The Effects of Simultaneous Communication on Production and Perception of Speech

Nicholas Schiavetti
State University of New York, Geneseo

Robert L. Whitehead
National Technical Institute for the Deaf

Dale Evan Metz
State University of New York, Geneseo

This article reviews experiments completed over the past decade at the National Technical Institute for the Deaf and the State University of New York at Geneseo concerning speech produced during simultaneous communication (SC) and synthesizes the empirical evidence concerning the acoustical and perceptual characteristics of speech in SC. Comparisons are drawn between SC and other modes of rate-altered speech that have been used successfully to enhance communication effectiveness. Of particular importance are conclusions regarding the appropriateness of speech produced during SC for communication between hearing and hearing-impaired speakers and listeners and the appropriateness of SC use by parents and teachers for speech development of children with hearing impairment. This program of systematic basic research adds value to the discussion about the use of SC by focusing on the specific implications of empirical results regarding speech production and perception.

The purpose of this article is to review empirical research on speech produced during simultaneous communication (SC) and to draw specific conclusions about the appropriateness of this speech pattern as a method for communicating in educational and social environments. As defined by Coryell and Holcolm (1997, p. 389):

Simultaneous Communication, or Sim-Com, is the practice of speaking and signing at the same time, following English word order. It is most often used in interactions between deaf individuals and those who are hearing. In combining the two modes (speaking and signing), English is the spoken language, and either PSE or one of the MCE systems (e.g., SEE II) is the signed component.

Empirical data on both production and perception of speech during SC may shed light on its effectiveness in fulfilling its two main purposes: serving as a means for communication between hearing and hearing-impaired people and enhancing speech and language development of hearing-impaired children. Although many issues should be considered in the evaluation of SC, including its role in education of the deaf, its influence on performance of manual tasks, and the linguistic integrity of messages conveyed, the discussion in this article is limited to the influence of SC on the production and perception of speech. The rationale for this limitation is twofold: We have completed a number of speech experiments that permit us to draw valid and reliable conclusions, and the answers to questions concerning other issues are beyond the purview of the research reviewed here and should be addressed by summaries of experimental research in other areas, such as linguistics, education, or motor learning. Reviews of research in some of these other areas of interest may be found in Akamatsu and Stewart (1989, 1998), Maxwell (1990), Moores (1996), and Strong and Charlson (1987).
The importance of studying SC is based on the accumulation over a number of years of well-documented evidence of its widespread use. For example, Woodward and Allen (1993) reported that over 80% of teachers in residential schools for deaf children and over 50% of nonresidential school teachers used speech and sign language at the same time in the classroom. Schildroth and Hotto’s (1993) survey revealed that the percentage of children taught through the use of combined sign and speech varied from year to year and with the type of educational setting (e.g., residential, day, or local schools) but averaged around 60% between 1984 and 1992. Moores’s (1996) review of Schildroth and Hotto’s data indicated that the percentages of children taught with combined speech and sign were 92% in residential schools, 38% in integrated local schools where auditory/oral instructional method predominated, and 72% in other school settings. In addition to these patterns of teachers’ use of SC, it is instructive to note that Tye-Murray, Spencer, and Woodworth (1995) demonstrated that children who used SC before cochlear implant surgery did not discontinue the use of sign language 2 years after receiving the implant. The children they studied used SC on an average of 83% of the words within their postsurgery spontaneous speech samples.

The use of SC is a controversial issue, and its pros and cons have remained the subject of continuing debate (Hyde, Power, & Leigh, 1998; Marmor & Petitto, 1979; Maxwell, 1990; Strong & Charlson, 1987). Vernon and Andrews (1990) outlined the advantages of SC in opening a variety of modes of communication for deaf children, providing more accurate representation of English than that provided by lipreading alone and contributing beneficial effects in psychosocial and linguistic development. The disadvantages of SC that they discussed included alterations in the linguistic integrity of both manual and oral forms of communication, abbreviations of English in the manual code, deletion of grammatical markers, and slowing of speech. They suggested that the increased cognitive demands of using SC and the inability to represent adequately and completely the structures of one mode in the other are two of the reasons for the difficulties of combining speech and manual communication but that the use of bimodal input in which the child is not forced to rely on one mode is widely accepted. Vernon and Andrews (1990) concluded that further research is needed to investigate the influence of SC on the acquisition of English. Although research on SC (Marmor & Petitto, 1979; Maxwell & Bernstein, 1985; Strong & Charlson, 1987) has provided much useful information regarding many of the advantages and disadvantages cited above, little research has been available until recently that documents the acoustical and perceptual quality of the developmental speech model presented to deaf and hearing-impaired children during SC.

Coryell and Holcolm (1997) have outlined a number of assumptions and areas of concern regarding SC. Foremost in their discussion, they stated (1997, p. 389):

A major assumption in the use of Sim-Com is that it is possible to represent spoken English manually (sign and speak at the same time) and that neither sign production nor the articulation of English deteriorates when combining the two modes. In this way, Sim-Com provides maximum exposure to oral skill development at no expense to content learning.

An important part of this assumption, according to Coryell and Holcolm (1997, p. 389), is that the "message remains intact" and, indeed, Maxwell and Bernstein (1985) have reported closely equivalent sign-speech performance during SC and stated (1985, p. 63) that "the results suggest a synergistic model for this bimodally coded form of English." On the other hand, Coryell and Holcolm (1997) have addressed a number of areas of concern about the use of SC and stated:

Questions abound as to Sim-Com’s effectiveness as a mode of communication or as a method of teaching English. Numerous studies have documented major compromises in the areas of clarity, ease, and expressibility, and include problems of cognitive overload, frequent omissions, a tendency to simplify the message, and distorted pace.

Research studies concerning the viability of the intact message assumption behind SC have examined in significant detail how the manual component of SC is negatively affected. For example, Marmor and Petitto
reported ungrammatical signed utterances in SC, and Strong and Charlson (1987) found deviations in signed utterances during SC that they classified into five different strategies employed by teachers to cope with the difficult task of attempting to communicate through two channels at once.

Although SC research has focused on a number of aspects of American Sign Language (ASL) and English language structure, it has not focused much attention on specific acoustic or physiologic parameters of speech production. Maxwell (1990) noted that the emphasis in SC research has historically been on English morphology and not on other levels of English structure, such as word sounds, intonation, and phrasing, and has pointed out that we need to examine these other variables as they relate to language acquisition. Empirical studies of segmental and suprasegmental features of speech produced during SC could provide useful information regarding possible effects of SC on the development of speech features by deaf and hearing-impaired children. Both theoretical and practical importance can be ascribed to knowledge of the speech process during SC. Mere slowing of speech with no violation of linguistic temporal rules would indicate that speakers attempt to approach simultaneity by approximating speech rate to the rate of the slower manual task, as suggested by Windsor and Fristoe (1991). Violations of linguistic temporal rules, however, would indicate that combining a manual task with a speech task could jeopardize speech intelligibility and provide an inappropriate model for the development of hearing-impaired children.

Several previous studies in the literature provide a rationale for the starting point of this speech research. Our investigations of the potential effects of SC on speech were predicated on prior reports of slower speech rate during SC, although there were scant data in this literature on the specific nature of segmental and suprasegmental temporal changes in SC speech. For example, Huntington and Watton (1984) described the reduced rate of speech of three teachers using speech and signed English in a Total Communication classroom; commented anecdotally on disruptions in the normal rhythm of speech, modifications to the vowel formant frequencies, and absent consonant-to-vowel formant transitions; and concluded that the hearing-impaired children were not exposed to the typical prosodic and segmental features of speech that are necessary for developing oral communication skills.

