CLINICAL LINGUISTICS & PHONETICS, 2002, VOL. 16, NO. 6, 403–424

Acoustic phenotypes for speech–genetics studies: toward an acoustic marker for residual /s/ distortions

HEATHER B. KARLSSON†, LAWRENCE D. SHRIBERG†, PETER FLIPSEN, JR†‡ and JANE L. MCSWEENY†

†Phonology Project, Waisman Center, University of Wisconsin-Madison, Madison, WI, USA
‡University of Tennessee, Knoxville, TN, USA

(*Received 23 August 2001; accepted 6 December 2001*)

Abstract

Findings in a prior study series indicate that acoustic markers may have the requisite sensitivity and specificity to discriminate speakers with histories of several types of speech disorders, one of which is posited to be genetically inherited. The present study in this series compares acoustic data from three groups of adolescent speakers. Group 1 speakers had residual dentalized /s/ distortions in conversational speech and histories of significant age-inappropriate deletion and substitution errors. Group 2 speakers also had residual dentalized /s/ distortions in conversational speech, but their speech histories were limited to dentalized distortions of /s/ and other fricatives/affricates. Group 3 speakers had typical speech on assessment and no histories of speech errors. Owing to the limited number of perceptually dentalized /s/ tokens produced by Groups 1 and 2 speakers in a phrase-level speech task, acoustic analyses were completed on /s/ tokens transcribed as correct for speakers in all groups. Moments analyses of /s/ spectra in three words with /s/-initial clusters yielded statistically significant differences and consistent trends for mean spectral frequency and spetcral variance for Group 1 compared with Group 2 speakers. These findings for perceptually normal /s/ tokens are interpreted as additional support for the potential of acoustic markers to discriminate speakers' speech–error histories. The discussion considers possible developmental and normalization correlates of the acoustic findings for speakers with each of the two types of speech–error histories studied in this paper.

Keywords: Articulation, genetics, moments analysis, phonology, speech disorders.

Address correspondence to: Lawrence D. Shriberg, The Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705, USA. e-mail: shriberg@waisman.wisc.edu

Clinical Linguistics & Phonetics
ISSN 0269-9206 print/ISSN 1464-5076 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/02699200210128954

Introduction

The research context for the present study is the hypothesis that at least one subtype of child speech disorder of currently unknown origin is genetically transmitted (Shriberg, Austin, Lewis, McSweeny and Wilson, 1997). As reviewed elsewhere, this hypothesis motivates the need for valid phenotype markers to classify the speech status of family members whose speech error histories in genetics studies may be unavailable or unreliable by recall report (Shriberg, 1991, 1993; Shriberg and Austin, 1998; Flipsen, Shriberg, Weismer, Karlsson and McSweeny, 1999, 2001; Shriberg, Flipsen, Karlsson and McSweeny, 2001).

To review the general problem, consider the four possible speech histories of a family member who is older than 9 years of age at the time a genetics study is conducted. One possibility is that a relative of the *proband* (the person identified as positive for the disorder) could have a history of significant speech delay of genetic origin. A second possibility is that this relative could have a significant speech delay not of genetic origin. Prior reports have suggested that in addition to an inherited (genetic) subtype of speech delay, there may be at least three other classes of aetiological origins of significant speech delay. Such classes include subtypes suspected to be associated with early recurrent otitis media with effusion, deficits in speech–motor control, or associated with psychosocial challenges (Shriberg *et al.*, 1997). A third possibility is that this family member may have a history of speech errors limited to distortions of /r/, /s/ or /l/, without the across-feature consonant deletions and substitutions, and the language deficits typically associated with early speech delay (Shriberg and Kwiatkowski, 1994). Epidemiological support for such feature- and error-limited forms of child speech–sound disorder, sometimes termed *lisping* or *lalling* (or an *articulation* disorder rather than a *phonological* disorder) is provided elsewhere (Lewis and Shriberg, 1994; Shriberg, 1994, 1999). For example, proportionally more males than females have histories of speech delay, whereas sex ratios for speakers with histories of only /s/ distortions appear to be equal or may even favour females. Finally, a fourth possibility is that the family member may have no history of a speech disorder.

Note that a family member who presents with residual speech–sound distortion errors at the time of speech testing may have any of the first three speech-disorder histories described above: speech delay presumed to be inherited, one of several types of speech delay presumed not to be inherited, or speech disorder limited to single-sound or sound class distortions. Moreover, a family member who does not have residual speech errors when tested may have one of the four speech histories, because he or she may have completely normalized errors associated with any of the first three histories, or may have no history of speech disorder. If we assume that only family members with prior speech delay presumed to be inherited should be classified as *affected* in speech–genetics studies, how can researchers determine the appropriate classifications for speakers who either have residual distortions that reflect one of the three speech–error histories or who do not have residual speech errors at the time of testing?

One possible method to identify and correctly classify the speech histories for family members with and without residual distortion errors may reside in fine-grained instrumental analysis of their speech. Specifically, close acoustic analysis of their correctly or incorrectly produced tokens might yield markers with sufficient sensitivity and specificity to discriminate among each of the four speech histories.

The data in the present paper were collected as part of a larger study exploring this possibility, with the current focus on production of English /s/. Because there is no current method to differentiate children with the proposed genetically inherited subtype of speech delay from those with the other non-genetic etiologies posited above, methods in the present study are limited to a contrast of the speech of children with histories of speech delay of *any* possible aetiological subtype (i.e. the first and second histories above) and the speech of children with histories of only distortion errors (the third history above). The following sections provide brief overviews of articulatory and acoustic phonetic information relevant to possible differences in /s/ production in speakers with histories of speech delay versus histories of feature-limited distortion errors.

Typical /s/ production

The cardinal sibilant fricative in many languages, /s/ is estimated to comprise approximately 8% of the consonant tokens in English conversational speech (cf. Shriberg and Kent, 1995). It has a widely distributed privilege of occurrence in English phonotactic rules and a central role in English morphological structure. Although acquired relatively early in phonemic development, perhaps due to the press from these distributional and morphophonemic factors, some children have early dentalized and/or lateralized distortions of /s/ that have been observed to persist throughout adulthood (Lewis and Shriberg, 1994). Alternative explanatory sources for the acquisition and persistence of such errors include: its challenging articulatory characteristics, involving motor control of intrinsic and extrinsic musculature; its salient auditory–perceptual characteristics, relative to the ease with which it can be discriminated from other sounds (e.g. voiceless fricatives) in a language; and its linguistic features, including phonotactic, allophonic and morphophonemic characteristics.

Articulation of /s/ is typically achieved by raising the tongue blade against the roof of the mouth to form a narrow constriction in the vicinity of the alveolar ridge (Stevens, 1998). Frication is produced primarily at the point at which the emerging airstream meets the lower incisors, which yields a dipole noise source (Stevens, 1998: 398). In a cineradiographic study of 10 young adults with typical speech and typical occlusion, Subtelny, Oya and Subtelny (1972) reported that mandibular and tongue tip position for /s/ in the sentence ‘Sister Suzy saw Sam’ were consistent across the different vowel contexts. The size and length of the lingua-dental constriction were consistent (1 mm constriction and 4 mm long), but tongue body position and lip opening were more variable.

