CLINICAL LINGUISTICS & PHONETICS, 2001, VOL. 15, NO. 8, 603–630

Acoustic phenotypes for speech-genetics studies: reference data for residual /ɝ/ distortions

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

†Phonology Project, Waisman Center, University of Wisconsin-Madison, USA
‡University of Tennessee, Knoxville, USA
§Speech Motor Control Laboratories, Waisman Center, University of Wisconsin-Madison, USA

(*Received 8 March 2001; accepted 28 May 2001*)

Abstract

An eventual genetic account of at least one subtype of child speech disorders may require the use of acoustic markers to phenotype children and family members. Acoustic markers have the potential for sensitivity and specificity that are not available using auditory-perceptual procedures such as phonetic transcription. Prior reports on the use of acoustics in speech-genetics research have described methods to complete large-scale acoustic analyses, addressed relevant technical and linguistic sampling issues, and provided an acoustic reference database for /s/ production in adolescents. The present paper addresses additional methodological issues and demonstrates how reference data might be used for speech-genetics studies of another class of frequently persisting speech errors—distortions of the English rhotics /r/ and /ɝ/. Findings support the need to adjust acoustic reference data by age and gender, and to subgroup reference data by rhotic phoneme and phonetic context. We describe a *z* score procedure that accommodates these methodological needs. A companion paper uses these data to develop an acoustic phenotype marker of residual /ɝ/ distortions for speech-genetics research.

Keywords: Articulation, genetics, phonology, speech disorders, speech pathology.

Introduction

Measurement needs in speech-genetics research

Emerging research increasingly supports the likelihood that at least one subtype of speech-language disorder of currently unknown origin is genetically transmitted

Address correspondence to: Lawrence D. Shriberg, PhD, 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 © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/02699200110069410

(Matheny and Bruggemann, 1973; Tallal, Ross and Curtiss, 1989; Tomblin, 1989; Lewis and Thompson, 1992; Lewis, Cox and Byard, 1993; Shriberg and Kwiatkowski, 1994; Dixon, Matheny and Mohr, 1995; Felsenfeld, McGue and Broen, 1995; Lahey and Edwards, 1995; Bishop, North and Donlan, 1996; Van Der Lely and Stollwerck, 1996; Felsenfeld and Plomin, 1997; Shriberg, Aram and Kwiatkowski, 1997; Spitz, Tallal, Flax and Benasich, 1997; Fisher, Vargha-Khadem, Watkins, Monaco and Pembrey, 1998; Rice, Haney and Wexler, 1998; Tomblin and Buckwalter, 1998; Lai, Fisher, Hurst, Levy, Hodgson, Jeremiah, Povey, Jamison, Green, Vargha-Khadem and Monaco, 2000). As in other areas of genetics research, the eventual success of research on the genetics of child speech-sound disorders—identification of the genetic loci and modes of transmission of what are currently termed 'developmental phonological disorders'—will require several types of epidemiologic data, including accurate lifespan information on the onset, severity, error patterns, and course of normalization of speech-sound errors as they aggregate in families (cf. Shriberg, 1993; Shriberg, Gruber and Kwiatkowski, 1994; Gruber, 1997; 1999).

For the goals of speech-genetics studies, phenotype research will require investigators to address three interrelated measurement needs that heretofore have not been studied. The primary measurement need is for phenotyping methods that resolve validity and reliability constraints associated with narrow phonetic transcription of speech. Limitations in auditory-perceptual systems for description and classification of segmental and suprasegmental behaviours have been discussed from several perspectives (e.g. Shriberg and Lof, 1991; McSweeny and Shriberg, 1995; Cucchiarini, 1996; Kent, 1996), with reviews underscoring problems of construct validity, examiner reliability, and retest measurement stability. Essentially, there are no auditory-perceptual 'gold standards' for the types of articulatory errors that might be included in phenotype criteria for speech disorders, such as clinical standards to define dentalized, lateralized, derhotacized, labialized, or velarized productions of consonant and vowel sounds. Without such auditory-perceptual standards to guide cross-laboratory research, the validity and reliability of phenotypes that require information on such behaviours as part or all of the relevant inclusionary or exclusionary criteria for the phenotype cannot be assessed or assumed. The premise motivating the present study is that acoustic methods may have the requisite sensitivity and reliability to quantify the types of phonemic and subphonemic information needed for speech-genetics research (cf. Bond and Wilson, 1980; Maxwell and Weismer, 1982; Weismer, 1984; Hewlett, 1988; Forrest, Weismer, Elbert and Dinnsen, 1994; Gibbon, 2000).

A second methodological need in speech-genetics research is for valid phenotype markers to classify the prior and current speech status of family members whose speech error histories may be unavailable and/or unreliable by recall report (Parlour, 1991; Tomblin, Freese and Records, 1992; Lewis and Freebairn, 1993). Direct assessment of the speech of family members is useful, but may not be sufficient to determine affection status because prior speech errors may have corrected (i.e. false negatives). Central to the present concern, moreover, is the situation in which family members may have persisting errors that appear to be similar to the target disorder, but have different etiologic origins (i.e. false positives). Thus, in addition to adequate sensitivity and reliability, phenotype markers must have the requisite specificity to discriminate speakers with the target disorder from speakers with a similar appearing, but different disorder.