Data were lacking in the literature, however, regarding the degree to which speech produced during SC conformed to the known temporal and spectral characteristics of English consonants and vowels. For example, it was not known to what extent consonant-vowel temporal relationships, voice onset time (VOT) patterns, anticipatory fricative durations, consonant spectral moments, second formant transitions, or the influence of consonant environment on vowel durations were maintained in speech produced during SC. An especially compelling reason to study temporal and spectral characteristics of speech during SC is the potential for rate changes to affect these characteristics of the speech signal in a manner that could influence the perception of specific linguistic contrasts.

Although researchers such as Huntington and Watton (1984) have raised questions regarding the quality of the speech model presented to deaf children during SC, it has been pointed out that temporal characteristics of speech produced by parents talking to their children (so called parentese behavior [cf., Owens, 2001]) and of speech made deliberately clear for the hearing impaired (Picheny, Durlach, & Braida, 1986) show similar durational changes to those found in SC. Research has also shown that some acoustic contrasts are actually enhanced when speech rate is slowed in synthetically rate-altered speech (Volaitis & Miller, 1992), in parentese speech that parents use to talk to their children (Swanson, Leonard, & Gandour, 1992), and in the experimental condition wherein Picheny et al. (1986) instructed talkers to deliberately alter their behavior to “speak clearly for hearing-impaired listeners.”

Such acoustic contrast enhancements could be advantageous for speech development of hearing-impaired children if they made important phonetic contrasts more salient to children, as appears to occur in parentese and speaking clearly. Hyde et al. (1998) have raised the possibility that there may be a trading relationship between perceived naturalness and understandability of SC that indicates that research on the acoustical and perceptual characteristics of speech

(1979) reported ungrammatical signed utterances in SC, and Strong and Charlson (1987) found deviations in signed utterances during SC that they classified into five different strategies employed by teachers to cope with the difficult task of attempting to communicate through two channels at once.

Although SC research has focused on a number of aspects of American Sign Language (ASL) and English language structure, it has not focused much attention on specific acoustic or physiologic parameters of speech production. Maxwell (1990) noted that the emphasis in SC research has historically been on English morphology and not on other levels of English structure, such as word sounds, intonation, and phrasing, and has pointed out that we need to examine these other variables as they relate to language acquisition. Empirical studies of segmental and suprasegmental features of speech produced during SC could provide useful information regarding possible effects of SC on the development of speech features by deaf and hearing-impaired children. Both theoretical and practical importance can be ascribed to knowledge of the speech process during SC. Mere slowing of speech with no violation of linguistic temporal rules would indicate that speakers attempt to approach simultaneity by approximating speech rate to the rate of the slower manual task, as suggested by Windsor and Fristoe (1991). Violations of linguistic temporal rules, however, would indicate that combining a manual task with a speech task could jeopardize speech intelligibility and provide an inappropriate model for the development of hearing-impaired children.

Several previous studies in the literature provide a rationale for the starting point of this speech research. Our investigations of the potential effects of SC on speech were predicated on prior reports of slower speech rate during SC, although there were scant data in this literature on the specific nature of segmental and suprasegmental temporal changes in SC speech. For example, Huntington and Watton (1984) described the reduced rate of speech of three teachers using speech and signed English in a Total Communication classroom; commented anecdotally on disruptions in the normal rhythm of speech, modifications to the vowel formant frequencies, and absent consonant-to-vowel formant transitions; and concluded that the hearing-impaired children were not exposed to the typical prosodic and segmental features of speech that are necessary for developing oral communication skills.

Data were lacking in the literature, however, regarding the degree to which speech produced during SC conformed to the known temporal and spectral characteristics of English consonants and vowels. For example, it was not known to what extent consonant-vowel temporal relationships, voice onset time (VOT) patterns, anticipatory fricative durations, consonant spectral moments, second formant transitions, or the influence of consonant environment on vowel durations were maintained in speech produced during SC. An especially compelling reason to study temporal and spectral characteristics of speech during SC is the potential for rate changes to affect these characteristics of the speech signal in a manner that could influence the perception of specific linguistic contrasts.

Although researchers such as Huntington and Watton (1984) have raised questions regarding the quality of the speech model presented to deaf children during SC, it has been pointed out that temporal characteristics of speech produced by parents talking to their children (so called parentese behavior [cf., Owens, 2001]) and of speech made deliberately clear for the hearing impaired (Picheny, Durlach, & Braida, 1986) show similar durational changes to those found in SC. Research has also shown that some acoustic contrasts are actually enhanced when speech rate is slowed in synthetically rate-altered speech (Volaitis & Miller, 1992), in parentese speech that parents use to talk to their children (Swanson, Leonard, & Gandour, 1992), and in the experimental condition wherein Picheny et al. (1986) instructed talkers to deliberately alter their behavior to “speak clearly for hearing-impaired listeners.”

Such acoustic contrast enhancements could be advantageous for speech development of hearing-impaired children if they made important phonetic contrasts more salient to children, as appears to occur in parentese and speaking clearly. Hyde et al. (1998) have raised the possibility that there may be a trading relationship between perceived naturalness and understandability of SC that indicates that research on the acoustical and perceptual characteristics of speech
during SC is important to obtain a better understanding of its efficacy as a speech model. This possible trading relationship highlights another area of important research—the measurement of speech quality during SC. Although a number of anecdotal reports of degraded speech quality during SC have been made, only recently has speech science research addressed this issue (Windsor & Fristoe, 1991), and it is important to apply the systematic research methodologies advocated by Onslow and Ingham (1987) to the study of speech quality during SC.

Method
Research Questions
The two general research questions addressed in the experiments were:

1. Are there temporal or spectral acoustical differences in speech produced during SC compared with speech produced alone?

2. Are there perceptual differences in the intelligibility or quality of speech produced during SC compared with speech produced alone?

The outline of the research review will follow the general research questions, covering acoustic speech production characteristics first, followed by information on perception of speech produced during SC.

Design
The experiments were designed to balance the relationship of internal and external validity with emphasis on the sine qua non of internal validity in our initial studies with subsequent research undertaken to extend external validity (Campbell & Stanley, 1966). We concentrated first on internal validity in order to compare our results with classic speech and hearing science laboratory data (e.g., House, 1961; House & Fairbanks, 1953; Lisker & Abramson, 1964; Peterson & Barney, 1952) to answer the above research questions on firm empirical grounds, then to examine the generality of some of our results through direct and systematic replication. Further research is still needed to extend generalizability beyond the laboratory to everyday situations to assess the degree to which constrained experimental results generalize to natural interactions.

Readers are referred to the original studies in the reference list for specific methodological details concerning the experienced and inexperienced SC users who served as speakers, the listeners in the perception studies, the equipment and procedures employed in each experiment, and the target words that were used to elicit specific speech production characteristics for acoustic and perceptual analysis. Recordings were made in a sound-attenuating room with high-quality audio equipment, and acoustic measures were made by digitizing the audio recordings and analyzing the resultant speech wave forms and spectra with a laboratory computer. Carrier sentences containing the target words were spoken in random orders, with the order of experimental conditions counterbalanced across speakers in each experiment. Perceptual analysis was accomplished by means of various speech intelligibility measures, word identification tests, and speech quality scaling procedures in which listeners audited the speech samples in a sound-treated audiometric test booth. Appropriate measures of reliability were applied to each dependent variable (e.g., test-retest, internal consistency, inter- and intraobserver agreement, etc.), and results indicated sufficient and in many cases substantial reliability.