In an electropalatographic study of nine typically-speaking 6–14-year-old children, Fletcher (1989) reported that the sibilant groove for /s/ and /z/ was narrower than for /ʃ/, /tʃ/ and /dʒ/. Fletcher also proposed two developmental trends across this age range: contact place moved posteriorly with age, and lingua-palatal contact reduced with age. More recently, in magnetic resonance imaging studies of fricatives in four adults, Narayanan, Alwan and Haker (1995) described alveolar strident production as characterized by concave cross-sectional shaping for the tongue body with speaker-dependent tongue tip position (apical or laminal). More anterior medial grooving and lateral lingua-palatal contact was found for the apical than for the laminal production. Tongue shape asymmetries were also noted by subject and by sound. Ultrasound studies by Stone and Lundberg (1996) have characterized the two primary features of /s/ production as tongue grooving and bilateral palatal contact of tongue margins. Thus, in studies using several instrumental approaches, the central features of typical /s/ production appear to be a narrowly grooved tongue, alveolar or laminal tongue tip orientation, generally consistent constriction size, bilateral tongue contact at the palate, stability across at least some vowel contexts and possible changes in articulatory features across the life span.

Jongman, Wayland and Wong (2000) provided the most comprehensive recent acoustic data using moments analysis to characterize English fricatives as produced by 20 typically speaking adults. Jongman *et al.* concluded that each of the four spectral moments provided useful information to discriminate among the four English voiceless fricatives /s/, /ʃ/, /θ/ and /f/. Specifically: moment 1 (M1: mean frequency) was useful to distinguish high-frequency /s/ from lower frequency /ʃ/ and mid-frequency /f/ and /θ/; increased moment 2 (M2: variance) values were associated with decreasing sibilance, useful for distinguishing /s/ and /ʃ/ from /f/ and /θ/; moment 3 (M3: skew) yielded information about the concentration of low frequency energy that is typical of /ʃ/ and that distinguishes it from the relatively higher frequency concentration of energy typical of /s/; and moment 4 (M4: kurtosis) values were generally greater for /s/ than obtained for other fricatives, indicating the presence of more clearly defined high frequency energies.

In a study of 26 typically-speaking 9–15-year-old adolescents that provided the reference data for the present study, Flipsen *et al.* (1999) reported several findings relevant for acoustic studies of /s/ production. Acoustic data were based on speakers' productions of 12 words containing singleton /s/ and /s/ in clusters in word-initial, word-medial (singleton only) and word-final position. Spectral moments analysis indicated that phonetic context did have an effect on the moment values, particularly M2 values for the onsets and offsets (20 ms) of /s/ spectra. M2 values were depressed, apparently due to the effects of surrounding vowels that lower mean frequency and increased variance. Therefore, the frication midpoint was least associated with the place/manner characteristics of adjacent vowels and consonants. There were also significant sex-related effects particularly for M1 and M3 values, but there were no significant differences associated with age for speakers in this age range. The sex-related effects for the M1 and M3 data were most likely related to cavity size, which presumably is smaller for female speakers in this age range. M4 data did not provide useful information for the questions posed in the Flipsen *et al.* study.

Atypical /s/ production

Instrumental studies of atypical /s/ production in children and adults also report reliable differences not observed in phonetic transcription that may be important for theory and practice. In an electropalatographic study of three children with speech sound disorders, Hardcastle and Morgan (1982) reported individual differences in the tongue positioning and making of stops in /s/ clusters that were not observed using acoustic–perceptual procedures. In a review of contemporary studies using EPG techniques, Gibbon (1999) described children with 'undifferentiated lingual gestures'. She noted that such gestures yielded speech productions that neutralized phonemic contrasts due to limited control of the tongue, particularly at the lateral margins.

A limited number of studies have investigated the acoustic correlates of dentalized /s/ production in children's speech. Daniloff, Wilcox and Stephens (1980) studied two children with dentalized /s/, two with lateralized /s/, two with ‘other’ /s/ errors and two typically speaking children, all 6–8 years of age. Compared with the typical /s/ productions, the dentalized /s/ productions of these children were characterized by flatter, higher frequency and less intense noise spectra. Additionally, the investigators noted fewer and reduced phonetic context effects in the /s/ productions of the children with dentalized /s/. Baum and McNutt (1990) compared the ‘frontal’ /s/ productions of five children aged 5–6 years and five children aged 7–8 years with the /s/ productions of 10 age-matched controls. Findings indicated that the dentalized /s/ productions had lower centroid frequencies than the /s/ productions judged to be normal. There is no clear explanation for the different findings in these two studies, although likely sources include differences in both subject samples and methodologies. In a study that included typically speaking adults, children with typical /s/ production and children with misarticulated /s/ (many of whom presumably produced dental /s/), Weismer and Elbert (1982) reported more temporal variability in /s/ production for the children with misarticulated /s/ compared with the /s/ productions of typically speaking children. Differences were greater in /s/ clusters compared with /s/ singleton contexts. Miccio, Forrest and Elbert (1996) reported that the fricative productions of children with disordered phonologies who have acquired a contrast between sibilant and non-sibilant fricatives are acoustically distinguishable from the sibilant productions of their typically speaking peers.

As suggested in the findings received to this point, there is good evidence for reliable differences in articulatory behaviours that may not be observable by auditory–perceptual and/or acoustic methods. Stoner, Gately and Rivers (1987) found no perceptual or acoustic differences between apical and dorsal /s/ productions in a group of eight adults. Forrest, Weismer, Hodge, Dinnsen and Elbert (1990) used moments analysis to determine whether the acoustic profiles of children who substituted [t] for /k/ differed from the profiles of typically speaking children. Findings indicated that the stops produced by children with stop substitutions differed in spectral shape and mean frequencies from the stops produced by the typically speaking children. Moreover, values on these two variables differed for one speech-disordered child with an emerging /k/ compared with values for speech-disordered children without indications of emerging /k/. The child’s /k/ productions remained acoustically distinguishable from the typically speaking children’s /k/ productions, even when his /k/ sounds were perceptually transcribed as ‘normal’. A subsequent study of the spectral contrasts for /t/ and /k/ indicated that productions from children who had recently acquired the t/k contrast in the initial position were spectrally distinguishable from the productions of typically speaking controls (Forrest, Weismer, Elbert and Dinnsen, 1994). When they had acquired the t/k contrast in all word positions, however, the formerly misarticulating children produced /t/ and /k/ sounds that were spectrally indistinguishable from the /t/ and /k/ produced by their typically speaking peers.

A final articulatory–perceptual consideration relevant for the present study concerns effects associated with coupling of the front and back cavities (ahead of and behind the constriction) in the vocal tract (Kent, Dembowski and Lass, 1996). Coupling produces an output spectrum that is affected by the resonances of the back cavity and of the constriction itself. These resonances converge (are reinforced) when the cavities are uncoupled. Thus, perceptually important resonances are reinforced when there are small constrictions as the result of fricative production. With excess coupling, acoustic cues may be harder to perceive. Consequently, a non-tight constriction (as in dentalized productions) could have different qualities and resonances from those produced with a tight constriction. For this reason, differences in the central tendency of spectral moments may occur for a perceptually normal /s/ produced with a consistently more coupled production. McGowan and Nittrouer (1988) noted that, in comparison with adults, children tend to produce /s/ and /ʃ/ with greater coupling of the front and back cavities. Such productions provide more input from the glottis to the noise spectrum, resulting in greater amplitudes in the lower frequencies, which potentially lowers measures of central tendency.