On the working assumption that acoustic methods may have the sensitivity, specificity and reliability needed to identify and classify probands and affected relatives in speech-genetics research, the third methodological need is for lifespan acoustic reference data. As in all genetic epidemiologic research, lifespan reference data are needed in speech-genetics studies to define the range of both typical and atypical behaviour for speakers of each sex at all ages. Adolescent-age data are especially needed to cover the extended period during which speakers may finally normalize early and persistent speech-sound distortions. Epidemiologic data from prior work suggest that this time period might extend from approximately 9 years of age until 12 years of age or later (Shriberg and Austin, 1998). Adolescent-aged acoustic reference data have recently been reported for /s/ productions (Flipsen, Shriberg, Weismer, Karlsson and McSweeny, 1999). The Flipsen *et al.* (1999) report provides acoustic reference data for typical /s/ production in 26 9–15 year-old children, including descriptive statistical data obtained from moments analysis. Associated research (Karlsson, 1999; Karlsson, Shriberg, Flipsen and McSweeny, 2001) provides preliminary support for a possible acoustic marker for the subtype of residual /s/ distortions posited to be genetically transmitted. The present paper addresses the need in speech-genetics research to characterize another class of sounds that are frequently in error in clinical populations, US English rhotics.

Acoustic characteristics of English rhotics

The US English consonant /r/, and vowels /ɝ/ and /ɚ/, are among the last of the phonemes to be acquired by children and are among a small set of speech sounds (together with /s/, /z/ and /l/) that may persist as residual distortion errors in adolescents and adults (Shriberg, 1993; 1999; Lewis and Shriberg, 1994). Findings from the 14 acoustic studies of US English rhotics since approximately the middle of the past century (Peterson and Barney, 1952; Angelocci, Kopp and Holbrook, 1964; Lehiste, 1964; Klein, 1971; Dalston, 1975; Lindau, 1985; Chaney, 1988; Espy-Wilson, 1992; Weismer, Martin, Kent and Kent, 1992; Hagiwara, 1995; Hillenbrand, Getty, Clark and Wheeler, 1995; Westbury, Hashi and Lindstrom, 1998; Guenther, Espy-Wilson, Boyce, Matthies, Zandipour and Perkell, 1999; Lee, Potamianos and Narayanan, 1999; Espy-Wilson, Boyce, Jackson, Narayanan and Alwan, 2000) are summarized in table 1. Examination of table 1 indicates a number of methodological differences across studies (highlighted below) that constrain consensus on the acoustic characteristics of English rhotics.

Speaker sample

Beginning with the first two columns in table 1, the external validity of findings in the 14 studies is limited by sample size, with only five studies including more than 20 speakers (Peterson and Barney, 1952; Klein, 1971; Hillenbrand *et al.*, 1995; Westbury *et al.*, 1998; Lee *et al.*, 1999). Although the age range in these studies spans 2 to 80 years, no one study provides data for the entire period of speech acquisition. The Lee *et al.* data do not extend to preschool children; information for older adults is available only in Weismer *et al.* (1992). Data are available for each sex in several studies that included adults (e.g. Peterson and Barney, 1952; Hagiwara, 1995), however, only Lee *et al.* provide child data by sex.

Table 1. *Major design variables in studies reporting data on the acoustic description of /r/ and /ɝ/*

Study	Speaker sample: Groups[a]	Speaker sample: Ages[b]	Linguistic sample: Task	Linguistic sample: Target and environment	Acoustic variables
Peterson and Barney (1952)	33 men 28 women 15 children (SNR)	ANR ANR ANR	Isolated productions of *heard*	/ɝ/	f_0, F1, F2 and F3 frequency and amplitude
Lehiste (1964)	5 men	ANR	Single words in sentence frame	Initial and final /r/ (singleton), /ɝ/ and /ɚ/	F1, F2 and F3 frequency
Angelocci *et al*. (1964)	18 adolescent boys	11–14 yrs	Reading sentence containing *heard*	/ɝ/	f_0, F1, F2 and F3 frequency and amplitude
Klein (1971)	24 children (SNR)	2–4 yrs	Imitation of single words	Initial /r/ (singleton)	F2 and F3 transition onset and offset frequency; F2 transition duration and rate
Dalston (1975)	3 men 2 women 6 boys 4 girls	19–27 yrs 18 yrs 3–4 yrs 3–4 yrs	Single words (reading or picture naming)	Initial /r/ (singleton)	Duration and frequency of F1, F2 and F3; transition duration and rate; F2/F1 and F3/F1 ratio
Chaney (1988)	2 boys 2 girls	4–5 yrs overall	Single words in sentence frame (picture stimuli)	Initial /r/ (singleton and cluster)	F1, F2 and F3 frequencies
Espy-Wilson (1992)	2 men 2 women	ANR ANR	Single words in structured phrases	Initial (singleton and cluster) and final/r/	F1, F2, F3 and F4 frequency (linear and bark); F1, F2 and F3 transition rates