Speech Production During SC
This section will review research concerning the acoustic characteristics of speech produced during SC. Specific topics examined include: (a) temporal characteristics, (b) spectral characteristics, (c) supra-segmental characteristics, and (d) replication of these results with inexperienced signers.

Temporal Characteristics of Speech

Speech Segment and Pause Durations. To quantify the temporal characteristics of speech produced during SC, the first experiment (Whitehead, Schiavetti, Whitehead, & Metz, 1995) was designed to examine speech timing characteristics for sentences produced under three separate conditions: speech alone (SA); speech combined with sign language; and speech combined with sign language and fingerspelling by
experienced signers. The specific temporal characteristics were (a) sentence duration; (b) consonant-vowel-consonant (CVC) word duration; (c) vowel duration in CVC words; (d) pause duration before and after CVC words; and (e) consonantal effects on vowel duration. The speech samples consisted of the carrier sentence, “I can say ____ again” with eight CVC words containing the vowels /a/ or /I/ and a final plosive consonant that was systematically varied to examine the effect of voicing and place of articulation on preceding vowel duration. To study fingerspelling combined with speech, the phonemes were orthographically and phonemically consistent (e.g., the vowel /a/ was always fingerspelled with the letter O and the vowel /I/ with the letter I). Speakers uttered each sentence under three conditions: (a) SA, (b) speech combined with sign language for every word in the phrase, and (c) speech combined with sign language for all words in the sentence except the experimental CVC word, which was fingerspelled.

Statistical analysis of the dependent variables revealed that the speech/sign/fingerspelling condition was longest in sentence, word, vowel, and pause duration, followed by the speech/sign condition, then SA. Of most importance, the vowel /a/ was always longer than the vowel /I/, and vowel duration was always longer before final voiced consonants than before final voiceless consonants, two phonetic patterns that would be predicted from normative vowel data (House, 1961; House & Fairbanks, 1953). These results indicate that although SC speech was elongated in duration, the temporal rules for vowel production matched the characteristics seen in classic speech science literature. The finding of longer sentence duration in SC as an indication of speech rate change was replicated in all of the subsequent studies, demonstrating the robust nature of the effect of SC on the temporal macrostructure of speech. Temporal microstructure was further examined by considering the specific temporal characteristics delineated below.

**Voice Onset Time.** To study the degree to which slower speech during SC conformed to normative data regarding temporal voicing patterns of English and the degree to which any SC effects would be generalized across place of articulation and vowel context, a study (Schiavetti, Whitehead, Metz, Whitehead, & Mignerey, 1996) was designed to examine the effects of SC on the VOT of the voiced and voiceless labial, alveolar, and velar plosives /p/, /t/, /k/, /b/, /d/, and /g/ preceding the high and low vowels /I/ and /ae/. VOT is the time interval from the release of a plosive consonant burst to the initiation of vocal cord vibration at the beginning of the following vowel in the syllable and is usually measured from the acoustic record of the sound spectrogram.

Sentence duration and VOT were significantly longer in the SC condition, and VOT was longer for voiceless than for voiced plosives, as would be expected from normative data (Klatt, 1975; Lisker & Abramson, 1964). A two-way interaction effect was also found in which SA and SC differed in VOT for voiceless but not voiced plosives, a finding that indicates increased VOT contrast between voiced and voiceless plosives, which could facilitate their perceptual identification. These results are consistent with the findings of Volaitis and Miller (1992) and of Picheny et al. (1986) that voiceless plosives increased in VOT more than voiced plosives during other rate-altered conditions. In addition, VOT varied systematically with both voicing and place of articulation, and these results were generalized across the vowels, as demonstrated in the classic studies of the temporal rules of English (Klatt, 1975; Lisker & Abramson, 1964; Volaitis & Miller, 1992). VOT was longer for voiceless than voiced plosives and increased from front to back places of articulation, in good agreement with Klatt’s explanation (1975) that the place of articulation effect was most likely based on physiological constraints of varying duration of glottal closure and plosive burst. These SC VOT findings show consistent maintenance of the temporal voicing rules of English, despite temporal elongations during SC.

**Final Consonant Voicing.** Vowel duration plays a pivotal role in perception of postvocalic consonants: One salient cue to final consonant voicing is carried by the duration of the preceding vowel. Vowel duration preceding voiced consonants is typically 50–100 milliseconds (msec) longer than vowel duration preceding voiceless consonants (House, 1961), and the ratio of
vowel durations preceding voiceless versus voiced consonants is typically in the neighborhood of 0.60 to 0.70. In addition, final consonant voicing can often be perceived by listeners after digital removal of the consonant from a syllable if appropriate vowel duration occurs in the speech sample (Raphael, 1972). Metz, Schiavetti, Lessler, Lawe, Whitehead, and Whitehead (1997) examined this phenomenon in speech produced during SC. Experienced signers produced speech samples containing six words with voiceless and six with voiced final consonants embedded in a carrier sentence, and durations of the vowels preceding the target words were analyzed acoustically. Results showed that duration of vowels preceding voiced consonants was about 50–65 msec longer than vowel duration preceding voiceless consonants in both SA and SC conditions, with equivalent voiced to voiceless vowel duration ratios of 0.72 in SA and 0.75 in SC.

**Anticipatory Fricative Duration.** Whitehead, Schiavetti, Metz, and Farinella (1999) examined the effect of vowel environment on the duration of preceding fricative consonants to determine whether the normal anticipatory scanning effect would occur in SC. A co-articulatory effect in contextual speech normally occurs in which the speaker scans forward from fricative consonants in anticipation of the next vowel and compensates for speech gesture movement time with greater fricative duration preceding high and front vowels than for low and back vowels (Schwartz, 1969). The durations of alveolar and palatal fricatives preceding four different vowels were examined to measure the degree of anticipatory scanning effect in SA versus SC. The speech samples consisted of eight experimental words that varied in the initial fricative (alveolar /s/ and palatal /ʃ/) and the following vowel (high front /i/, low front /ae/, low back /a/, and high back /u/) embedded in the blank slot in the carrier sentence. Speakers produced each sentence under two conditions: SA and speech combined with sign language for all words in the sentence except the experimental word, which was fingerspelled. Fricative consonant duration was measured by visually isolating the initiation of the aperiodic wave form following the diphthong /eɪ/ in the word *say* and marking the location with a cursor. The cursor was then moved to the end of the aperiodic wave form, and the temporal interval between the two cursors was taken as the value for fricative consonant duration. Results showed significant effects of communication mode on fricative duration, with longer fricative durations in the SC than in the SA condition. This result would be predicted from previous research showing longer phoneme segments in SC (Whitehead et al., 1995). Anticipatory scanning was the same in both conditions, however, because longer fricative durations preceding high and front vowels than low and back vowels were found in both the SA and SC conditions. This result would be predicted from normative data of Schwartz (1969) if SC speech followed the linguistic temporal rules of spoken English. Thus, SC speech presented both normal vowel anticipation in fricative co-articulation and more salient fricative production during SC because of the increased fricative duration.

