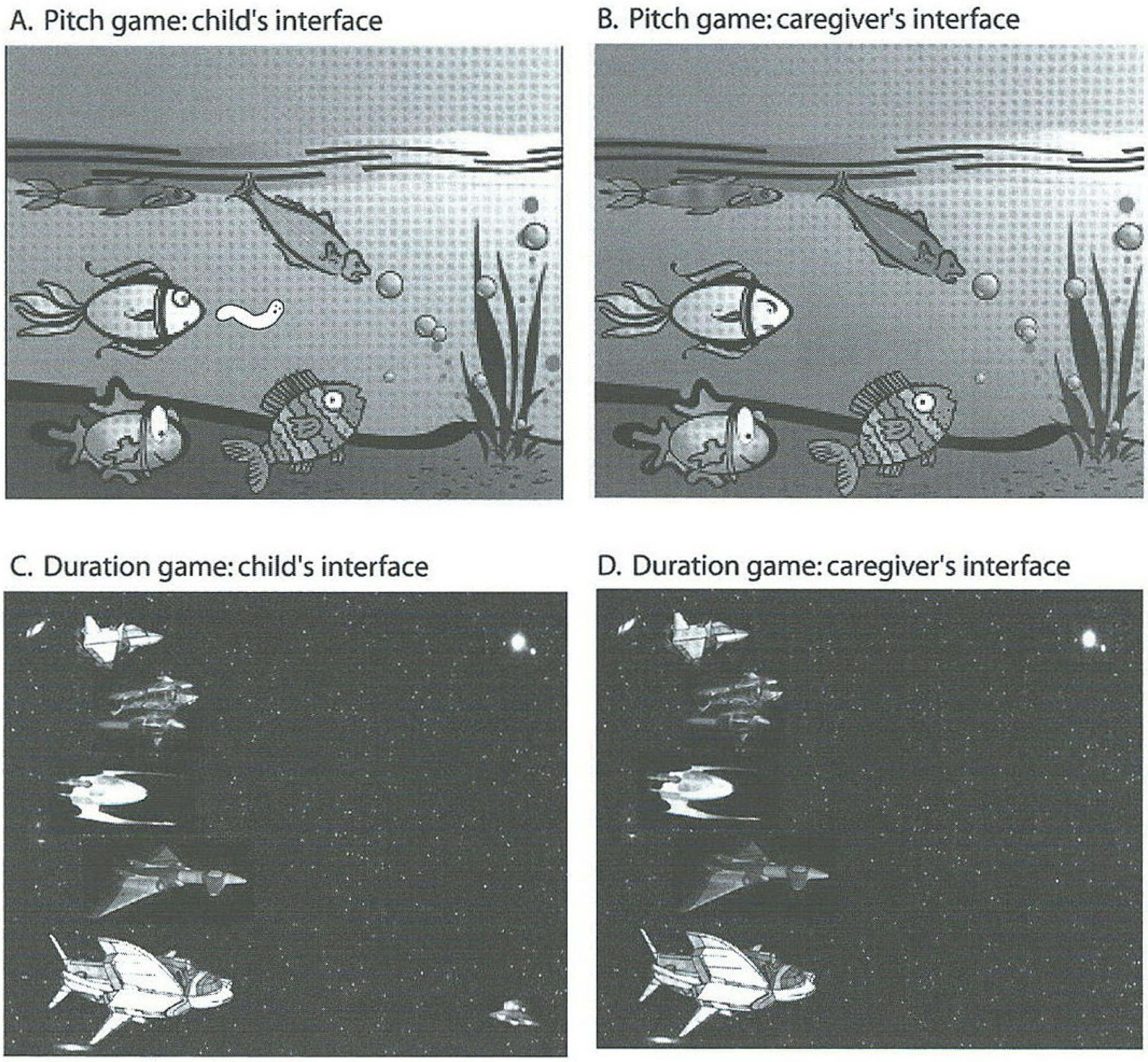

Figure 1. Screenshots of the pitch (A, B) and duration (C, D) games. In the pitch game, each fish is associated with a distinct pitch category. While the worm in child's interface (A) indicates which pitch target is being requested, the caregiver's interface (B) only includes the fish, as her task is to identify the child's intention. Similarly in the duration game, each spaceship is associated with the target duration lengths and the UFO in the child's interface (C) signals which length to produce.

above-chance level (20%) for three of the five pitch targets for three dyads (Speakers 1, 2, and 4), and for two of the pitch targets for the remaining dyad (Speaker 3). Midlevel pitch was identified most accurately except for Speaker 3, where falling pitch had the highest accuracy.