Table 1. *(Continued)*

Study	Speaker sample		Linguistic sample		
	Groups[a]	Ages[b]	Task	Target and environment	Acoustic variables
Weismer *et al.* (1992)	15 men	68–80 yrs	Isolated production of *row*	Initial /r/	Onset frequency, duration, and frequency change of F1 and F2 transitions
Hagiwara (1995)	6 men 9 women	18–26 yrs overall	Single words in a sentence frame	Initial and final /r/ (singletons) and /ɝ/	F1, F2 and F3 (linear and bark) frequency
Hillenbrand *et al.* (1995)	45 men 48 women 27 boys 19 girls	ANR ANR 10–12 yrs 10–12 yrs	Isolated production of *heard*	/ɝ/	Duration and frequency of f_0, F1, F2, F3 and F4
Westbury *et al.* (1998)	23 men 30 women	18–37 yrs overall	Isolated production of *row* and *right*	Initial (singleton) /r/	F1, F2 and F3 frequency
Guenther *et al.* (1999)	3 men 2 women	ANR ANR	Nonsense words in a structured phrase	Medial (singleton and cluster) /r/	F3 frequency
Lee *et al.* (1999)	29 men 27 women 229 boys 207 girls	25–50 yrs 25–50 yrs 5–18 yrs 5–18 yrs	Production of *bird* in a structured phrase	/ɝ/	f_0, F1, F2 and F3 frequency
Espy-Wilson *et al.* (2000)	2 men 2 women	ANR	Single words in a sentence frame	Initial /r/, /ɝ/ and /ɚ/[c]	F1, F2, F3 and F4 frequency

[a] SNR = sex not reported.
[b] ANR = ages not reported.
[c] Data presented were collapsed across all contexts.

Linguistic sample

The entries in the middle columns in table 1 indicate the variety of speech tasks and targets sampled in the 14 acoustic studies. All studies evoked production of English words in structured tasks to control for linguistic context. In six of the 14 studies the rhotic sound was embedded in single words spoken in isolation (Peterson and Barney, 1952; Klein, 1971; Dalston, 1975; Weismer *et al.*, 1992; Hillenbrand *et al.*, 1995; Westbury *et al.*, 1998), which limits generalization of findings to rhotic sounds as they may be produced in decontextualized speech.

The entries in table 1 also indicate that few acoustic data are available on rhotic sounds in differing linguistic environments. Data on final /r/ are provided in three studies (Lehiste, 1964; Espy-Wilson, 1992; Hagiwara, 1995), and these data are limited to adult speakers. Data on cluster contexts are available in three studies (Chaney, 1988; Espy-Wilson, 1992; Guenther *et al.*, 1999), with the Chaney data the sole source of information on /r/ clusters in children. Few studies have examined more than a single context for the same set of speakers, making it difficult to construct across-context comparisons.

Acoustic variables

The studies listed in table 1 have used two approaches to characterize the acoustics of English rhotics: description of the formant frequencies obtained at some point within the sound (11 of 14 studies) and/or description of the transitions between the rhotic sound and adjoining segments (five of 14 studies). Procedural differences within each approach make it difficult to compare data on the same acoustic variable. For example, investigators describing the formant frequencies for rhotic sounds have reported measurements based on the 'steady-state portion' (Peterson and Barney, 1952; Dalston, 1975; Hagiwara, 1995; Hillenbrand *et al.*, 1995; Lee *et al.*, 1999), 'onset of the steady-state' (Angelocci *et al.*, 1964), the point at which 'F2 and F3 are the closest to each other' (Lehiste, 1964), the point at which 'F3 is the lowest' (Espy-Wilson, 1992; Guenther *et al.*, 1999; Espy-Wilson *et al.*, 2000), the 'centre of the formant band' (Chaney, 1988), and the 'onset of phonation' (Westbury *et al.*, 1998). It is unclear whether such alternative measurement loci yield comparable formant data.

Variability in procedural conventions in acoustics studies has also been evident in studies examining formant transitions. Acoustic variables have included transition durations, transition extents, and rates of formant frequency change—either into or out of /r/ or /ɝ/. Klein (1971) reported a single value for transition duration, presumably assuming that formant changes occur in unison, but most investigators have tended to report separate values for each formant. Espy-Wilson (1992) reported the difference in formant frequencies between /r/ and adjacent vowels, whereas most studies use the onset and offset frequencies of the transition itself to calculate frequency changes. Chaney (1988) discussed F2 transitions but reported only group differences, not mean values for her groups. Studies also differ in the data available for each formant. Specifically, Espy-Wilson (1992), Hillenbrand *et al.* (1995) and Espy-Wilson *et al.* (2000) included data for F1, F2, F3 and F4; 11 of the 14 studies provided data on F1, F2 and F3; Weismer *et al.* (1992) reported data for F1 and F2; Klein (1971) reported data for F2 and F3; and Guenther *et al.* (1999) reported data for F3.

As indicated in the right-most column in table 1, all investigators have reported linear (untransformed) frequency data; Espy-Wilson (1992) and Hagiwara (1995) also reported Bark-transformed formant values. The use of Bark transformations was presumably motivated by the possibility that the human auditory system may be more sensitive to relationships among formants than to absolute formant values, and in particular, to differences on a Bark scale (Syrdal and Gopal, 1986). Both Espy-Wilson (1992) and Klein (1971) reported untransformed differences between formants. Dalston (1975) reported untransformed formant ratios in an attempt to allow for comparisons between child and adult data.

Goals of the current study

Acoustic correlates of age and sex during adolescence

The current report attempts to add to our understanding of the English rhotics for the purposes of speech-genetics research by addressing two questions that are especially relevant to child speech-sound disorders. The first is the question of potential acoustic differences associated with age and sex during the adolescent period. The period for normal speech-sound acquisition is thought to have its endpoint at approximately 9 years of age, but this period may be temporally shifted upwards for some individuals with speech-sound disorders (Shriberg *et al.*, 1994). For the latter children, the process of developing fully normal segmental production may continue well into the adolescent period.