**Manual Task Effect on Speech Timing.** Two studies examined specific effects of the manual task on speech during SC, one on length of fingerspelling task and one on the complexity of the signs used. Whitehead, Schiavetti, Whitehead, and Metz (1997) studied the effect of sign task variations on selected temporal features of speech by examining the effects of three independent variables: (a) mode (SA vs. SC), (b) sign task demand (base vs. elaborated signs), and (c) type of sign movement (kinetic vs. morphokinetic) on five dependent variables (a) word duration, (b) sentence duration, (c) diphthong duration, (d) inter-word interval before the signed word (IWIB), and (e) interword interval after the signed word (IWIA). The speech samples investigated consisted of a carrier sentence presenting 20 base words and 20 elaborations of these base words. The base words were represented by simple signs with minimal movement. The elaboration words were represented by signs that were extensions of the base sign, accomplished by movement with maintenance of the handshape of the base sign (kinetic elaboration) or movement accompanied by a change in handshape and/or number of hands (morphokinetic elaboration). Because sign task experimental conditions required different length words,
absolute sentence and word durations could not be used as the dependent variables. Therefore, ratios were calculated between base and elaboration conditions to index relative change in word and sentence durations.

The results showed increased duration ratios of sentences containing words with elaboration signs versus those containing words with base signs for SC but not for SA. The absolute values of the anticipatory durations of the diphthong and interword interval preceding the experimental words showed similar temporal elongations during SC, and this effect was greater for morphokinetic than for kinetic signs. Interword interval after the experimental words showed only the previously reported effect of SC versus SA with no influence of the sign task demand or type of movement. The greater effect of increased sign task demand and movement before the sign probably indicates the influence of anticipation of the more difficult task the speaker is about to encounter.

Schiavetti, Whitehead, Whitehead, and Metz (1998) examined the effect of fingerspelling difficulty (indexed as four graduated levels of fingerspelling task length from low to high) on six dependent variables: (a) sentence duration, (b) experimental word duration, (c) diphthong duration in the word before the experimental word, (d) IWIB, and (e) IWIA.

The speech samples consisted of 16 experimental words embedded in the blank slot of a carrier sentence. Four groups of four experimental words were composed to represent samples of words increasing in fingerspelling task length, operationally defined as an increase in the number of letters orthographically displayed per word by beginning with an initial base word and adding three longer suffixes to increase the word length. Thus, level 1 of task length consisted of base forms of the four experimental words and levels 2, 3, and 4 of task length consisted of the base forms with suffixes of increasing length added. The stimulus material consisted of the four base words care, trust, truth, and talk and their expansions into words of increasing length by suffix additions (e.g., care, careless, carelessly, carelessness). The results showed that sentence and word duration increased systematically with increased fingerspelling task length. The steeper slope of the function for SC than for SA revealed a significant interaction effect between experimental condition and fingerspelling task length, in which the influence of orthographic word length increased segmental duration more rapidly in SC than in SA. Diphthong duration, IWIB, and IWIA were longer in SC and increased systematically with fingerspelling task length in SC but did not vary with increased word length in the SA condition.

Both of these studies, then, indicated that increasing the demands of the manual task caused increased speech segment and pause durations as speech rate slowed even more when the manual task became longer or more complex. The magnitude of the effect was even larger for increased fingerspelling tasks than for increased sign task demands.

Spectral Characteristics of Speech

Fricative and Stop Spectral Moments. Spectral moment analysis is a procedure where fast Fourier transforms (FFTs) are computed for a speech sound spectrum, each FFT is treated as a random probability distribution, and the first four moments of the distribution (mean, variance, skewness, and kurtosis) are computed (Nittrouer, 1995). Spectral moments characterize the gross shape of the spectrum: Mean (first moment) reflects central concentration, variance (second moment) reflects the dispersion, skewness (third moment) reflects symmetry, and kurtosis (fourth moment) reflects the peak shape of the spectrum. The first, third, and fourth spectral moments can be used to accurately classify word initial voiceless obstruents and distinguish fricative place of articulation (Nittrouer, 1995), whereas the second moment is usually feckless in differentiating consonants.

Kardach, Wincowski, Metz, Schiavetti, Whitehead, and Hillenbrand (2002) investigated whether changes in speaking rate during SC would have an effect on the spectral moments of selected obstruct consonants (front and back fricatives and stops). The speech materials were based on the stimuli used by Nittrouer (1995): Twelve consonant-vowel (CV) syllables were constructed consisting of the front and back fricatives and stops. The speech materials were based on the stimuli used by Nittrouer (1995): Twelve consonant-vowel (CV) syllables were constructed consisting of the front and back, fricative and stop consonants /s/, /ʃ/, /tʃ/, and /k/ followed by the vowels /i/, /a/, and /u/ embedded in the carrier sentence. Results indicated no significant
effect of communication mode (SA vs. SC) on first, third, or fourth moment of any of the consonants in any of the vowel contexts. Significant effects were found for manner of articulation (fricative vs. stop) on first, third, and fourth moment and for place of articulation on first and third moments. The spectral moment values were generally consistent in both magnitude and direction with the adult data reported by Nittrouer (1995). In general, fricatives had higher first and fourth moments and lower third moments than stops; front consonants had higher first moments and lower third moments than back consonants; and first, third, and fourth moments for each individual consonant were virtually identical in the SA and SC conditions. Despite their slowing of speech during SC, the speakers maintained the spectral integrity of the obstruent consonants, as reflected by the comparability of the spectral moments between the SA and SC experimental conditions, and the authors concluded that SC does not disrupt the obstruent spectral patterns of spoken English.

Static and Dynamic Measures of Vowel Spectrum. Schiavetti, Metz, Whitehead, Brown, Borges, Rivera, and Schultz (2004) examined static and dynamic acoustic characteristics of vowels in an experiment designed to compare SC data with previously published data on vowel production from both a static and a dynamic spectral perspective (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). Maintenance of accurate vowel spectral information in SC would have strong implications for its appropriateness as a speech model for deaf and hearing-impaired persons. As Hillenbrand et al. (1995, p. 3099) stated: “There is now a solid body of evidence indicating that dynamic properties such as duration and spectral change play an important role in vowel perception.” The speech materials were from the same word list that Hillenbrand et al. (1995) constructed for their vowel study and was designed to elicit utterances that would permit analysis of a wide variety of English vowels, including those that had been studied by Peterson and Barney (1952). Ten vowels (/i, ï, e, æ, a, ə, o, u, ʌ, ɜ/) were recorded in a /hVd/ context embedded in the carrier sentence “I can say ____ again” in both SA and SC modes of communication.

Vowel duration, fundamental frequency, and formant frequency measurements made at the 20%, 50%, and 80% points of the duration of the vowel in the manner reported by Hillenbrand et al. (1995) were analyzed. Static midpoint vowel frequencies were in general agreement with the data of Peterson and Barney (1952) and of Hillenbrand et al. (1995) for adult men and women, and there were no significant differences in fundamental, first formant, and second formant frequency between the SA and the SC condition for any of the 10 vowels.

In addition to the traditional static analysis, Schiavetti et al. (2004) presented a dynamic perspective of the vowel formants by plotting spectral change from the 20% point to the 80% point of vowel duration for the first and second formants, using a method similar to that employed by Hillenbrand et al. (1995). The formant excursions between these two points were plotted on a modified Peterson-Barney acoustical vowel diagram to indicate how the first and second vowel formants varied as the speakers moved from the beginning toward the end of each vowel. Results revealed dynamic spectral patterns for both the SA and SC conditions that were similar to the patterns shown in the analysis of Hillenbrand et al. (1995). In most cases, first and second formants of the vowels moved away from each other in the acoustical vowel diagrams, a pattern that should result in perceptual salience of vowels that are normally close together in F1 and F2 at the nucleus. Despite the temporal slowing of speech during SC, the speakers maintained the same pattern of static and dynamic vowel acoustics in SC as they did in SA and consistently demonstrated vowel formant relationships that characterized place and height of vowels in both the SA and SC conditions in much the same manner as has been reported in previous research (Hillenbrand et al., 1995; Peterson & Barney, 1952).