Figure 3 shows the mean F0 for each pitch level for each speaker. Several of the F0 patterns appeared consistent with the accuracy data. For example, Speaker 1 showed a distinctly lower starting point for the low pitch target; caregiver accuracy for this target was above chance level (40%). For Speaker 2, the high pitch target, which had 40% perceptual accuracy, had a higher F0 at four of the five time points. This speaker also had a distinctly lower F0 pattern, at the 25% and 50% time points, for the low pitch target. For Speaker 3, the falling pitch target had a distinct descending contour and caregiver accuracy was 60%. However, the acoustic data also show considerable overlap in the F0 patterns for many of the other targets across speakers, thus rendering accuracy data difficult to interpret. For example, although Speaker 1 had 100% accuracy for the midlevel pitch, the acoustic pattern for this target did not appear distinct from the other targets in terms of mean F0. Indeed, the midlevel pitch appeared to serve as the "default" for several of the caregivers; the higher accuracy for this target was not easily interpretable from the F0 data. It is possible that the caregivers were responding to cues other than F0, for this and other targets. Anecdotally, the third

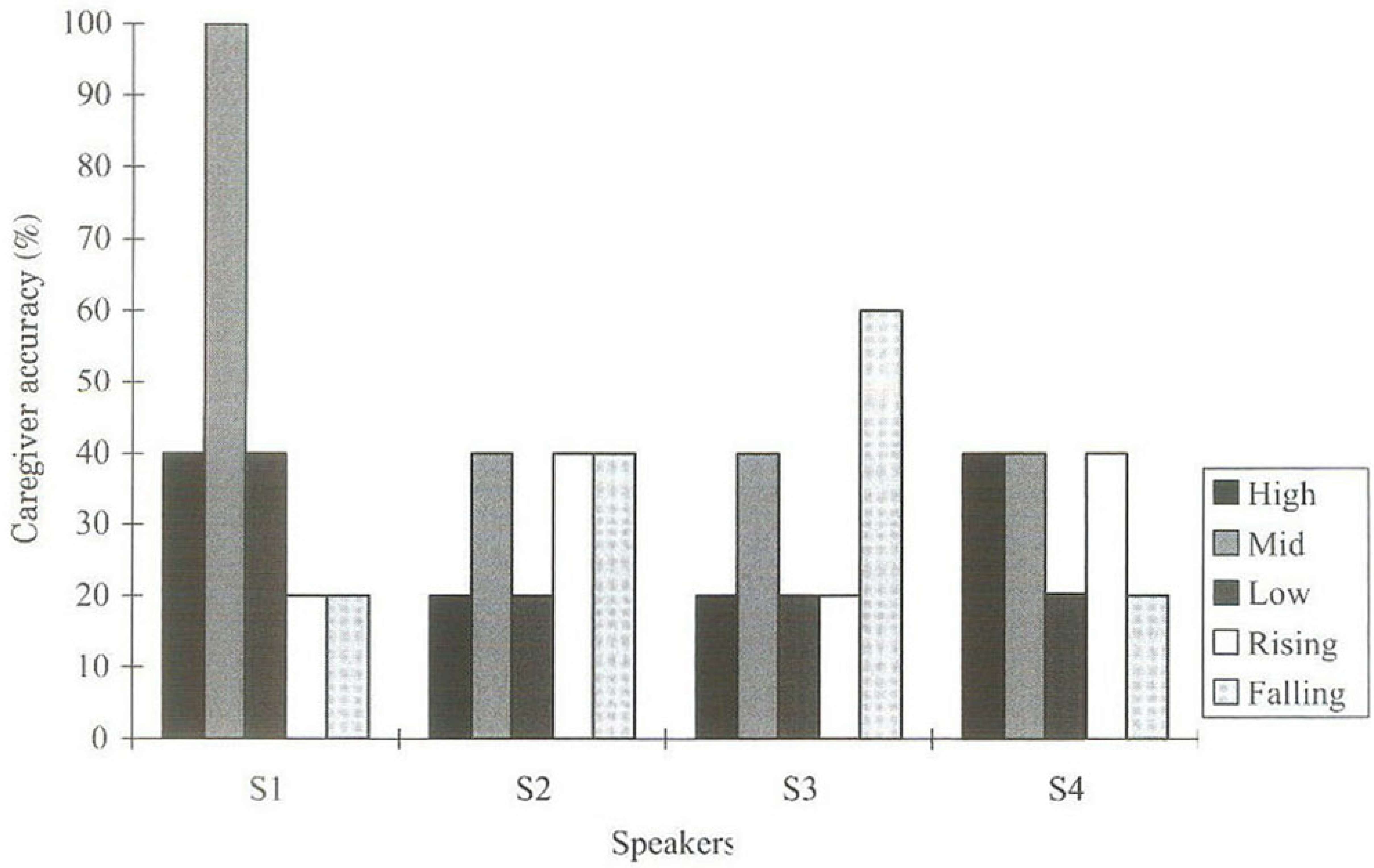

Figure 2. Caregiver accuracy for each pitch target.