It is reasonable to expect that modifications in the size and shape of the vocal tract that accompany puberty (Kent and Vorperian, 1995) may manifest acoustically as differences associated with both age and sex. The report cited previously (Flipsen *et al.*, 1999) found significant sex differences in the acoustics of /s/ in adolescent speakers. The resonant nature of the English rhotics suggests that the need to control for age and sex differences in vocal tract structure may be especially crucial. Three of the studies in table 1 included adolescent speakers (Angelocci *et al.*, 1964; Hillenbrand *et al.*, 1995; Lee *et al.*, 1999), and only Lee *et al.* provide data spanning the entire period of adolescence. Consistent with expectations associated with the physiological changes during development, Lee *et al.* reported significant age trends (decreasing frequency values with increasing age) and different patterns of change for each sex.

Acoustic correlates of phonetic context

The second question to be addressed in the current study concerns the effects of linguistic context on the production of English rhotics. Of the three studies in table 1 that included adolescents (Angelocci *et al.*, 1964; Hillenbrand *et al.*, 1995; Lee *et al.*, 1999), each limited the analysis to productions of vocalic /ɝ/ from a single target word. Reports that have examined across-context variations in other age groups have suggested significant differences across the various English rhotic forms. Lehiste, for example, reported differences among at least three rhotic variants (initial, vocalic, final) noting, for example, that '...the formant frequencies [F1, F2, F3] of initial allophones of /r/ were consistently lower for every speaker than those of any other potential variant' (p. 58). Hagiwara (1995) reported similar findings for F1 and F2 but not for F3. Olive, Greenwood and Coleman (1993, p. 225) noted distinctly lower values for F3 in prevocalic versus postvocalic /r/ (what they termed 'dark' and 'light' /r/, respectively).

A previous report (Flipsen *et al.*, 1999) indicated a need to retain acoustic data for /s/ separately by phonetic context. Relative to rhotic sounds, variations across linguistic contexts may have particular implications for the identification of disorder in speakers of non-standard dialects. For example, speakers of African-American vernacular English (AAVE) typically neutralize vocalic /ɚ/ to /ə/, but use standard English versions of initial and medial consonantal /r/. Accordingly, samples of rhotic productions from speakers of AAVE should only be obtained from prevocalic and intervocalic contexts (Stockman, 1996) to preclude invalid conclusions about the articulatory correctness of AAVE speech samples pooled across all contexts.

Method

Current study

Participants

The primary data for this study were obtained from speech tokens produced by 26 typically speaking children (14 males, 12 females), who ranged in age from 9;7 to 15;2 (years;months), with a mean age of 12;4 (SD = 1;9). There was no significant age difference between female and male speakers, $t(22) = 0.01$, $p = 0.99$. Details on the recruitment procedures and information on the participants are provided in the prior report on the acoustics of /s/ (Flipsen *et al.*, 1999). Essentially these were typically speaking children nominated by teachers as classroom-matched controls for adolescent speakers with residual speech-sound disorders. All speakers had a Midwestern English dialect.

The speech task

A speech production task to assess production of /r/, /ɝ/, and /s/ was administered to each of the 26 children as part of a 90-minute assessment battery. Stimuli for the speech task consisted of five randomized lists of 24 words (120 total tokens) produced in the carrier phrase 'Say ____ again'. Words were presented live by the examiner, who read from a typed list; the child could not see the list or the examiner's face during this task. Speakers were asked to repeat the target in the carrier phrase while maintaining loudness within a preset range as indicated by the VU meter on the tape-recorder. The examiner monitored the participants' alertness and performance and asked children to repeat a phrase if the target appeared not to be understood, was produced incorrectly, or contained obvious interword pauses or dysfluencies.

The present study involved acoustic analysis of the 60 items in the speech task that contained rhotic sounds. These 60 tokens consisted of five repetitions each of the following 12 words: *bird*, *burg*, *burr*, *ride*, *rude*, *rebel* (noun), *rebel* (verb), *pried*, *cried*, *tried*, *crude* and *prude*. These 12 words sampled consonantal /r/ or vocalic /ɝ/ in four canonical forms, four word positions, and in several consonant and vowel contexts.

Acoustic analysis

Acoustic analyses of participants' responses to the speech task were accomplished by two trained research assistants, each of whom was randomly assigned to half of the participants. The assistants, who had completed a course in speech acoustics, followed a well-developed protocol for the analyses (Flipsen, Tjaden, Weismer and Karlsson, 1996). Using a Sony 5000EV tape-recorder as the input source, tokens were digitized using a Sound Blaster AWE32 PNP A/D soundcard connected to a Pentium-based PC. The signal was sampled at 22 kHz with 15 bits of quantization, a stop-band attenuation of −72 dB, and low-pass filtered at 9.8 kHz using the record utility of the software program CSpeech (Milenkovic, 1996). The target word in the interval from the start of /eɪ/ in 'say' to the closure for /g/ in 'again' was isolated and stored. Words were eliminated during digitization if they were mispronounced or incomplete, included extraneous paralinguistic (laughs, yawns) or linguistic (epenthetic consonant or vowel) sounds, included dysfluencies, or obvious interword pauses, or if the rhotics had insufficient formant energies for analysis needs. Pauses, defined as any period of silence 250 ms or longer (Miller, Grosjean and Lomanto, 1984), were measured from the wide-band spectrograms generated with a bandwidth of 500 Hz. To ensure that there was sufficient acoustic energy present in both F2 and F3 of /r/ and /ɝ/, tokens were evaluated during both digitization and subsequent measurements and rejected if both formants could not be reliably tracked throughout their entire duration from the preceding segment to the following segment.