Rationale and statement of the problem

For the purposes of the speech–genetics studies described above, the description of dentalized /s/ productions using the method of moments analysis might provide information on speakers' speech histories. As reviewed, acoustic analysis might be sensitive to subphonemic differences that are not detected using auditory–perceptual methods. Presumably, several articulatory parameters could be reflected only in acoustic data, including instability in the underlying representation of /s/ targets and/or lack of precision in tongue placement or grooving during /s/ production. Specifically, the acoustic signal may be sensitive to small changes in the configuration of the tongue and the extent to the coupling or decoupling of the cavities ahead of and behind the constriction formed by /s/ production. Additionally, any differing contributions from glottal sources have the potential to affect average values in the spectral moments. Prior findings (Flipsen *et al.*, 1999) and pilot data (Karlsson, 1999) suggest that controls are required for possible phonetic context effects in the speech stimuli, and that M1 and M2 are most likely to be sensitive to developmental differences in the acquisition of /s/.

Methods

Participants

Recruitment

A total of 122 potential participants for the current study were ascertained from three sources. The first source was a pool of 9–17-year-old children who had been treated for speech delay at a university phonology clinic 5–10 years before the present study. For the present study, the children's parents were contacted by mail and by a follow-up telephone interview. Of an original group of 89 children, 58 (65%) children and their parents were located, and all agreed to participate in an assessment session.

The second source of potential participants included 38 9–16-year-old children who were identified by speech–language clinicians in the Madison Metropolitan School District as having speech-error histories limited to speech-sound distortions. The goal was to identify children within the same age range as those in the first sample, but whose speech-error histories were limited to distortions of the English sibilants (primarily /s/ and /z/) or the English rhotics (/r/, /ɝ/ and /ɚ/).

The third set of potential participants was a group of 26 9–15-year-old typically speaking children with no history of speech disorder. These speakers were nominated by their teachers as classroom-matched controls for the 38 adolescent speakers from the second sample.

Study groups

Several exclusionary criteria were used to assign all potential participants to one of the following three study groups: Group 1: prior speech delay; Group 2: prior speech distortions; and Group 3: controls.

The primary inclusionary criterion for Group 1 participants was a history of prior speech delay that included dentalized /s/ productions. Each such subject history was documented by clinical records indicating treatment at a university speech clinic approximately 10 years before the present study. Potential participants whose histories or current production of /s/ in conversational speech included lateralized /s/ productions were excluded to avoid the possibility of unknown sources of variance in the acoustic data. Candidates were not excluded if they had distortion errors in other sound classes, including derhotacized rhotics and/or velarized pre-vocalic liquids. Of the original 58 children with prior speech delay, 13 children with a history of dentalized /s/ errors were identified. Of these 13 children, four speakers were excluded because they also had lateralized distortions.

The primary inclusionary criterion for Group 2 was a referral from a school speech-language therapist who had documented residual errors in speech therapy for each participant. Exclusionary criteria were production of lateralized /s/, or a history of a more general speech delay as reported by the child or accompanying caregiver, or as suspected by the examiner in view of the assessment information. Of the 14 eligible participants with a history of producing dentalized /s/, two were excluded due to concurrent production of lateralized /s/, and three were excluded due to the possibility of prior undetected speech delay.

As shown in table 1, these inclusionary and exclusionary criteria resulted in the retention of nine participants with prior speech delay for Group 1, including four males aged 9 (years); 1 (month) to 14;6, and five females aged 9;8 to 15;7. These criteria also resulted in the retention of nine participants for Group 2, including two males aged 11;4 and 12;1 and seven females aged 9;4 to 14;5. The typically speaking participants in Group 3 included 12 females aged 9;7 to 15;0 years and 14 males aged 9;7 to 14;10 (cf. Flipsen *et al*., 1999). Note that the participants selected for Groups 1 and 2 were not required to have one or more dentalized /s/s during any of the speech tasks administered for the present study.

Assessment protocol

The speech data for the current report were obtained from two tasks, which were part of a 90-min assessment protocol described elsewhere (Flipsen *et al*., 2001). All testing and phonetic transcription was conducted by the fourth author, an experienced speech–language examiner and research transcriber.

Conversational sample

A conversational sample was obtained in a quiet test suite. Instrumentation included a Sony 5000EV audiocassette recorder and Teac ME-50 microphone positioned so that mouth-to-microphone distance was approximately 10–15 cm. The conversational samples were transcribed to a 100 first-occurrence words criterion (Shriberg, Allen, McSweeny and Wilson, 2001) using well-developed conventions for narrow phonetic transcription (Shriberg, 1993; Shriberg and Kent, 1995).

Table 1. *Description of participants in the three study groups*

		Age (months)			Sex				PPVT-R[a] standard score	
Group	*n*	Mean	SD	Range	% Male	% Female	Prior speech delay	History of dentalized /s/ distortions	Mean	SD
1	9	152.8	29.4	109–187	44	56	yes	yes	106.7	22.1
2	9	144.5	17.8	112–173	22	78	no	yes	107.9	17.9
3	26	148.0	21.1	115–182	54	46	no	no	118.9	12.1

[a] Peabody Picture Vocabulary Test—Revised (Dunn and Dunn, 1981).

Speech task

Acoustic tokens were obtained from a speech production task designed to yield multiple types and tokens of several English phonemes produced in the carrier phrase 'Say_again'. A total of 24 target words was repeated five times and read out loud by the examiner, who was positioned so that the speaker could not see her or the word list during the task. Recordings were made using a head-mounted Shure SM-10A microphone connected to a Sony 5000EV taperecorder. The microphone was tilted toward the speaker and was positioned no more than 5 cm from the speaker's nose and approximately 4 cm from the speaker's lips. Children were asked to repeat the target in the carrier phrase while maintaining volume within a preset range as indicated by the VU metre on the taperecorder. The examiner monitored the participants' alertness and performance and asked speakers to repeat a phrase if the target appeared to be misunderstood or contained interword pauses or dysfluencies.

Acoustic analysis

Procedures

Acoustic analyses were completed by the first author and a research assistant, each of whom had completed a course in speech acoustics and were trained for the analyses using a procedural manual developed for the present and associated studies (Flipsen, Tjaden, Weismer and Karlsson, 1996). Each analyst was randomly assigned approximately half of the original pool of 122 speech samples. Tokens were first digitized using a Sony 5000EV taperecorder as the input source and a Sound Blaster AWE32 PNP A/D sound card connected to a Pentium-based PC. The signal was sampled at 22 kHz with 15 bits of quantization using the CSpeech (Milenkovic, 1996) record facility, including low-pass filtering at 9.8 kHz (see Flipsen *et al*., 1999, for constraints on data analysis associated with this filter condition) and a pass-band attenuation of -72 dB. Tokens were not digitized if they contained additional phonemes, substitutions for the primary vowels, dysfluencies or obvious interword pauses. Pauses, defined as any period of silence $\geqslant$250 ms (Miller, Grosjean and Lomanto, 1984), were measured from the wide-band spectrograms generated with a bandwidth of 500 Hz. Once confirmed as usable, the target word in the phrase interval from the start of /eɪ/ in *say* to the closure for /ɡ/ in *again* was isolated, digitized and stored.

Based on the methodological and substantive findings reviewed previously (Flipsen *et al*., 1999), the five words selected for analysis in the present study were *sin*, *soon*, *skin*, *spin* and *spoon*. Three considerations supported a decision to limit acoustic analysis to this subset of five words with /s/ in the initial position. First, sounds in word-initial position are important in phonological representations, yet may be difficult to discriminate phonetically due to the lack of formant cues preceding production. Second, representational challenges may be compounded in words with initial /s/ clusters, due to the short duration of /s/ in this phonetic context (Jongman, 1989). Third, preliminary analysis indicated that the strongest trends for the crucial comparisons of Group 1 with Group 2 speakers might be associated with word-initial /s/ tokens. The following criteria were used to segment the /s/ for analysis: (1) for *sin* and *soon*, from the last glottal pulse of /eɪ/ in *say* or in the beginning of frication (glottal pulse criterion was preferred unless frication was not continuous with the vowel), to the first glottal pulse of the vowel following /s/; and (2) for *skin*, *spin* and *spoon*, from the last glottal pulse of /eɪ/ in *say* or the beginning of frication, to the onset of closure for the stop consonant following /s/.