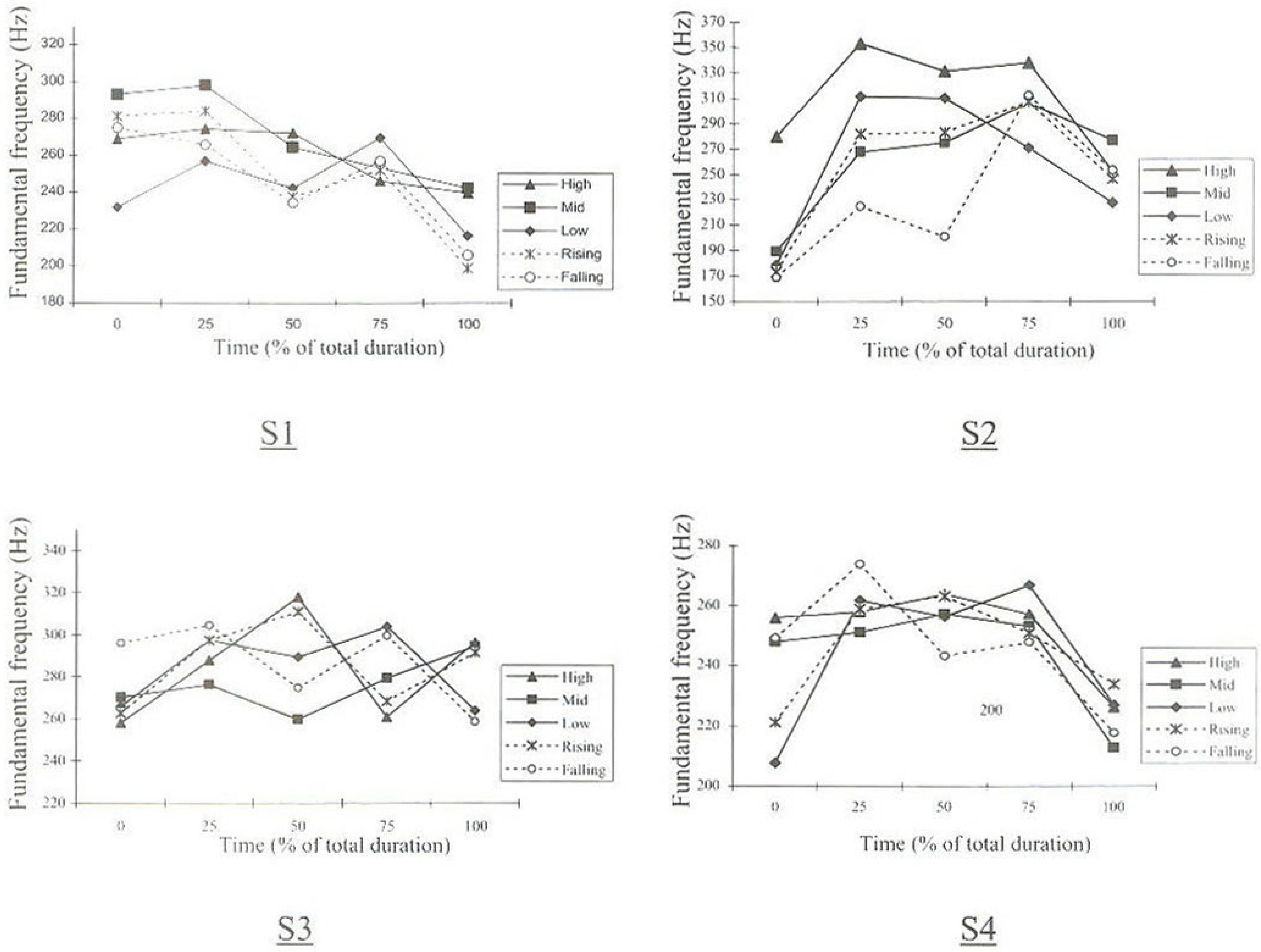

Figure 3. Mean F0 across five time intervals of the sustained /a/ production for each pitch targets by speaker.

author, who collected the data, noted that some of the boys appeared to be using gestures such as raising the upper body for a target high pitch (presumably subconsciously). Unfortunately, no video-recordings were available to confirm or further analyze this observation. It should be noted that the possibility of visual cues was not a factor in the study by Patel and Salata (2006), where the mothers were seated in a different room and had access only to acoustic information. It is possible that voice intensity or voice quality cues were also being exploited in the present study.

Khouw and Ciocca (2007) showed that F0 in the later part of the vocalic segment (62.5% and 75% time points) was crucial in identification of lexical tone, for nonimpaired Cantonese speakers and listeners. This is in contrast with the findings of the current study, where F0 information in the earlier part of the segment appeared to be more useful to listeners in identification. The difference in findings may be due to differences in the speech sample used (lexical tone in real words, versus pitch targets in sustained vowel productions). Ten adolescent nonimpaired speakers were employed in the Khouw and Ciocca (2007) study. The children with severe dysarthria involved in the current study may have used very different strategies in their attempts to achieve prosodic contrasts.

Caregiver accuracy for duration targets is shown in Figure 4. Accuracy was above chance for one of the five targets for Speaker 1, for four of the five targets for Speaker 2, for two targets for Speaker 3, and for all five targets for Speaker 4. The "shortest" duration was identified at above-chance level for all speakers except Speaker 1, and the "short" duration