Of 1,560 possible tokens (60 tokens × 26 subjects), 281 (18%) met one or more of the exclusionary criteria and were rejected. To be included in the statistical analyses, speakers needed to produce at least three acceptable tokens for that word because the token values for each speaker were to be averaged for each word. The yield after all exclusions was 1,216 (78%) useable tokens. Token loss was most frequently due to insufficient energy present in F3, a problem reported by other investigators (e.g. Hoffman, Stager and Daniloff, 1983; Huer, 1989). With one exception, the exclusions did not result in large per-speaker or per-target losses. The exception was for the word *rebel* (noun) in which only six male and eight female speakers produced at least three useable tokens. Because only 56 useable tokens (43% of the intended tokens) were available for *rebel* (noun), findings for this target word should be viewed with some caution.

For all 12 words, the formant frequencies of F1, F2 and F3 were calculated for /r/ and /ɝ/within the constriction interval (the region where F2 and F3 are closest together; the same locus was used by Lehiste, 1964). Within the constriction interval, the flat portion was identified visually on a spectrogram generated using a 500 Hz bandwidth, and both formant values were measured at the same point in time. When there was no flat portion (e.g. in /r/ cluster contexts in which the formants might rise immediately after the burst release of the stop), the point in time where F3 was lowest was used to make the measurements for all three formants. The frequencies were identified by first isolating the centre 20 ms of the flat portion (or a 20 ms window centred at the low point of F3 in the case of cluster contexts). CSpeech was then used to compute an LPC (Linear Predictive Coding) spectrum with 24 coefficients. Initially, only F2 and F3 were measured because data on F1 were not viewed as crucial to /r/ or /ɝ/ (i.e. F1 cannot be used to distinguish /r/ from other semivowels or vowels; Dalston, 1975). For completeness, however, information on F1 was obtained several months after the initial period of data reduction. By that time, one of the original research assistants was no longer available; therefore, all F1 measurements were made by a single assistant.

The Lee et al. *database*

The Lee *et al.* (1999) data were considered the secondary data source required to address the two goals of the present study. The critical contribution of these data was that the means and standard deviations reported were based on relatively large cell sizes and were reported separately for each sex. The Lee *et al.* database included the stressed rhotic vocalic /ɝ/, obtained on 1–2 tokens of the single target word *bird* produced by each speaker. Note that the word *bird* was also a stimulus item in the current study, which included fewer participants but a larger corpus of target words. In addition to a common target word, the study by Lee *et al.* and the current study were both based on data obtained using the similarly structured-phrase tasks, 'I say a ____ again' and 'Say ____ again', respectively.

Statistical design

Potential effects of physiological changes during puberty on the acoustics of the two English rhotic sounds were examined using three approaches. First, the formant frequency data were examined to identify possible age trends and sex differences. Second, relative relationships (differences and ratios) between F2 and F3 were tested to determine if they provided adequate control for age trends and sex differences. Formant difference (F3-F2) was used because Klein (1971) noted that '...the difference between the F2 and F3 origin discriminate reliably between good and poor attempts [at /r/]...' (p. 548). Formant ratio (F3/F2) was used because Dalston (1975) has suggested that ratios of the formant frequencies to each other might allow for comparisons between child and adult productions. The third approach was to control for observed age trends and sex differences by the use of z scores derived from data available from the study by Lee *et al.* (1999).

Consistent with the approach used in 10 of the 14 studies in table 1, acoustic analysis was conducted on data from three formant frequencies (F1, F2 and F3) obtained within the rhotic sound. Although 5 of the 14 studies in table 1 included analysis of formant transitions between the rhotic target and adjacent segments (Klein, 1971; Dalston, 1975; Chaney, 1988; Espy-Wilson, 1992; Weismer *et al.*, 1992), the current focus was on the independent characterization of rhotic sounds. Transitions were examined in this dataset but are not reported here. It should also be noted that although Bark transformations have been shown to help distinguish among the semivowels (Espy-Wilson), they were not used for the present study because Hagiwara (1995) has demonstrated that such transformations provide no advantage over untransformed values in normalizing sex differences.

Statistical comparisons in the current study were made with significance set at a liberal α level of 0.05, rather than a lower Bonferroni corrected value. The exploratory nature of the current study suggested that the need to minimize Type II errors was as important as the need to minimize Type I errors. Except as indicated, cell sizes in the current study limited analyses to univariate comparisons. Note also that to avoid mixing across-token variance with across-speaker variance, all analyses used one averaged value for each speaker.[1]

Reliability estimates

First estimate. Reliability estimates for the acoustic measures were completed as part of a larger project that included a total of 122 speakers and were carried out in two phases. Table 2 includes a summary of the findings. Immediately following the acoustic analysis, each assistant reanalysed 24 (3.9%) of the tokens from a randomly selected 2 (15.4%) of the approximately 13 speakers they had each analysed. Interjudge reliability was also estimated by having each assistant analyse 24 (3.9%) of the tokens randomly selected from two (15.4%) of the speakers originally measured by the other assistant. A total of 48 (3.9%) of all the tokens were