The /s/ segments for each speaker were analysed using the *moments* batch command function in CSpeech. Analyses were completed using a 20-ms Hamming window with a 10-ms step. For each token, three temporal points in the frication noise were identified from the moments output–onset, midpoint and offset. The onset was defined as the first analysis window in the output (i.e. the first 20 ms of the frication). The midpoint was defined as the middle analysis window (i.e. the middle 20 ms). In cases where a token contained an even number of analysis windows, one of the two middle windows was selected randomly. Offset was defined as the last analysis window (i.e. the last 20 ms).

Figure 1 provides graphic illustrations of segmentation and the quantitative data provided by moments analysis. Pilot studies indicated that information on M1 and M2 was relevant for the present question, whereas data on M3 and M4 were either not reliable due to data loss or were not relevant. For illustrative purposes, the three panels in figure 1 were obtained from the speaker in each group whose M1 and M2 data for the word *spin* best approximated the group means. The vertical lines in the exemplars labelled *A* in each panel illustrate the midpoints for /s/ within the fricative spectra in each spectrogram. The graphics labelled *B* in each panel are the fast Fourier transforms for this section of /s/, providing information on peak spectral energy. The lines labelled *C* in each panel are the traces produced for moments analysis. They illustrate the variance in spectral energy for this section of /s/ (M2), as well as the spectral peak. Notice the differences in the mean frequencies (B) in these expemplars from each of the three groups, as well as the differences in variance and spectral peaks (C).

Tokens

All usable responses to the speech task were phonetically transcribed by the fourth author. In contrast to speakers' articulation of /s/ in the conversational samples, which averaged approximately 9% dentalized distortions (mean: 8.9, SD: 7.2, range: 2–26) for Group 1 and 20% dentalized distortions (mean: 19.7, SD: 13.4, range: 0–41) for Group 2, most of the speech task /s/ tokens were transcribed as perceptually acceptable. For the /s/ initial tokens in the speech task, Group 1 participants produced approximately 7% dentalized distortions (mean: 7.4%, SD: 11.3, range: 0–10%) and Group 2 participants also produced approximately 7% dentalized distortions (mean: 7.4, SD: 7.4, range: 0–13%). Such notable speech differences in conversation versus imitative responses to citation tasks have been frequently reported in child speech–sound disorders (cf. Morrison and Shriberg, 1992). For the present speakers, most had almost fully normalized their /s/ distortions in both conversational speech and short phrases. Based on the reported individual differences among perceptually typical tokens reported in Forrest *et al.* (1994), the decision was made to analyse all the speech task tokens transcribed as correct. Perceptually correct /s/ initial tokens that included non-error diacritic symbols (e.g. palatalised, fronted, whistled) were excluded. These exclusions yielded 167 (74% of the analysable /s/ tokens) correct, diacritic-free /s/ initial tokens for Group 1, 186 (83% of the analysable /s/ tokens) for Group 2 and 619 (99.7% of the analysable /s/ tokens) for Group 3.

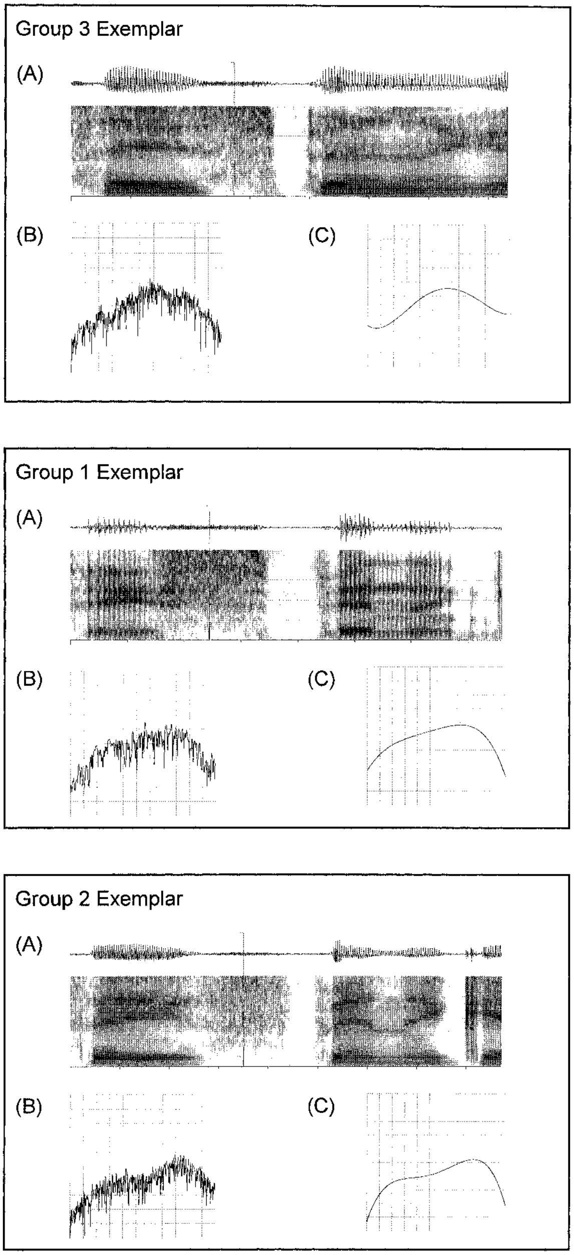

Figure 1. *Spectrograms, fast Fourier transforms and variance patterns for exemplars from one subject in each of the three study groups.*

Reliability estimates

Acoustics measures: first reliability estimate

Inter- and intrajudge agreement and standard errors of measurement were calculated for 20% of the /s/ tokens in all word positions from 15% of the participants. Complete descriptions of these agreement values are provided in Flipsen *et al.*

(1999). Interjudge agreement ranged from 0.10 to 0.67 kHz for M1 and 0.07 to 0.27 kHz for M2. Intrajudge agreement for Assistant 1 ranged from 0.35 to 0.73 kHz for M1 and 0.14 to 0.27 kHz for M2. Intrajudge agreement for Assistant 2 ranged from 0.20 to 0.56 kHz for M1 and 0.13 to 0.26 kHz for M2. Standard error of measurement values ranged from 0.16 to 0.55 kHz for M1 and 0.11 to 0.29 kHz for M2.

Acoustics measures: second reliability estimate

Four years following completion of the agreement estimate based on samples from the typical speakers (Flipsen *et al.*, 1999), a second reliability estimate was obtained. At that time, the research assistant was no longer available. The first author remeasured the first two tokens (40%) of each of the 10 target words from seven of the 122 (5.7%) speakers, three of whom had originally been analysed by the assistant. The sample included randomly selected speakers from each of the three groups, consisting of 22.2, 22.2 and 10.7% of the speakers from Groups 1, 2 and 3, respectively. Averaged standard errors of measurement ranged from 0.12 to 0.44 kHz for M1 and 0.11 to 0.16 kHz for M2. Averaged interjudge agreement values ranged from 0.12 to 0.59 kHz for M1 and 0.06 to 0.13 kHz for M2. Averaged intrajudge agreement values ranged from 0.12 to 0.36 kHz for M1 and 0.09 to 0.15 kHz for M2. Larger interjudge, intrajudge and standard error of measure values tended to occur in the frication onsets and offsets, which was expected due to the difficulty in identifying onsets and offsets. The inter- and intrajudge reliability and standard error of measurement data from the two estimates are regarded as adequate for the current purposes.