Consonant-Vowel Formant Transitions. The second formant frequency (F2) transition from consonant to vowel is an important perceptual cue to stop consonant place of articulation in spoken English. Sussman, McCaffrey, and Matthews (1991) used locus equations, or straight line regression fits, of the onset of the F2 transition plotted against F2 at the vowel midpoint as a metric for the relational invariance of
place of articulation and concluded that locus equation slopes and y-intercepts were effective metric descriptors for place of articulation. If changes in speech rate disrupt the transition of the second formant frequency during SC, the intelligibility of speech could suffer due to ambiguous place of articulation cues.

Baillargeon, McLeod, Metz, Schiavetti, and Whitehead (2002) used locus equations to examine second formant frequency changes for initial voiced stop consonants in speech produced by experienced sign language users using SA versus SC with the same speech materials that Sussman et al. (1991) used to elicit the voiced consonants /b/, /d/, and /g/ paired with 10 different vowels. The F2 onsets plotted as a function of F2 offsets for each consonant produced in SA and SC were fit with linear regression functions, and their slopes and intercepts were determined. The locus equations for both SA and SC were consistent with the functions reported by Sussman et al. (1991) for normal speakers, and the slopes and intercepts were virtually identical for each stop consonant in the SA and SC conditions. Despite the temporal slowing of speech during SC, speakers maintained the same relationship between second formant onset and offset frequencies in SC as they did in SA and consistently demonstrated locus equations that characterized place of articulation in the same manner as reported by (Sussman et al., 1991).

**Suprasegmental Speech Characteristics**

*Sentence Intonation and Syllable Stress.* Whitehead, Schiavetti, Metz, Gallant, and Whitehead (2000) compared speech performance in SA and SC conditions with the classic data in the literature on fundamental frequency contours (Lehiste, 1976) and syllables stress (Fry, 1958), prosodic characteristics of spoken English that are regarded as important parameters in the quality of the adult speech model presented to deaf children (Maxwell, 1990). Speech materials adapted from a study of suprasegmental characteristics of the speech of deaf persons (Metz, Schiavetti, Samar, & Sitler, 1990) were designed to elicit utterances that would permit acoustic analysis of three prosodic variables: (a) fundamental frequency contour differences between declarative and interrogative sentences, (b) vowel duration differences between stressed and unstressed monosyllables, and (c) vowel fundamental frequency differences between stressed and unstressed monosyllables.

Vowel duration and fundamental frequency differences between stressed and unstressed syllables as well as intonation contour differences between declarative and interrogative sentences were essentially the same in both SA and SC conditions. Sentence contours demonstrated the expected contrastive frequency changes between declarative and interrogative sentences (Lehiste, 1976) for both SA and SC conditions: Fundamental frequency increased at the end of interrogative sentences and decreased at the end of declarative sentences. Syllable stress pattern demonstrated the expected contrastive duration and frequency changes from unstressed to stressed syllables (Fry, 1958) for both SA and SC conditions: Vowel duration was longer and fundamental frequency was higher in stressed than in unstressed syllables. These results demonstrate that slower speech during SC does not violate the suprasegmental rules used by speakers to signal prosodic changes in syllable stress or sentence type.

**Systematic Replications with Inexperienced Signers.** Three systematic replications were undertaken with inexperienced signers to determine the extent to which the results with the experienced signers would be obtained with persons more representative of new parents or teachers who recently learned to sign. Undergraduate students in an introductory sign language class participated twice in the making of experimental SA and SC recordings: during the first week of class and 15 weeks later, at the end of the semester.

The importance of generalizing results beyond experienced signers is stressed in a number of recent research reports and is important for at least three reasons. First, inexperienced signers’ speech disruption during SC may be different from that seen with experienced signers. Second, many hearing-impaired children use SC with inexperienced signers, such as their parents, siblings, peers, or school teachers. For example, Woodward and Allen (1993) reported that 21% of residential school teachers of the deaf and 52% of nonresidential school teachers of the deaf rated their
sign language skills as “greatly inferior” to their English skills. Third, because Lodge-Miller and Elfenbein (1994) have shown that inexperienced signers tend to overestimate their signing ability, it is possible that signing may put more stress than they realize on their speech.

Speech Segment and Pause Durations. Whitehead, Schiavetti, Metz, and Farrell (1999) replicated the study of Whitehead et al. (1995) to quantify the temporal characteristics of speech produced during SC. Speech was elongated in duration when it was combined with signed English and fingerspelling in SC by inexperienced signers in much the same fashion as it was for experienced signers, but the magnitude of the effect was much greater for inexperienced signers. The magnitude of the sentence, word, and vowel duration effects for inexperienced signers was approximately 1.5–2.0 times greater than the increases for the experienced signers in the Whitehead et al. (1995) study. The interword interval duration increases for inexperienced signers in the first week of class were approximately 2–3 times the increase for experienced signers but in the last week of the class were 1.5–2.0 times as great. This decrease probably means that the students needed less planning and recovery time as they developed more simultaneity of speech and signs and would be consistent with the conclusions of Windsor and Fristoe (1991). It could also be speculated that inexperienced signers begin by signing between words in SC, thereby elongating the interword intervals more, then move toward more simultaneous signing and speaking, thereby elongating speech segment durations as they became more proficient manually. In fact, word duration increased from the first week to the last week of the class primarily because of significant prolongation of the vowel in the experimental CVC word. Because vowel prolongation accompanied the reduction of interword intervals as the students gained experience signing, they may have attempted to increase the simultaneity of speech and sign by prolonging the vowels in their speech while they signed the word. Further research employing videotaping of students throughout a class would be necessary to confirm this speculation and determine how inexperienced signers learn to produce speech and signs simultaneously over the course of time.

Although speech segment durations were considerably prolonged during both the first and last weeks of the sign language class, the inexperienced speakers maintained the temporal rules of spoken English. Vowels were consistently longer before a voiced consonant and shorter before a voiceless consonant, and the low vowel /a/ was longer than the higher vowel /I/. Both of these results are consistent with the classic vowel studies of House (1961), House and Fairbanks (1953), and Peterson and Lehiste (1960) regarding vowel height and consonant voicing effects on duration and agree with the findings of Whitehead et al. (1995) regarding vowel duration rules followed by experienced signers. It is, therefore, reasonable to conclude that even inexperienced signers using SC conform to these selected temporal rules of English, despite the extensive elongation of speech segment and interword interval durations.

Voice Onset Time. Schiavetti, Whitehead, Metz, and Moore (1999) replicated the study of Schiavetti, Whitehead, Metz, Whitehead, and Mignerey (1996) to examine the effect of SC speaking rate reduction on VOT with inexperienced signers and found that inexperienced signers elongated VOT in the same way as the experienced signers with appropriate VOT relationships across place of articulation. Furthermore, the effect of speech rate on VOT was not altered by the semester of sign language classes, even though sentence duration during SC was slightly decreased as the signers became more fluent in SC toward the end of the class. This result suggests that the rate effect on VOT is very robust, and even inexperienced signers maintain the appropriate VOT contrast between voiced and voiceless members of cognate pairs during SC. Interestingly, the experienced signers in the Schiavetti et al. (1996) study showed a more enhanced VOT contrast between voiced and voiceless plosives in SC because their voiceless VOT increased more than their voiced VOT as rate slowed. Further research is necessary to determine when in the sign language learning process beginning signers start to enhance the VOT contrast as experienced signers do, because this
would further improve the speech model for hearing-impaired children.