Table 2. *Reliability estimates for the acoustic measurements*

Type	First estimate								Second Estimate							
			Mean differences (KHz)			% differences					Mean differences (KHz)			% differences		
	No. of tokens	% of tokens	F1	F2	F3	F1	F2	F3	No. of tokens	% of tokens	F1	F2	F3	F1	F2	F3
Interjudge																
	48	3.9	—	34.0	61.3	—	2.2	2.8	72	5.9	—	39.4	50.0	—	2.5	2.1
Intrajudge																
Assistant 1	24	3.9	14.6	22.4	31.3	2.9	1.4	1.4	72	5.9	27.0	33.3	43.0	5.1	2.0	1.7
Assistant 2	24	3.9	—	14.3	36.5	—	1.0	1.9								

reanalysed. The tokens for both intra- and interjudge estimates included one randomly selected token from each of the 12 target words. Remeasurements were made using copies of the originally digitized files. As shown in table 2, intrajudge and interjudge agreement ranged from differences of 1.0% to 2.8% across the three formants (recall that F1 data were obtained by only one assistant).

Second estimate. Four years after completion of the acoustic analysis, a second reliability estimate was obtained to increase precision, although the second assistant was no longer available. The first assistant remeasured the first two tokens (40%) of each of the 12 target words from 12 of 122 (9.8%) of the speakers, six of whom had been originally analysed by each assistant. The sample included three of the 26 (11.5%) speakers used in the current study resulting in a remeasurement of 72 of the 1,216 (5.9%) tokens used in the current study. As shown in table 2, intrajudge agreement over the 4-year time span was within 5.1% of the original values for F1 and 2.0% and 1.7% of the original values for F2 and F3, respectively. Remeasurement differences for F2 and F3 were 2.5% and 2.1%, respectively. Across the 12 target words, difference values ranged from 0.9% to 4.7% (F2) and 1.1% to 3.3% (F3) of the original mean frequencies.

Reliability and measurement notes. Four of the 14 studies in table 1 provide reliability estimates (Peterson and Barney, 1952; Chaney, 1988; Hillenbrand *et al.*, 1995; Lee *et al.*, 1999), but the only data that are methodologically comparable to those obtained in the present study are those of Hillenbrand *et al.* These authors obtained interjudge mean differences between original and remeasurements of 25.2 and 28.7 Hz for F2 and F3, respectively, which reportedly represented differences of 1.0% to 2.0% from the original measurements. Comparisons of these values to interjudge data obtained in the first estimate for the current study suggests some possible concern for the reliability of the F3 values in the current study. Possible sources of the slightly lowered agreement in the present study include linguistic context and analysis methods. The phrases used in the current study likely resulted in shorter segment duration than those obtained by Hillenbrand *et al.* who used single words spoken in isolation. This difference would be further compounded by the size of the analysis window: Hillenbrand *et al.* used the entire steady-state portion of the rhotic whereas the present study used only the middle 20 ms. In any case, the mean difference values in the current study for F3 were less than 4% of the originally measured values, which is viewed as support for the measurement stability of the present data.

Results

Validation of the Lee et al. (1999) *data for the current study*

The first methodological need was to ensure that the speakers in the current study were sufficiently similar to those of Lee *et al.* (1999) to warrant use of the Lee *et al.* data to generate z scores for reference data. As described previously, Lee *et al.* included data for a common target word (*bird*) and both studies used similar speech sampling procedures. For the present purposes, S. Lee provided the authors unpublished individual token data for the word *bird*. A series of three t-tests were completed on F1, F2 and F3. The tests compared per speaker average values for the 26 speakers in the current study with per speaker average values for the 222 speakers from Lee *et al.* who were 9–15 years of age. Values for F2 and F3 were not significantly different from those reported in Lee *et al.* ($t(24)=0.18$ and $t(26)=0.42$, respectively, $p>0.05$), but values for F1 were significantly different ($t(31)=-7.66$, $p=0.000$).

Differences in the acoustic analysis methods used in Lee *et al.* (1999) compared to the current study likely accounted for at least some of the obtained differences for F1. Lee *et al.* used a fully automated procedure to estimate formant frequencies compared to the manual approach used in the current study. It is possible that there was a confounding of the signal in the current study due to the filter bandwidth chosen, resulting in a perceptual problem distinguishing F1 from f_0. The lower mean F1 value obtained in the current analysis (517 Hz vs. 600 Hz for Lee *et al.*) is consistent with this possibility; the corresponding mean values for F2 and F3 were much closer (1607 vs. 1599 Hz and 2129 vs. 2106 Hz, respectively). The fact that F1 has been demonstrated to be of limited value in the analysis of US English rhotics (Dalston, 1975; Espy-Wilson, 1992), combined with the finding of no significant differences for either F2 or F3, suggested that the formant data from the current study sample were sufficiently consistent with those of Lee *et al.* to consider using the Lee *et al.* data for the current purposes. Given these concerns about the F1 data, however, subsequent analyses were completed only for F2 and F3.

Question 1: Are acoustic characteristics of English rhotics associated with age and sex during the adolescent period?

Although Lee *et al.* (1999) observed age trends and sex differences in their original study, it was not clear whether these trends and differences occurred during the period of adolescence (i.e. 9–15 years, rather than the extended period of 5–18 years they examined). To address this question, data for 9–15-year olds from the data provided by Lee *et al.* for the target word *bird* were examined statistically. Using a single average value per speaker ($n=222$), ANOVAs were conducted using a General Linear Model testing age and sex as main effects. Age and sex were significant main effects for both F2 (age [$F(6, 214)=9.63$, $p=0.000$]; sex [$F(1, 214)=68.58$, $p=0.000$]) and F3 (age [$F(6, 214)=11.64$, $p=0.000$]; sex [$F(1, 214)=47.21$, $p=0.000$]).