Transcription reliability: conversational speech samples

Estimates of the reliability of transcription for the converstaional speech samples had been obtained as part of the parent study of residual speech errors that included all 122 children. Intrajudge agreement was estimated by having the transcriber retranscribe a randomly selected 10% (12 speakers out of 122) at least 12 months after original transcription. Based on the total sample of 2613 retranscribed words, point-to-point intrajudge agreement for consonants was 96.9% for broad transcription and 90.4% for narrow transcription. Agreement for vowels was 90.4% for broad transcription and 82.1% for narrow transcription.

Transcription reliability: speech task

Transcriptions of all words in the speech task were redone by the original transcriber for 12 of 84 (14.3%) children included in this study and a prior study of acoustic markers in speakers with histories of /ɝ/ distortions (Shriberg *et al.*, 2001). Based on the total sample of 1440 retranscribed words, point-to-point intrajudge agreement for narrow phonetic transcription of /s/, /r/ and /ɝ/ was 90.0, 93.7 and 89.7%, respectively. Initial, medial and final word position narrow transcription reliability was 85.7, 94.2 and 94.4% respectively.

Speaker normalization

Preliminary analysis of these data by speaker age indicated that there was a decrease in M1 values at approximately 132 months, possibly coincident with puberty and growth patterns. For these reasons, together with the sex effects reported in Flipsen *et al*. (1999) and the imbalance in the percentage of speakers of each sex in Groups 1 and 2 (table 1), *z* scores were computed for all participants. The means and SDs for the computations were obtained from the Flipsen *et al*. (1999) reference data for /s/ productions, using reference groups cross-classified by age (less or greater than 132 months) and sex (male and female). As these speakers comprised the control group (Group 3), they were normalized against their own subgroup means and SDs.

Results

Raw data and analysis of standardized data

Table 2 includes raw data (means, SDs) for each of the two spectral moments for each speaker group, with spectral data provided for the three sampling periods for the five /s/ words. As described above, these data are constrained by age and sex considerations, notably the possibility of the truncated high frequency energy data for females due to the 9.8 kHz filtering limitations. Notice that the M1 values for Group 1 are lower than those for Group 2, which is consistent with the greater number of males in Group 1. M2 means were more variable: Group 2 speakers tended to have higher M2 values than Group 1 speakers for the frication onset and offset, but they had lower values than Group 1 speakers for the data obtained at the frication midpoint.

Table 3 includes the primary findings based on the standardized values for each of the two spectral moments. Pair-wise, between-group comparisons using binary logistic regressions were completed for each word of the three sampling periods. Conventional significance levels for the obtained *p* values are indicated by asterisk(s). The pattern of statistically significant outcomes across the 72 pair-wise comparisons suggests the following two substantive findings.

First, the most robust acoustic differences for Group 1 compared with Group 2 speakers were the M1 comparisons at the frication midpoint (table 3: frication midpoint) in the /s/ cluster words *spin*, *spoon* and *skin*. The direction of this pattern of statistically significant differences indicates that the Group 2 speakers had larger (more positive) *z* scores compared with Group 1 speakers. As expected, the pattern of statistically significant differences for M1 at the frication midpoint also occurred for the comparisons of Group 2 speakers to the Group 3 control speakers, because the *z* score transformations were based on reference data from these typical speakers. In contrast, the M1 *z* scores of speakers in Group 1 did not differ significantly from the M1 *z* scores of the Group 3 control speakers at the frication midpoint. Thus, the perceptually correct /s/ tokens from Group 2 speakers had higher mean frequency energies at the midpoint than tokens from Group 1 speakers and tokens from Group 3 control speakers.

Second, the pattern of statistically significant findings in table 3 suggests that the M2 values of Group 1 speakers at the frication midpoint of /s/ also tended to differ from the M2 values for speakers in the other two groups, in the words *spin*, *skin* and *soon*. Group 1 speakers had higher *z* scores than both the Group 2 speakers and Group 3 control speakers. The M2 values for the Group 2 speakers were not significantly different from those of the Group 3 speakers, nor did they trend in a consistent direction. Thus, compared with the perceptually correct /s/ tokens produced by Group 2 and 3 speakers, the perceptually correct /s/ tokens produced by Group 1 speakers tended to have greater variance (M2), with spectral energies spread more broadly across the frequency spectrum.

Notice also in table 3, the pattern of statistically significant differences for the M1 and M2 values sampled at the frication onset and offset. Whereas the largest and most consistent between-group differences were obtained at the midpoint, all trends were less clear at the other sampling points. At frication onset, none of the Group 1–2 M1 comparisons and only two of the Group 1–2 M2 comparisons reached statistical significance, and the directions of these differences were inconsistent. For the offset samples, none of the Group 1–2 comparisons for either moment reached statistical significance. Consistent with the normalization design, the variances (M2) for both Groups 1 and 2 were significantly higher than those for Group 3 typical speakers for many words sampled at /s/ onset and /s/ offset.

Figure 2 provides a graphic illustration of the token-level, distributional data for the *z* score findings just reviewed for M1 at the frication midpoint. As shown in the top panel, the M1 *z* scores for /s/ for the Group 3 (control) speakers form an essentially normal distribution confirming the self-normalized mean of 0.0. The distribution of M1 *z* scores for /s/ produced by the Group 1 speakers (middle panel) is also fairly normal, with somewhat more skew to the left. In comparison to these two distributions, the distribution of M1 *z* scores for /s/ produced by the Group 2 speakers is moved rightward on the abscissa, with scores more notably skewed to positive values on the right.

Discussion

These preliminary data on two groups of speakers with differing speech-error histories are viewed as support for the validity and potential utility of acoustic markers in descriptive-explanatory research. Individual overlaps in *z* score values for M1 and M2 in this small data set were too great to provide a marker with strong sensitivity and specificity. However, findings support the need for continued study using both adolescent and older speakers who produce dentalized /s/ errors on structured evocation tasks such as the one used in this study. The following comments address possible articulatory correlates of the findings, and possible developmental considerations for subtypes of child speech disorders.

Articulatory correlates

The trend for higher standardized M1 values in the /s/ productions of the Group 2 speakers (figure 2) may be consistent with several alternative articulatory correlates. One possible correlate is that speakers with speech-error histories limited to /s/ distortions may have a more forward tongue carriage than speakers with histories of an 'across-the-board' speech delay (i.e. Group 1). A forward tongue carriage is consistent with a shorter front resonating cavity, which may result in higher spectral energies for /s/ productions. A second potential articulatory correlate could be reduced coupling of the front and back resonating cavities, which may lessen the contribution of back cavity lower frequency resonances. As reviewed previously, such effects are consistent with tighter constriction of the tongue-alveolar posture, and may characterize some types of dentalized /s/ productions.