was identified at above-chance level for all speakers except Speaker 3. The relatively high accuracy for these two target durations may be related to the difficulties which some speakers with cerebral palsy have with respiratory support and control (Hardy, 1983; Yorkston, Beukelman, & Bell, 1988). "Longest" duration was identified at above-chance level for two of the four speakers (Speakers 2 and 4).

The mean duration and standard deviation of the five duration targets by speaker are shown in Figure 5. Again, some of the perceptual data appeared relatively easily explained by the acoustic measures of duration. For example, caregivers of Speakers 2 and 4 had high perceptual accuracy for the "longest" duration category, which was corroborated by the acoustic data. Similarly, Speaker 2's caregiver had high accuracy for the "shortest" category and the child produced distinctly short vocalizations for this target. Conversely, relatively low perceptual accuracy for Speaker 1 was consistent with the lack of clear acoustic distinctions between targets. However, again, some of the acoustic data appeared inconsistent with the perceptual results. For example, the high accuracy (60%) for "shortest" duration in Speaker 3 did not coincide with the mean duration measures of that target. The acoustic results for duration need to be interpreted in the context of what we know about detectable differences in duration. For example, Moore (2003) reported that the detectable durations for duration discrimination were 4, 15, and 60 ms for durations of 10, 100, and 1000 ms, respectively. Several of the apparent differences in duration in the current study would not have been sufficient to allow accurate discrimination by a listener.

As shown in Figure 5, the standard deviations for the duration measures were high. Although not shown in Figure 3, standard deviations for the F0