**Final Consonant Voicing.** D’Avanzo, Graziano, Metz, Schiavetti, and Whitehead (1998) replicated the study of Metz et al. (1997) to examine the effect of SC speaking rate reduction on the duration of vowels preceding voiced and voiceless consonants. Inexperienced signers produced speech samples containing the same words with voiceless and voiced final consonants, and duration of the vowel preceding each target word was analyzed acoustically. Results showed that duration of vowels preceding voiced consonants was about 60–70 msec longer than vowel duration preceding voiceless consonants in both SA and SC conditions, with voiced to voiceless vowel duration ratios of 0.77 in SA and 0.79 in SC. Thus, the pattern of results was the same as for the experienced speakers, and the magnitude of the rate reduction effect was slightly larger.

All of these replication results are consistent with the conclusions of Windsor and Fristoe (1991) and Whitehead et al. (1995) that speech is merely slowed to approximate simultaneity of communication while the speaker is engaged in the slower manual task but that SC speech does not violate the specific temporal rules of spoken English that were studied. Although the speakers in the original studies were experienced in the use of SC, the replications have extended the generality of these findings to speakers with less experience. Therefore, the use of SC by parents or teachers should not raise a concern that a different speech model that violates English temporal rules is presented by inexperienced signers to hearing-impaired children.

**Perception of Specific Phonemic Contrasts**

**Perception of Final Consonant Voicing.** The study by Metz et al. (1997) referred to earlier regarding production of final consonant voicing also considered the effect of preceding vowel duration on perception of the voicing of syllable final consonants. Experienced signers showed vowel durations preceding voiced consonants in both SA and SC that were 50–65 msec longer than vowels preceding voiceless consonants, with voiced-to-voiceless vowel duration ratios of 0.72 in SA and 0.75 in SC. Metz et al. (1997) digitally edited those words to remove the final consonants and played them to normal listeners, who completed a forced-choice word recognition task. Of the 3,840 word identifications made by the listeners, 69.2% were correct for speech produced alone, and 77% were correct for speech produced during SC, but the difference between the two groups was not statistically significant. These results indicate that perception of final consonant voicing was not impaired by the duralional changes accompanying the typically slower speech rate of SC. D’Avanzo et al. (1998) replicated the study of Metz et al. (1997) with inexperienced signers and found similar results: of the 4,800 word identifications made by their listeners, 77% were correct for speech produced alone, and 74% were correct for speech produced during SC. The difference between the two groups was not statistically significant.

**Perception of Vowels.** Schiavetti et al. (2004) examined normal listeners’ identification of vowels in speech produced by experienced signers using SA versus SC and compared their data with previously published data on vowel identification (Hillenbrand et al., 1995) as an indication of degree of external validity of the results. Vowel perception was analyzed by means of a multiple-choice vowel identification test in which five listeners audited the speech samples provided by each of the 12 speakers and marked the /hVd/ words that they heard on each trial. Student volunteers who had completed an undergraduate phonetics course were recruited as listeners, as suggested by Hillenbrand et al. (1995), to control for vowel identification errors due to unfamiliarity with phonetic and orthographic vowel symbols.
Vowel identification was essentially the same in SC as in SA, and correct identification percentages for each vowel in both SA and SC conditions were comparable with those of Hillenbrand et al. (1995) and Peterson and Barney (1952). The pattern of vowel identification errors analyzed in confusion matrices to compare listeners’ perceived vowels to speakers’ intended vowels were very similar for SA and SC, and both were very similar to the confusion matrices reported by Hillenbrand et al. (1995) and Peterson and Barney (1952). Most errors were close to the intended vowel phonetically, and the most errors in all the data sets were concentrated around the vowels /a/ and /i/. Despite the temporal slowing of speech in SC, listeners were able to identify the speakers’ vowels as well as they identified vowels produced during SA. These results are consistent with previous research showing maintenance of phonetic rules of spoken English during SC (Schiavetti et al., 1996; Whitehead et al., 1995; Whitehead, Schiavetti et al., 1999; Whitehead, Whitehead et al., 1999) and indicate that speech produced during SC does not disrupt perception of the vowel pattern of spoken English, supporting its appropriateness as a speech model for deaf and hearing-impaired children and as a mode of communication with deaf and hearing-impaired adults.

Intelligibility of SC Speech to Hearing-Impaired Listeners. Whitehead, Schiavetti, MacKenzie, and Metz (2004) examined the overall intelligibility to hearing-impaired listeners of speech produced in SA versus SC. Speech intelligibility analysis was accomplished by means of Boothroyd’s (1985) forced-choice phonetic contrast test that permits analysis of eight different phonetic contrasts (vowel height and place; initial consonant place, continuance, and voicing; and final consonant place, continuance, and voicing) as well as prediction of phoneme, word, and sentence intelligibility from the overall phonetic contrast intelligibility score. Twelve severely to profoundly hearing-impaired students at the National Technical Institute for the Deaf audited recordings made by four experienced signers, and correct overall phonetic contrast percentages were computed for SA and SC conditions. The eight phonetic contrast errors made by the listeners in each condition were analyzed, and predictions of phoneme recognition, isolated word recognition, and sentence context word recognition measures of speech intelligibility were made from the phonetic contrast data using Boothroyd’s (1985) intelligibility prediction equations.

The results revealed similar means and standard deviations for SA and SC conditions on overall phonetic contrast percentages and all three of the predicted intelligibility phoneme, word, and sentence measures, indicating no difference in intelligibility between speech produced in the SA and the SC conditions. The individual phonetic contrast errors across listener-speaker pairs showed very similar patterns for both the SA and the SC conditions, with consonant place of articulation and continuance errors and vowel place errors being the most common errors in both conditions. Vowel place errors would be expected with hearing-impaired listeners because vowel place is cued by the higher frequency second formant, and consonant place and continuance errors would also be expected from the high-frequency spectral place cues, second formant transition cues, and continuance timing characteristics that might be perceptually difficult for these listeners. Despite the rate reduction, speech produced during SC was as intelligible overall to listeners as speech produced during SA. These results are consistent with previous research that has revealed the slowing of speech with maintenance of phonetic rules of spoken English. Because speech intelligibility was preserved for hearing-impaired listeners, it is reasonable to conclude, therefore, that SC is appropriate for use as a means of communication because the spoken message passes from speaker to listener with the same accuracy as is obtained with SA and can be used as an accurate speech model to present to hearing-impaired children and as an accurate mode of communication with hearing-impaired adults.

Intelligibility of SC Speech under Difficult Listening Conditions. The foregoing studies of perception of speech produced during SC used quiet listening conditions to compare data with normative speech perception studies in the speech and hearing science literature. The degree to which intelligibility is maintained under difficult listening conditions, such as speech in noise and filtered speech, is an important indicator of the
appropriateness of a developmental speech model and an efficient communication method. If researchers could determine the degree to which speech during SC maintains the acoustic redundancy necessary to resist intelligibility degradation by masking or frequency distortion, the external validity of SC research would be extended to common realistic listening conditions, such as noisy classrooms or electronic transmission systems. MacKenzie, Schiavetti, Whitehead, and Metz (in press) examined the intelligibility of speech produced in SA and SC when the speech was presented to normal hearing listeners under two experimental listening conditions: speech in noise and filtered speech.