The M2 data support the possibility of additional or alternative articulatory correlates in relation to those described for M1. The M2 data indicate that the correct /s/ tokens from Group 1 speakers had larger variance than tokens produced by Groups 2 and 3 (control) speakers, attesting that Group 1 speakers produced /s/

Table 2. *Non-standardized spectral data (KHz) for participants in the three study groups*

		Frication onset							
						Group 3			
		Group 1		Group 2		Females		Males	
Word	Moment	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Spin	1	4.32	1.05	4.77	1.40	5.35	1.28	4.73	0.72
	2	1.85	0.36	1.87	0.36	1.83	0.037	1.46	0.26
Spoon	1	4.60	1.05	5.01	1.00	5.28	1.25	4.69	0.81
	2	1.74	0.41	2.06	0.36	1.76	0.34	1.39	0.30
Skin	1	4.36	0.99	4.58	1.32	5.26	1.35	4.77	0.72
	2	1.78	0.40	2.03	0.44	1.71	0.32	1.48	0.24
Sin	1	4.49	0.98	4.67	1.39	4.93	1.14	4.83	0.72
	2	1.88	0.43	1.83	0.41	1.86	0.30	1.54	0.24
Soon	1	4.33	0.97	4.48	1.25	5.13	1.16	4.68	0.61
	2	1.72	0.41	1.87	0.37	1.78	0.34	1.38	0.27
		Frication midpoint							
Spin	1	6.50	0.98	7.18	0.65	7.14	1.03	5.84	0.70
	2	1.52	0.39	1.41	0.26	1.39	0.21	1.31	0.29
Spoon	1	6.52	0.88	7.26	0.74	6.85	1.09	5.73	0.85
	2	1.52	0.29	1.38	0.32	1.41	0.24	1.24	0.24
Skin	1	6.57	0.84	7.27	0.62	7.09	0.85	6.07	0.64
	2	1.56	0.47	1.40	0.31	1.52	0.34	1.26	0.27
Sin	1	6.83	0.94	7.25	1.05	7.39	0.85	6.35	0.67
	2	1.50	0.39	1.41	0.31	1.44	0.32	1.22	0.25
Soon	1	6.64	0.80	7.07	0.29	7.03	0.90	5.89	0.84
	2	1.37	0.29	1.26	0.29	1.27	0.26	1.26	0.31
		Frication offset							
Spin	1	5.40	1.12	6.12	1.09	6.35	1.06	5.35	0.70
	2	1.34	0.37	1.38	0.36	1.48	0.29	1.47	0.29
Spoon	1	4.90	1.18	5.56	1.21	5.86	1.29	4.33	0.64
	2	1.27	0.31	1.41	0.41	1.16	0.14	1.27	0.36
Skin	1	5.61	1.14	6.48	1.21	6.35	1.06	5.35	0.70
	2	1.79	0.40	1.66	0.41	1.48	0.29	1.47	0.29
Sin	1	5.13	1.43	5.29	1.07	5.31	1.00	4.90	0.71
	2	1.86	0.31	1.96	0.45	1.93	0.37	1.56	0.40
Soon	1	4.41	1.20	4.68	0.80	4.35	0.84	3.95	0.50
	2	1.71	0.57	1.68	0.40	1.52	0.29	1.30	0.28

with more diffuse distribution of spectral energies. Such speech effects might be viewed as the result of increased coupling of the cavities anterior and posterior to the constriction, whereby additional variance is due to the addition of back cavity resonance in the region of F3 (Stevens, 1998: 402). The Jongman *et al.* (2000) study of /s/ production in typically speaking adults posited that larger M2 values were

Table 3. *Inferential statistical findings for comparisons of standardized (z score) values for participants in groups 1, 2 and 3*

			Groups 1–2			Groups 1–3			Groups 2–3		
			Mean *z*			Mean *z*			Mean *z*		
Moment	Frication	Word	Group 1	Group 2	BLR[a]	Group 1	Group 3	BLR	Group 2	Group 3	BLR
1	Onset	spin	−0.572	−0.309	0.312	−0.572	−0.0001	0.006**	−0.309	−0.0001	0.138
		spoon	−0.273	−0.072	0.355	−0.273	−0.0002	0.156	−0.072	−0.0002	0.678
		skin	−0.444	−0.407	0.859	−0.444	0.0002	0.021*	−0.407	0.0002	0.095
		sin	−0.266	−0.067	0.435	−0.266	−0.0001	0.150	−0.067	−0.0001	0.731
		soon	−0.431	−0.386	0.856	−0.431	−0.0001	0.031*	−0.386	−0.0001	0.044*
	Midpoint	spin	−0.116	0.312	0.040*	−0.116	−0.0003	0.544	0.312	−0.0003	0.079
		spoon	0.176	0.6927	0.002**	0.176	0.0000	0.240	0.6927	0.0000	0.000***
		skin	−0.083	0.453	0.013**	−0.083	0.0003	0.651	0.453	0.0003	0.013*
		sin	−0.132	0.099	0.435	−0.132	−0.0002	0.480	0.099	−0.0002	0.613
		soon	0.229	0.348	0.856	0.229	−0.0001	0.220	0.348	−0.0001	0.054
	Offset	spin	−0.363	−0.083	0.231	−0.363	−0.0003	0.051	−0.083	−0.0003	0.670
		spoon	−0.265	0.095	0.094	−0.265	−0.0000	0.156	0.095	−0.0000	0.599
		skin	−0.221	0.321	0.061	−0.221	−0.0002	0.271	0.321	−0.0002	0.095
		sin	−0.046	0.017	0.811	−0.046	−0.0003	0.813	0.017	−0.0003	0.929
		soon	0.525	0.455	0.826	0.525	−0.0001	0.032*	0.455	−0.0001	0.013*

Table 3. *(Continued)*

			Groups 1–2			Groups 1–3			Groups 2–3		
			Mean *z*			Mean *z*			Mean *z*		
Moment	Frication	Word	Group 1	Group 2	BLR[a]	Group 1	Group 3	BLR	Group 2	Group 3	BLR
2	Onset	spin	0.381	0.276	0.591	0.381	0.0001	0.049*	0.276	0.0001	0.134
		spoon	0.346	1.188	0.004**	0.346	0.0002	0.081	1.188	0.0002	0.000***
		skin	0.320	1.037	0.012*	0.320	−0.0001	0.097	1.037	−0.0001	0.016*
		sin	0.433	0.159	0.288	0.433	0.0002	0.028*	0.159	0.0002	0.400
		soon	0.431	0.617	0.576	0.431	0.0001	0.036*	0.617	0.0001	0.003**
	Midpoint	spin	0.668	0.050	0.018*	0.668	0.0001	0.002**	0.050	0.0001	0.780
		spoon	0.574	0.114	0.066	0.574	0.0000	0.004**	0.114	0.0536	0.551
		skin	0.414	−0.156	0.050*	0.414	0.0000	0.048*	−0.156	0.0000	0.395
		sin	0.431	0.124	0.171	0.431	0.0000	0.028*	0.124	0.0000	0.485
		soon	0.340	−0.065	0.045*	0.340	−0.0000	0.077	−0.065	−0.0000	0.709
	Offset	spin	0.625	0.799	0.594	0.625	0.0000	0.007**	0.799	0.0000	0.000***
		spoon	0.186	0.721	0.063	0.186	−0.0000	0.329	0.721	−0.0000	0.001***
		skin	0.760	0.459	0.213	0.760	0.0001	0.000***	0.459	0.0001	0.016*
		sin	0.150	0.333	0.363	0.150	0.0002	0.412	0.333	0.0002	0.070
		soon	0.791	0.552	0.464	0.791	0.0000	0.001***	0.552	0.0000	0.006**

[a] Binary logistic regression.
$*p<0.05$; $**p<0.01$; $***p<0.001$.