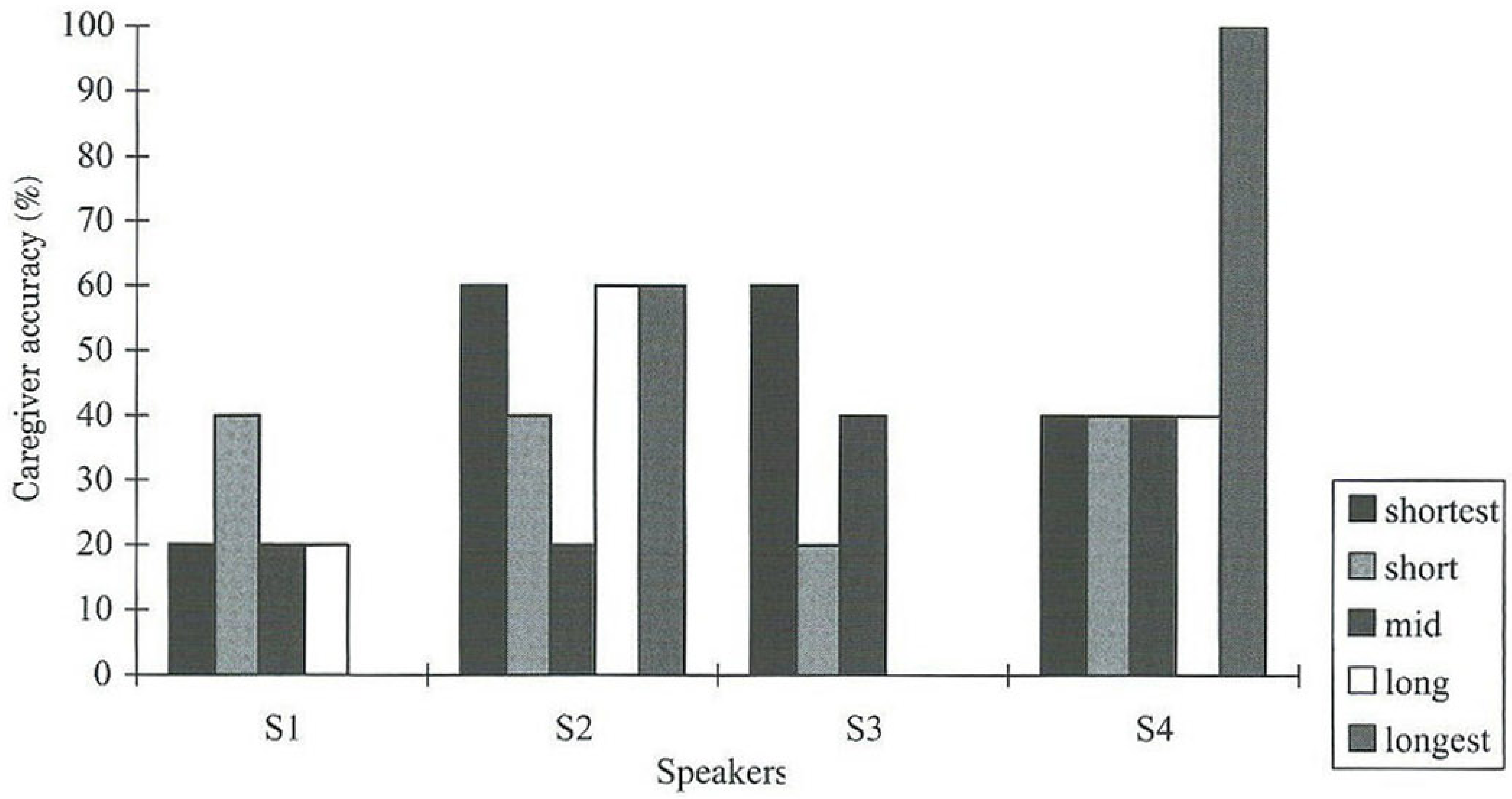

Figure 4. Caregiver accuracy for each duration target.