To examine age trends in the current study, a series of linear regressions was conducted on the formant frequency data for each of the 12 target words. Because Anderson-Darling tests of the normality of the distributions of formant values for each word indicated several non-normal distributions ($p<0.05$), the data for all acoustic variables were rank-transformed prior to computing the regressions of each variable on age (Conover and Iman, 1981). Consistent with the above findings for the Lee *et al.* (1999) data, significant linear trends ($p<0.05$) were observed for all 12 words for F2, and 10 out of 12 words for F3. The regressions were all in the negative direction (i.e. formant frequencies decreased over the age range), with the significant regressions accounting for 19–48% common variance. To examine sex differences, a series of Wilcoxon-Mann-Whitney tests were carried out on each of the 12 target words. Again, consistent with the findings of Lee *et al.*, statistically significant differences ($p<0.05$) for sex of speaker were obtained for six out of 12 words for F2 and eight out of 12 words for F3. These findings confirmed the need for acoustic data on US English rhotics to be controlled for both age and sex differences in this age range, as described in the following subsections.

Derived variables. To examine the possibility that relative differences between F2 and F3 may control for absolute differences associated with age and sex, the above analyses were repeated using F3−F2 (i.e., F3 *minus* F2) and F3/F2 (i.e., F3 *divided by* F2). Using data from each eligible token, speaker averages on each of these derived variables were generated for each of the 12 target words. Relative to age trends, significant linear regressions were obtained on age for only one out of 12 words (*burg*) for F3−F2 and none of the 12 words for F3/F2. Relative to sex, significant differences were obtained for only one of the 12 words (*burg*) for F3−F2 and none of the 12 words for F3/F2.

These findings suggested that the use of either F3−F2 or F3/F2 might control for the differences in absolute formant magnitudes associated with age and sex. However, the small sample size of the current dataset raised the possibility that limited statistical power may have constrained potentially significant effects. Using individual token data provided by S. Lee (personal communication), F3−F2 and F3/F2 values were calculated for each token. Means and standard deviations for each age (5–18 years and adults) and sex were calculated and are provided in the Appendix. To be consistent with the prior analysis of the original formants for data from the current study, a single average value for each speaker in the 9–15 year-old range ($n = 222$) was used to conduct ANOVAs on both derived variables using a General Linear Model testing age and sex as main effects. For F3−F2, age ($F(6, 214) = 3.68$, $p = 0.000$) was a significant main effect but sex ($F(1, 214) = 1.83$, $p = 0.178$) was not. For F3/F2, neither age ($F(6, 214) = 1.99$, $p = 0.068$) nor sex ($F(1, 214) = 1.14$, $p = 0.286$) was a significant main effect. These findings suggested that F3/F2 provided statistical control over age trends and sex differences in this age range, whereas F3−F2 did not. However, because the trend in the analysis of F3/F2 for the 9–15 year-olds was in the direction of a main effect for age and because the need for acoustic characterization of rhotic sounds may include the entire developmental period, the analysis was repeated for the entire dataset from Lee *et al.* ($n = 433$). Findings yielded significant main effects for both age ($F(14, 408) = 7.14$, $p = 0.000$) and sex ($F(1, 408) = 6.66$, $p = 0.010$) for F3−F2. For F3/F2, significant main effects were observed for age ($F(14, 408) = 2.64$, $p = 0.001$), but not for sex ($F(1, 408) = 0.15$, $p = 0.695$). Thus a conservative approach for the reference data was adopted in which it was deemed appropriate to provide statistical control for both age and sex.

z scores. The availability of the large reference dataset reported by Lee *et al.* (1999) suggested a potential way to control for age and sex differences during puberty. By normalizing all of the formant frequency values relative to specific age and sex expectations (i.e. generating *z* scores), differences associated with vocal tract structure should be minimized. For each of the 26 speakers in the current study, individual mean values for F2, F3, F3−F2 and F3/F2 in the word *bird* were converted to *z* scores using the means and standard deviations reported by Lee *et al.* as well as the data presented in the Appendix. To be consistent with the procedure used by Lee *et al.*, age at last birthday was used as the basis for identifying the appropriate reference group. The obtained distributions of *z* scores were compared to the corresponding values for *bird* reported for typical speakers in Lee *et al.* (assuming $M = 0$, $SD = 1.0$) using a series of four *t*-tests. Not surprisingly, and consistent with the above comparisons of the original F2 and F3 formant frequencies, none of the distributions was significantly different from those reported in Lee *et al.* for

F2 ($t(22)=0.388$) or F3 ($t(22)=1.043$), or from those in the Appendix, for F3−F2 ($t(22)=-0.967$) or for F3/F2 ($t(22)=-0.637$).

To evaluate whether the conversion to *z* scores provided the needed control for age trends, a series of linear regressions were conducted on *z*F2, *z*F3, *z*F3−F2 and *z*F3/F2 from the target word *bird* on age. None of the regressions was significant. Control for sex differences was evaluated using Wilcoxon-Mann-Whitney tests on each of the four *z* score variables. None of the comparisons was significant.