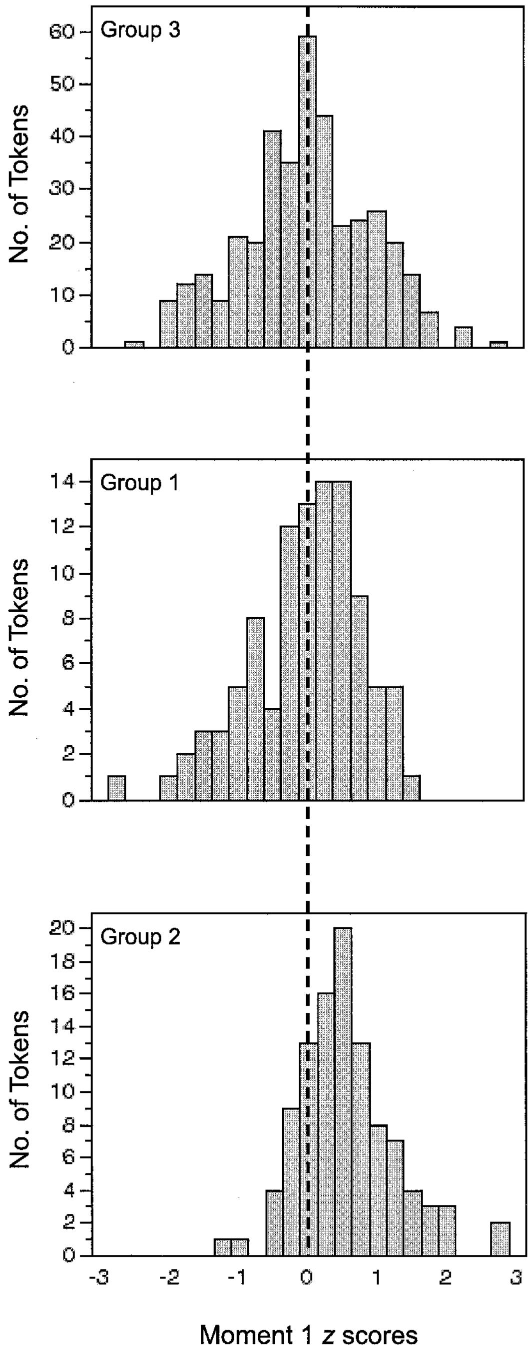

Figure 2. *z score data pooled across* spin, spoon *and* skin *for subjects in each of the three study groups.*

consistent with the perception of decreased sibilance. If such effects occurred for Group 1 speakers in the present study, they were not sufficiently salient to cause the transcriber to classify the /s/ tokens as distorted.

To summarize, two possible articulatory correlates of the acoustic findings are proposed for further study. First, although the /s/ productions of the Group 2 speakers were perceptually transcribed as within the normal range, their M1 values may be consistent with differences in tongue carriage. Second, although the /s/ productions of Group 1 speakers were also perceptually transcribed as within the normal range, their M2 values were consistent with articulatory gestures associated with reduced sibilance and potential differences in constriction shape.

Coarticulation at frication onset and offset

As noted above, statistical differences observed at the frication midpoint did not consistently occur for comparisons involving frication onsets and offsets. Nittrouer, Studdert-Kennedy and McGowan (1989) have suggested that coarticulation decreases as children mature, positing that children initially organize their speech into syllables and only later into phonemic segments (Nittrouer, 1995). Presumably, any contribution of contiguous vowel energy (for onsets in all words and offsets in *sin* and *soon*) could impact both M1 and M2 values. Alternatively, as noted by Jongman (1989), the fricative /s/ reaches its amplitude peak relatively late past frication onset. This may contribute to the possibility of perceptual confusion and productive differences, particularly for /s/ production onsets.

Developmental and normalization correlates

Whatever the articulatory correlates underlying the acoustic differences observed in the present study, explanatory accounts of potentially important differences in speech acquisition and normalization in Group 1 versus Group 2 speakers require consideration. A prior discussion of findings for children with comparable histories of residual /ɝ/ distortions proposed one possible account (Shriberg *et al*., 2001). The prior study reported significant differences in the F3-F2 characteristics of residual derhotacized /ɝ/ productions in adolescents with histories of generalized speech delay, compared with adolescents with speech error histories limited to rhotic distortions. Children in the former group were comparable with Group 1 speakers in the present study, and speakers in the latter group were comparable with the present Group 2 speakers. The children in these studies were assessed with the same protocol used in the present study, and the control speakers were the same children (Group 3) as those in this study. The F3-F2 marker used to discriminate derhotacized /ɝ/ tokens from the two speaker groups had high sensitivity and specificity. Although the present acoustic findings differentiating speakers with the two histories of residual /s/ distortions are not as compelling, the following outlines the basic concepts of a possible explanatory account.

The higher spectral energies for the /s/ productions of speakers in Group 2 suggest that the speakers may have a tongue posture and/or constriction type that is more aberrant than Group 1 speakers relative to typical /s/ production. As suggested for speakers with comparable histories restricted to errors on rhotics, such children might produce distortions on /s/ and possibly on other fricatives and affricates from their very earliest attempts to articulate these sounds. One possible reason that their perceptually correct /s/ productions at adolescence are less like those of typically speaking children is that the early appearing tongue postures were overlearned and hence resistant to change. As proposed for children with histories of only rhotic distortions, relevant explanatory constructs for such persistent speech-motor errors and differences could be drawn from a number of literatures, including diverse perspectives in dynamical systems and second language acquisition (cf. Shriberg *et al*., 2001). Essentially, the thesis proposed in this account is that it is difficult to modify frequently occurring inappropriate behaviours acquired early, at a time when a system is less mature and is self-organizing phonetically and phonologically.

In contrast to the above hypothesis concerning the antecedents of /s/ production in speakers with Group 2 histories, consider the speech histories of children with significant speech delay as described in many sources in the last 30 years. Children with mild to severe speech delay typically have early deletions of /s/ and other fricatives, particularly deletion of /s/ in word-initial clusters (McLeod, 1999). They also have substitutions of stops for /s/, particularly in word-initial singleton contexts. Thus, in contrast to the early /s/ distortions posited for Group 2 speakers, children with speech delay may not realize /s/ targets (i.e. they may not be present in their phonetic inventories) until a later age. By that point in time, children's phonological systems and speech-motor maturity may be characterized by essentially normal tongue postures for /s/ and other fricatives. As indicated by the term *phonological disorder*, the problem in speech delay has been more associated with cognitive-linguistic constraints than with deficits involving speech-motor control. Some classification systems, in fact, reserve the term *phonological disorder* for children with Group 1 histories, using *articulation disorder* to classify children with Group 2 histories. Although we do not endorse this proposed nosological dichotomy, we do hypothesize significant etiological differences in the two speech histories (cf. Shriberg *et al*., 1997; Shriberg *et al*., 2001).

Conclusions

Findings from the present study support the potential utility of acoustic markers for speech-genetic needs, as well as for other research needs in child speech disorders. From a methodological perspective, our findings suggest that spectral moments analysis provides a useful analytic approach toward the identification of such markers, and that use of word-initial clusters may provide the best motoric context to assess the precision of /s/ articulation. Findings further suggest that a midpoint frication sample may be most sensitive to differences in the spectral characteristics of /s/ tokens perceived as correctly articulated. Further research should include larger, more socio-demographically diverse groups of speakers with well-documented speech histories. A primary need is to assess speakers whose residual dentalized /s/ productions in conversational speech samples also occur under the controlled conditions of speech tasks. As in the F3-F2 marker reported in the prior study of adolescents with residual rhotic distortions, perhaps a derived marker based on data from two or more spectral moments will have sufficient sensitivity and specificity to adequately discriminate the speech histories in question for older speakers with persistent /s/ distortions or perceptually normalized speech. Continued acoustic studies and planned physiologic studies pursue these challenging questions.