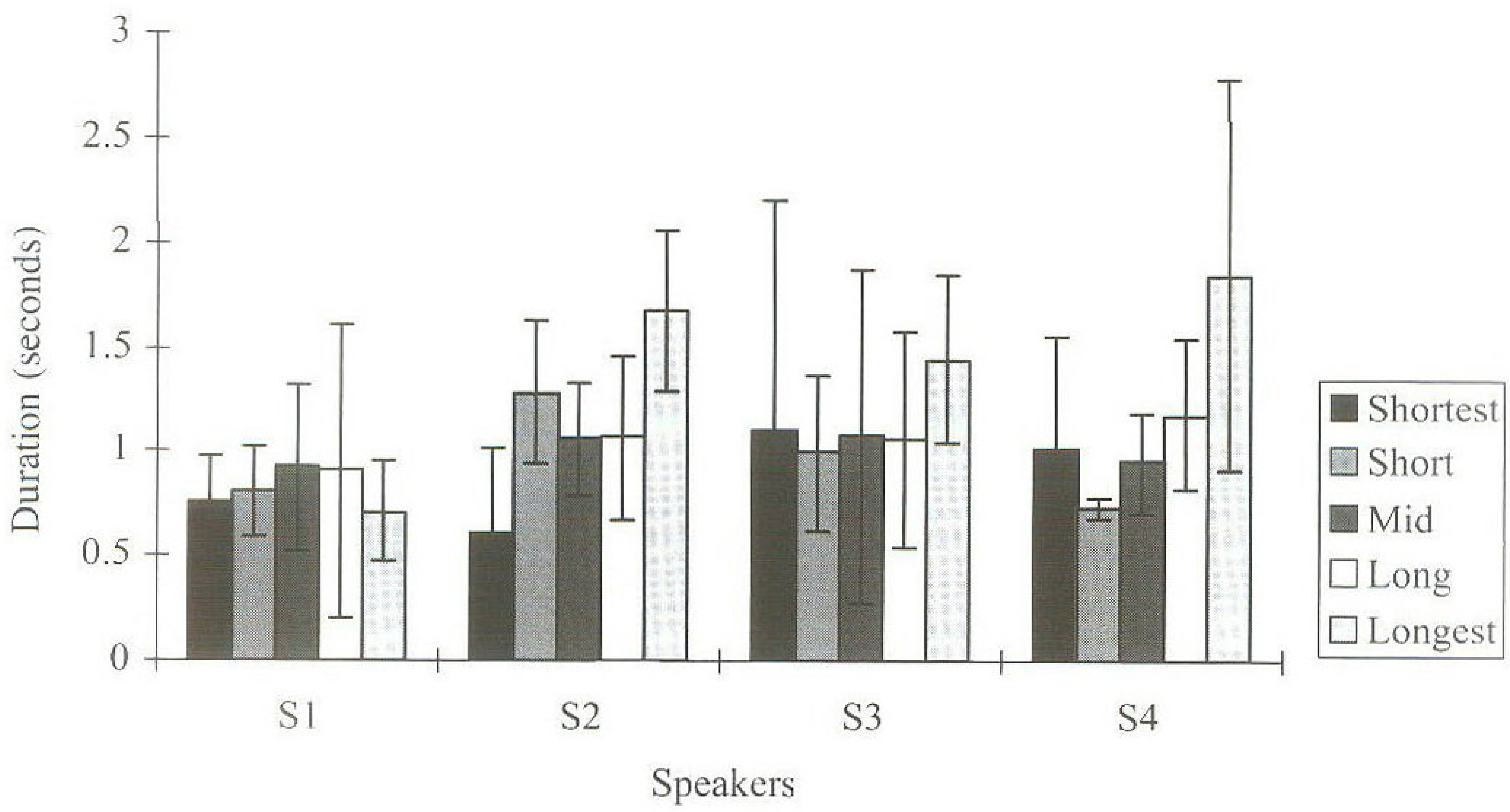

Figure 5. Mean duration and standard deviation of each duration target by participant.

measures were also high. The high degree of both intraparticipant variability and interparticipant variability in both production and listener accuracy was not unexpected. Speakers with dysarthria as a result of cerebral palsy, as well as other causes, are known to be heterogenous and to demonstrate variable productions (Patel & Salata, 2006; Weismer & Liss, 1991). Despite this variability, each of the severely dysarthric speakers in this study demonstrated the ability to use several prosodic targets contrastively, as demonstrated by their caregivers' accuracy in identifying a number of the pitch and duration targets for each speaker.

The caregivers in the current study appeared to demonstrate higher accuracy in identifying their children's target pitch productions than the caregivers in the study of Patel and Salata (2006), based on the number of categories identified at above-chance levels. The caregivers in the current study identified two to three categories at above-chance levels (from the five target pitch levels), whereas the caregivers in the earlier study identified one or two, from a target three categories, across five speaker dyads. As noted above, dysarthric speakers are not homogenous and there was no attempt to control for severity or other variables across the two studies. However, it is possible that Cantonese listeners are better than English listeners in identifying pitch targets, due to their experiences with a tonal language. Francis and Ciocca (2003) found that Cantonese and English listeners performed similarly in an experiment involving pitch perception. However, their study involved discrimination (of lexical tones) rather than identification (of pitch in sustained phonations). Further investigation of this issue would be of both theoretical and practical interest.