The same method and speech recordings used by Whitehead et al. (2004) to examine SC intelligibility to hearing-impaired listeners were used to examine the effects of filtering and noise on intelligibility of speech produced during SC. Listeners were 24 hearing student volunteers, 12 of whom were randomly assigned to listen to speech and white noise presented at a signal-to-noise ratio of 0 dB and 12 of whom were randomly assigned to listen to low-pass filtered speech at a cutoff frequency of 600 Hz and a slope of 24 dB per octave. Results indicated no difference in intelligibility between speech produced in the SA and SC conditions for either the speech in noise or the filtered speech listening condition: Means were similar for SA and SC conditions on the overall phonetic contrast percentages and all three of the predicted intelligibility phoneme, word, and sentence measures. The individual phonetic contrast error patterns were very similar for both the SA and the SC conditions in both the noise and the filtered listening conditions and were similar to the error patterns shown by the hearing-impaired listeners. Despite the temporal slowing, speech produced during SC maintained as much acoustical redundancy as SA, as evidenced by its comparable resistance to intelligibility degradation by masking and filtering.

Quality of Speech Produced During SC. Huntington and Watton (1984) reported that three teachers using SC showed disruptions in the normal rhythm of speech and concluded that SC may not always expose deaf and hearing-impaired children to the typical prosodic and segmental features of normal speech that are necessary for the development of good oral communication skills. However, empirical research had not directly measured how natural speech quality was disrupted during specific manual tasks, such as fingerspelling, or documented the relationship between specific temporal speech characteristics and perception of speech quality or naturalness during SC. A valid and reliable speech naturalness measurement scale was developed for stuttering research by Martin, Haroldson, and Triden (1984) and was later used to study relationships among temporal speech characteristics and perceived speech naturalness of posttreatment stutterers by Metz, Schiavetti, and Sacco (1990), who found that temporal elongations similar to those reported by Whitehead et al. (1995) for SC were correlated with perceived speech naturalness.

Schiavetti et al. (1998) examined the effect of SC fingerspelling task difficulty (indexed as four graduated levels of fingerspelling task length from low to high) on temporal characteristics and perceived naturalness of speech and reported the correlations between five temporal characteristics of speech production and perceived speech naturalness in a multiple regression analysis designed to predict speech naturalness from the temporal characteristics. Speech during SC was rated more unnatural than during SA for all fingerspelling task length levels, and speech unnaturalness increased as fingerspelling task length increased in the SC condition but not in SA. The multiple regression analysis, using the five temporal characteristics of speech as predictor variables and the scaled speech naturalness as the predicted variable, indicated that sentence duration (reflecting speech rate) showed the highest correlation with speech naturalness, with moderate correlations between the other temporal characteristics and speech naturalness. The correlations among the other temporal measures were low to moderate, indicating that these variables were good candidates for entry into a regression analysis as predictor variables that could complement each other in the prediction of speech naturalness. This regression analysis demonstrated a highly significant relationship of these selected temporal characteristics of speech production and perceived speech naturalness. The results regarding speech naturalness and its relationship to temporal characteristics of speech were quite clear: Speech during SC was perceived as significantly
more unnatural than during SA, and perceived unnaturalness increased systematically with fingerspelling task length in the SC mode. These findings were consistent with the results of Metz, Schiavetti, and Sacco (1990), who reported significant correlations between sentence duration, sentential pause, and vowel duration in the speech of nonstutterers and posttreatment stutterers and the perceived naturalness of their speech. Although the results of both of these studies demonstrated the salience of temporal features of speech for perception of speech naturalness, other acoustical characteristics of speech, such as spectral and amplitude changes, need to be investigated to understand more completely the acoustical factors that influence listeners’ perception of speech naturalness during SC.

Conclusions

The results of the program of research described above led to the following conclusions:

1. Speech produced during SC was slower than speech alone, with specific increases observable in pause, consonant, vowel, word, and sentence durations.
2. Speech produced during SC did not violate any temporal rules of spoken English.
3. Speech produced during SC was perceived as less natural than speech alone, but specific consonants and vowels were equally perceptible in both speech alone and SC, and the overall intelligibility of speech produced during SC was equivalent to speech alone.
4. Temporal changes in speech produced during SC were similar to temporal changes observed in parentese and when speaking clearly for the hearing impaired.
5. Speech produced during SC did not demonstrate any acoustical or perceptual characteristics that would disqualify it as an appropriate developmental speech model for hearing-impaired children or as an accurate mode of communication with hearing-impaired adults.

The sum of these conclusions demonstrates that speech produced during SC presents an appropriate speech model because of its equivalence in acoustic phonetic contrasts, phoneme perception, and speech intelligibility to SA. However, the issue of speech quality deterioration due to temporal changes in SC is still somewhat disconcerting. The important speech quality issue concerns the degree to which the temporal disruptions and reduced speech naturalness can be tolerated as unwanted by-products of the potential communication improvement with SC. As Huntington and Watton (1984, p. 141) have stated: “The way a teacher talks must be as normal as possible. Slight concession regarding rate of utterances would, we feel, not be injurious, but general adherence to normal prosodic practice is imperative.” Strong and Charlson (1987, p. 381) were decidedly more critical of SC when they stated: “If a primary purpose of total communication programs is to make English more accessible to deaf children than is possible through lipreading, reading and writing alone, SC is quite possibly impairing if not preventing the realization of this goal.” They even suggested that the distortions in both manual and spoken channels occasioned by SC may be partly responsible for failures in the improvement of reading levels of deaf children over the years.

However, temporal disruptions and reduced speech naturalness have been accepted as unwanted but tolerated by-products of potentially improved communication in at least one other area. Yorkston, Hammen, Beukelman, and Traynor (1990) reached this conclusion when they found increased intelligibility but somewhat decreased naturalness in the speech of persons with dysarthria using rate-reduction treatment strategies. Yorkston et al. (1990, p. 558) stated: “In the clinical setting a slight reduction in speech naturalness may be an acceptable price to pay for a substantial improvement in intelligibility.” The important question for SC concerns the relative degrees to which it improves communication and disrupts speech naturalness.

This concept needs to be considered in relation to the results of Hyde et al. (1998), who found that deaf and hearing-impaired individuals judged that speech produced by teachers using simulated SC sounded more unnatural but more “understandable” than simulated SA speech. The results of Whitehead et al. (2004) of actual intelligibility tests showed that the SC and SA samples were equivalent in intelligibility. This result suggests the possibility that the listeners in the Hyde
et al. (1998) study may have judged SC as more understandable because the SC speech samples were easier to listen to at the slower rate. Future research needs to address this issue via comparisons of scaling and word recognition measures of intelligibility of SA and SC speech samples.

These intelligibility results also need to be discussed in comparison with the findings of Picheny et al. (1986) for speech produced “clearly” for hearing-impaired listeners. There is a meaningful behavioral difference between the condition that Picheny et al. studied and the SC condition, yet both conditions involve similar rate alterations. In the “speaking clearly” condition, speakers deliberately altered their speech to provide a clearer signal to a hearing-impaired listener, whereas in SC, speakers inadvertently reduced speech rate in the attempt to maintain simultaneity between spoken and manual communication (Windsor & Fristoe, 1991). The results of Picheny et al. (1986) indicated that speech intelligibility increased approximately 15% from their “conversational” to their “clear” conditions, and the acoustical characteristics of their clear conditions were similar but not identical to changes in SC reported by Schiavetti et al. (1996, 1999) and Whitehead et al. (1995), such as lengthened VOT, pause, and phoneme durations.