Acknowledgements

The authors thank Chad Allen, Roger Brown, Ernest Colantonio, Katherina Hauner, Ray Kent, Paul Milenkovic, Rachel Riely, Gary Weismer, David Wilson and the clinicians in the Madison Metropolitan School District for their assistance. Preparation of the paper was supported by a grant from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, DC00496.

References

BAUM, S. R. and MCNUTT, J. C., 1990, An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, **18**, 51–63.

DANILOFF, R. G., WILCOX, K. and STEPHENS, M. I., 1980, An acoustic-articulatory description of children's defective /s/ productions. *Journal of Communication Disorders*, **13**, 347–363.

DUNN, L. M. and DUNN, L. M., 1981, *Peabody Picture Vocabulary Test—Revised* (Circle Pines: American Guidance Service).

FLETCHER, S. G., 1989, Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research*, **32**, 736–748.

FLIPSEN, P., JR, SHRIBERG, L. D., WEISMER, G., KARLSSON, H. B. and MCSWEENY, J. L., 1999, Acoustic characteristics of /s/ in adolescents. *Journal of Speech, Language, and Hearing Research*, **42**, 663–677.

FLIPSEN, P., JR, SHRIBERG, L. D., WEISMER, G., KARLSSON, H. B. and MCSWEENY, J. L., 2001, Acoustic phenotypes for speech-genetics studies: Reference data for residual /ɝ/ distortions. *Clinical Linguistics and Phonetics*, **15**, 603–630.

FLIPSEN, P., JR, TJADEN, K., WEISMER, G. and KARLSSON, H., 1996, Acoustic analysis protocol. Technical Report No. 4. Phonology Project [www.waisman.wisc.edu/phonology], Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

FORREST, K., WEISMER, G., ELBERT, M. and DINNSEN, D. A., 1994, Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children. *Clinical Linguistics and Phonetics*, **8**, 267–281.

FORREST, K., WEISMER, G., HODGE, M., DINNSEN, D. A. and ELBERT, M., 1990, Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics and Phonetics*, **4**, 327–340.

GIBBON, F. E., 1999, Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech Language and Hearing Research*, **42**, 382–397.

HARDCASTLE, W. J. and MORGAN, R. A., 1982, An instrumental investigation of articulation disorders in children. *British Journal of Disorders of Communication*, **17**, 47–65.

JONGMAN, A., 1989, Duration of frication noise required for identification of English fricatives. *Journal of Acoustical Society of America*, **85**, 1718–1725.

JONGMAN, A., WAYLAND, R. and WONG, S., 2000, Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, **108**, 1252–1263.

KARLSSON, H., 1999, An acoustic description of /s/ productions in two subtypes of child speech disorders. Unpublished Master's thesis, University of Wisconsin, Madison.

KENT, R. D., DEMBOWSKI, J. and LASS, N. J., 1996, The acoustic characteristics of American English. In N. J. Lass (ed.), *Principles of Experimental Phonetics* (St Louis: Mosby), pp. 185–225.

LEWIS, B. A. and SHRIBERG, L. D., 1994, Life span interrelationships among speech, prosody-voice, and nontraditional phonological measures. Miniseminar presented at the annual meeting of the American Speech–Language–Hearing Association, New Orleans, LA, November.

MCGOWAN, R. S. and NITTROUER, S., 1988, Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *Journal of the Acoustical Society of America*, **83**, 229–236.

MCLEOD, S., 1999, Children's acquisition of consonant clusters. Unpublished doctoral dissertation, University of Sydney.

MICCIO, A. W., FORREST, K. and ELBERT, M., 1996, Spectra of voiceless fricatives produced by children with normal and disordered phonologies. In T. Powell (ed.), *Pathologies* *of Speech and Language: Contributions of Clinical Linguistics and Phonetics* (New Orleans: ICPLA), pp. 223–236.

MILENKOVIC, P., 1996, CSpeech, v.4 [computer program] (Madison: University of Wisconsin-Madison, Department of Electrical Engineering).

MILLER, J. L., GROSJEAN, F. and LOMANTO, C., 1984, Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica*, **41**, 215–225.

MORRISON, J. A. and SHRIBERG, L. D., 1992, Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, **35**, 259–273.

NARAYANAN, S. S., ALWAN, A. A. and HAKER, K., 1995, An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, **98**, 1325–1347.

NITTROUER, S., 1995, Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, **97**, 520–530.

NITTROUER, S., STUDDERT-KENNEDY, M. and MCGOWAN, R. S., 1989, The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, **32**, 120–132.

SHRIBERG, L. D., 1991, Towards a phenotype for developmental phonological disorders. Invited Paper presented at Genetics: Progress and Promise for Communication Sciences and Disorders, American Speech–Language–Hearing Association, Atlanta, GA, November.

SHRIBERG, L. D., 1993, Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*, **36**, 105–140.

SHRIBERG, L. D., 1994, Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders*, **4**, 38–53.

SHRIBERG, L. D., 1999, Epidemiologic and diagnostic profiles for five developmental phonological disorders. Paper presented at the Annual Convention of the American Speech–Language–Hearing Association, San Francisco, CA, November.

SHRIBERG, L. D., ALLEN, C. T., MCSWEENY, J. L. and WILSON, D. L., 2001, PEPPER: programs to examine phonetic and phonologic evaluation records [computer program] (Madison: Waisman Center, University of Wisconsin-Madison).

SHRIBERG, L. D. and AUSTIN, D., 1998, Comorbidity of speech–language disorder: implications for a phenotype marker for speech delay. In R. Paul (ed.), *The Speech–Language Connection* (Baltimore: Paul H. Brookes), pp. 73–117.

SHRIBERG, L. D., AUSTIN, D., LEWIS, B. A., MCSWEENY, J. L. and WILSON, D. L., 1997, The speech disorders classification system (SDCS): extensions and lifespan reference data. *Journal of Speech and Hearing Research*, **40**, 723–740.

SHRIBERG, L. D., FLIPSEN, P., JR, KARLSSON, H. B. and MCSWEENY, J. L., 2001, Acoustic phenotypes for speech genetics studies: An acoustic marker for residual /ɝ/ distortions. *Clinical Linguistics and Phonetics*, **15**, 631–650.

SHRIBERG, L. D. and KENT, R. D., 1995, *Clinical Phonetics*, 2nd edn (Boston: Allyn & Bacon).

SHRIBERG, L. D. and KWIATKOWSKI, J., 1994, Developmental phonological disorders. I: A clinical profile. *Journal of Speech and Hearing Research*, **37**, 1100–1126.

STEVENS, K. N., 1998, *Acoustic Phonetics* (Cambridge, MA: MIT Press).

STONE, M. and LUNDBERG, A., 1996, Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America*, **99**, 3728–3737.

STONER, R., GATELY, G. and RIVERS, D. L., 1987, An investigation of apical and dorsal /s/ production. *Texas Journal of Audiology and Speech Pathology*, **13**, 40–42.

SUBTELNY, J. D., OYA, N. and SUBTELNY, J. D., 1972, Cineradiographic study of sibilants. *Folia Phoniatrica*, **24**, 30–50.

WEISMER, G. and ELBERT, M., 1982, Temporal characteristics of 'functionally' misarticulated /s/ in 4- to 6-year-old children. *Journal of Speech and Hearing Research*, **25**, 275–287.