In summary, the results of this study showed that children with severe dysarthria have some ability to use prosodic cues (pitch and duration) contrastively, and in a way that can be identified by their communication partners. These results confirm and extend the findings of Patel and Salata (2006) to Cantonese speakers with dysarthria. The addition of rising and falling pitches in the current study provides a useful addition to the pitch targets already explored (Patel & Salata, 2006). Although not detailed in this article, combination targets (pitch plus duration) could also be employed. The dimension of intensity was not investigated in this study but would be another variable (alone or in combination with other targets) to employ.

The ability to control prosodic cues could be exploited to enhance communication between children with dysarthria and their communicative partners. Perhaps more excitingly, this ability could be developed as means for these children to interact with AAC devices (Patel & Roy, 1998). Researchers or clinicians could follow a protocol similar to that employed in this study to determine how many distinct contrasts an individual child could produce. These distinct categories could then be associated with specific messages or particular icon locations on an AAC device. Along these lines, Patel and colleagues are developing software to examine the impact of training/learning on refining prosodic abilities as children use voice as an input to control an AAC device.

Acknowledgments This study is based on an undergraduate dissertation completed by the third author, under the supervision of the first author, with the collaboration and assistance of the second author. We are grateful to Bill Kuo for his assistance in modifying the computer program. Thanks to Valter Ciocca and Joan Ma for their advice on the acoustic analysis. Finally, sincere thanks to the eight participants.

Address correspondence to Tara L. Whitehill, Ph.D., Division of Speech and Hearing Sciences, University of Hong Kong, Bonham Road, Hong Kong.
e-mail: tara@hku.hk

REFERENCES

Boersma, P., & Weenink, D. (2000). *Praat: A system for doing phonetics by computer* (version 4.0.30) [Computer software]. Retrieved February 1, 2007, from http://www.praat.org/

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America, 114*, 1611–1621.

Hardy, J. C. (1983). *Cerebral palsy*. Englewood Cliffs, NJ: Prentice-Hall.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35,* 104–117.

Le Dorze, G., Ouellet, L., & Ryalls, J. (1994). Intonation and speech rate in dysarthric speech. *Journal of Communication Disorders, 27*, 1–18.

Leuschel, A., & Docherty, G. J. (1996). Prosodic assessment of dysarthria: Effects of sampling task and issues of variability. In D. A. Robin, K. M. Yorkston, & D. R. Beukelman (Eds.), *Disorders of motor speech: Assessment, treatment and clinical characterization* (pp.155–180). Baltimore: Paul H. Brookes Publishing.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego: Academic Press.

Patel, R. (2002a). Phonatory control in adults with cerebral palsy and severe dysarthria. *Augmentative and Alternative Communication, 18*, 2–10.

Patel, R. (2002b). Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast. *Journal of Speech, Language, and Hearing Research, 45*, 858–870.

Patel, R. (2003). Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. *Journal of Speech, Language & Hearing Research, 46*(6), 1401–1415.

Patel. R. (2004). The acoustics of contrastive prosody in adults with cerebral palsy. *Journal of Medical Speech-Language Pathology, 12*, 189–193.

Patel, R., & Roy, D. (1998). Teachable interfaces for individuals with severe physical and speech disabilities. *American Association for Artificial Intelligence Workshop on Assistive Technology* (pp. 40–47), Madison, WI.

Patel, R., & Salata, A. (2006). Using computer games to mediate caregiver-child communication for children with severe dysarthria. *Journal of Medical Speech-Language Pathology, 14*, 279–284.

Patel, R., & Watkins, C. (2007). Stress identification in speakers with dysarthria due to cerebral palsy: An initial report. *Journal of Medical Speech-Language Pathology, 15*(2), 149–159.

Weismer, G., & Liss, J. (1991). Acoustic/perceptual taxonomies of speech production deficits in motor speech disorders. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 245–270). Baltimore: Paul H. Brookes.

Yorkston, K. M., Beukelman, D. R., & Bell, K. R. (1988). *Clinical management of dysarthric speakers.* London: Taylor & Francis.

Yorkston, K. M., Beukelman, D. R., Minifie, F., & Sapir, S. (1984). Assessment of stress patterning. In M. McNeil, J. Rosenbek, & A. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 131–162). Austin, TX: Pro-Ed.