Further research is needed to determine the specific acoustical similarities and differences between rate-altered speech during SC and rate-altered speech produced “clearly” for hearing-impaired listeners and how these acoustical differences might affect intelligibility. Speech production research should examine differences between the acoustical changes observed between conversational and clear conditions and between SA and SC conditions to determine why clear, rate-altered speech became more intelligible and why SC rate-altered speech maintained the same level of intelligibility. For example, one important variable that has not been examined in SC research so far is amplitude, and it is possible that amplitude changes in specific speech sounds during speaking clearly that are lacking in SC may have important effects of speech intelligibility. There also may be important interactions between amplitude and duration changes that could influence intelligibility. Amplitude is particularly difficult to measure during SC research because of the need to keep constant the microphone-to-mouth distance and because of artifacts caused by hand movements near the microphone. New research strategies need to be developed to overcome these problems in future research so the role of specific amplitude changes in facilitating intelligibility can be examined carefully.

An interesting area for future research would be the possibility that rate-altered speech during SC could be modeled on speech produced “clearly” for hearing-impaired listeners. Adding characteristics of clear speech to SC might produce a more intelligible speech model for use with children and adults with hearing impairment. Such future research should examine the specific spectral and temporal acoustic characteristics of vowels and consonants and suprasegmental factors that have been shown in previous research to be related to the speech intelligibility of hearing-impaired speakers (Monsen, 1978; Metz et al., 1990). Although research has demonstrated that rate reduction does not decrease speech intelligibility and may increase it, research is needed to understand why intelligibility is maintained in some conditions and improved in others.

The importance of this notion can be inferred from Maxwell’s (1990) discussion of the training position concerning SC. Some authors believe that SC is not well implemented as a model for hearing-impaired children who are developing speech but could be better if people were trained to use SC. As Maxwell (1990, p. 373) has stated:

If teachers and parents lack sufficient skills, then, practically speaking, MCE/SC cannot accomplish what it is supposed to do. The possibility is that MCE/SC, in whatever specific form selected, might be more effective and successful if teachers and parents were better trained in its performance.

One possible approach to effective training to improve SC use could be to use the data of Picheny et al. (1986) to teach people how to speak clearly for hearing-impaired persons while signing. This might not be a difficult task because both speaking clearly and SC involve speech rate reduction. If we determine what caused clear rate-altered speech to become more intelligible and what caused SC rate-altered speech to maintain the same level of intelligibility, modeling SC
on speaking clearly might produce an improved SC speech model to use in training parents and teachers. In looking toward the future of SC, Maxwell (1990, p. 373) stated:

If one believes that the purpose of simultaneous communication is to represent fully and exactly the structure of the English spoken and that the availability of such sign input is critical to the English language acquisition of deaf children, then this training solution would hold great attraction. To date, the discussion has been focused almost exclusively on the morphological level of fit between speech and sign. We would like to see more research on other levels, so that a decision to improve training could achieve consensus about the level of fit that is the goal of training and so that the practitioners would be more aware of what is lost or distorted and could make decisions about how to teach those aspects of language.

An important goal for the future, then, is to focus a program of comparative research on the level of the comparative acoustical and perceptual characteristics of SC and speaking clearly so that we can improve the quality and intelligibility of SC speech in training programs for new SC users.

Another area of interest for viewing the results of this SC research concerns the relation of the results to the so-called motherese or parentese speech used by adults when addressing small children. Owens (2001) has summarized the characteristics of parentese and concluded that parents’ speech modifications may aid children’s language acquisition by bringing utterances into the “processing range” of the child and facilitating communication between adults and children. Many of the characteristics of parentese shown in Owens’s (2001) Table 7.4 are remarkably similar to the acoustical characteristics of speech produced during SC described above, especially the reduced rate, increased segment duration, and longer pause durations. Further research is needed to determine the effects of parentese and the role of the specific acoustical characteristics in accomplishing these effects. The results of this research will help to illuminate the purposes of the acoustical changes seen in both SC and parentese.

In addition, Owens (2001, p. 219) has pointed out that: “if adults simplify their language in order to be understood, then these modifications reflect cues coming from the child. Apparently, however, adults are not conscious of their modifications, nor are they attempting to teach language.”

Parentese, then, appears to be somewhat more like SC than speaking clearly in that the acoustical changes are not intentional; that is, they are not the result of a deliberate attempt to increase intelligibility. Thus, comparisons of both SC and parentese with speaking clearly may be informative in helping to determine the training needs of speakers using SC. If we can discover acoustical changes in speaking clearly that do not occur in either SC or parentese, we may identify candidate behaviors to target in training programs aimed at improving the intelligibility of speech during SC. For example, the influence of slower parental speech on VOT needs to be examined to determine whether parental rate reduction increases the absolute value of consonant voicing contrasts and influences children’s development of perception and production of the voicing feature. Perhaps an interaction between SC and parentese occurs in which VOT is not lengthened in parentese as much as it would be in combined SC and parentese. A large number of similar research questions could be asked about the influence of SC on the temporal characteristics of parental speech and what effect combined SC and parentese would have on the speech model presented to developing hearing-impaired children.

These questions also suggest the possibility of research regarding speech naturalness and its relationship to temporal characteristics of speech during SC, parentese, and speaking clearly. Hyde et al. (1998) have discussed the apparent paradox between perceived speech naturalness and improved intelligibility in SC, and this paradox may apply to parentese and speaking clearly, as well. Data are needed to address the similarity of temporal alterations in speech produced during SC, speaking clearly for the hearing impaired, and parentese speech and to determine their influence on the perception of speech naturalness. Such research also needs to determine whether speech quality can be kept from deteriorating in rate-altered speech and whether the acoustical characteristics of speaking
clearly may have implications for helping to maintain speech naturalness during SC. This is important from the training perspective discussed by Maxwell (1990) because speech quality measures may have an important influence on the acceptance of SC. Documenting the relation of SC to these other modes may help to determine the relative advantages and disadvantages of using SC in considering its quality versus its usefulness. Future research on the use of SC needs to be directed at the empirical analysis of its relative benefits in facilitating communication between hearing and hearing-impaired persons because the decision to adopt SC should be based on a logical and empirical analysis of its relative benefits in improving communication and its relative disadvantages as an “unnatural-sounding” manner of speaking.

In the meantime, it is reasonable to draw the practical conclusion from the empirical research reviewed here that SC is appropriate for use in communication between hearing and hearing-impaired persons and is an appropriate speech model for hearing-impaired children who are developing speech. This interim conclusion is based on acoustical data that indicate that speech produced during SC does not deviate from temporal and spectral characteristics that are important determinants of intelligibility and on perceptual results that show equivalent intelligibility of SC and SA. It is true that the quality of speech during SC is perceived as more unnatural than speech alone, primarily because of temporal elongations introduced in rate alteration, but this reduction of speech naturalness may be a tolerable price to pay for the advantages of SC. Future research and development should be aimed at finding methods to improve both the intelligibility and the quality of speech produced during SC and at developing training programs for SC users to improve their speech production while simultaneously signing.

References


Received October 20, 2003; revisions received January 15, 2004; accepted February 3, 2004.