Conclusion. In response to Question 1, we conclude that there were significant age and sex effects in the original formant data, and that these effects could not be controlled by using the two derived variables (i.e. F3−F2, F3/F2). We suggest that *z* scores based on the Lee *et al.* (1999) age-by-sex data provide the needed standardized scores for studies in which participants include multiaged adolescent children of both sexes.

Question 2: are acoustic characteristics of English rhotics associated with linguistic context?

The second goal of the current study was to assess associations among the production of US English rhotics and linguistic context for speakers in this age range. Using the same information as described above (i.e. the Lee *et al.* (1999) original data and the values in the Appendix), F2 and F3 frequencies and F3−F2 and F3/F2 values were converted to *z* scores for each of the 12 target words produced by each of the 26 speakers in the current study. A series of 48 Wilcoxon-Mann-Whitney tests were conducted comparing male to female distributions of *z*F2, *z*F3, *z*F3−F2 and *z*F3/F2 values for each of the 12 target words. Only one of the comparisons (*z*F3 in *crude*) was statistically significant, which is consistent with chance expectation. This supported a decision to base all subsequent analyses on data pooled across sex. Group mean values for *z* scores for the formant frequencies and derived variables for each of the target words are shown in table 3.

Original formants. A series of pairwise comparisons was carried out using Wilcoxon-Mann-Whitney tests for differences in *z*F2 and *z*F3 for each of the 12 target words with every other target word. Probability outcomes are shown in table 4, with values for *z*F2 in the bottom half of the table (not bold) and values for *z*F3 in the top half of the table (bold). Of 66 comparisons for *z*F2, 35 (53%) were significant ($p<0.05$), and of 66 comparisons for *z*F3, 31 (47%) were significant. In both cases, these values were considerably more than the uncorrected chance expectation of 5% significant comparisons. This suggested that there were reliable differences among the 12 target words relative to their original formant frequencies.

Beginning with the vocalic /ɝ/ targets (*bird*, *burg* and *burr*), there were no significant differences among them on either *z*F2 or *z*F3 (all *p*s>0.05). The vocalic /ɝ/ forms differed significantly from three of the four word-initial singleton /r/ targets (the exception was *rebel* [verb]) for *z*F2, but the pattern was reversed for *z*F3. The vocalic targets were significantly different from the word-initial cluster targets (*pried*, *cried*, *tried*, *crude*, *prude*) on five of the 15 comparisons for *z*F2 (all involving *cried* and *tried*) and on all 15 comparisons for *z*F3.

Three of the four word-initial singleton /r/ targets (*ride*, *rude* and *rebel* [noun]) patterned strongly together. They did not differ from each other on either *z*F2 or

Table 3. *Normalized (z score) descriptive statistics for F2, F3, F3−F2, and F3/F2 across the 12 /r/ and /ɝ/ target words produced by the 26 typically speaking adolescents*

Target word	*n*	*z*F2[a,b]	*z*F3[a,b]	*z*F3−F2[a,c]	*z*F3/F2[a,c]
bird	23	0.11 (1.36)	0.25 (1.15)	.17 (0.82)	0.12 (0.88)
burg	24	−0.13 (1.21)	0.11 (1.14)	0.21 (0.89)	0.20 (0.84)
burr	24	−0.26 (1.45)	−0.21 (1.22)	−0.07 (0.75)	−0.01 0.78)
ride	18	−1.57 (1.96)	−0.01 (1.01)	1.43 (1.25)	2.27 (2.21)
rude	17	−1.70 (1.60)	0.27 (1.77)	1.91 (1.87)	2.78 (2.15)
rebel (noun)	14	−1.53 (1.82)	−0.30 (1.12)	1.90 (1.26)	2.78 (2.07)
rebel (verb)	23	0.48 (1.81)	1.23 (1.46)	1.08 (1.02)	0.92 (1.15)
pried	26	0.41 (1.34)	0.84 (1.14)	0.64 (0.14)	0.44 (0.14)
cried	25	0.81 (1.40)	1.00 (1.13)	0.51 (0.75)	0.22 (0.82)
tried	23	1.19 (1.60)	1.41 (1.42)	0.67 (0.91)	0.25 (0.85)
crude	25	0.07 (1.37)	−1.03 (1.23)	1.23 (1.48)	1.19 (1.66)
prude	26	−0.32 (1.19)	0.89 (1.25)	1.41 (0.25)	1.46 (0.27)

[a] Cell entries are group means (and standard deviations).
[b] *z* scores derived from each speaker's mean formant data using the means and standard deviations (by age and sex) reported by Lee *et al.* (1999) for the target *bird.*
[c] *z* scores derived from each speaker's mean formant data using the means and standard deviations (by age and sex) shown in the Appendix.

*z*F3, but they did differ significantly from all the cluster targets on *z*F2 and on 11 of the 15 comparisons on *z*F3 (the exceptions all involving *rebel* [noun]). The singleton target *rebel* (verb) differed significantly from the other three singleton targets on *z*F2 and from two of the three (the exception being *rebel* [noun]) on *z*F3. The verb form of *rebel* did not differ from any of the cluster targets on either formant.

The cluster targets appeared to form two subgroups based on their vowel nuclei: those containing the /a͡ɪ/ diphthong (*pried*, *tried*, *cried*) did not differ from each other on either formant, and the same was true for those containing the /u/ vowel (*crude*, *prude*). The latter two targets differed significantly from the first three on three of the 10 comparisons for *z*F2 but on none of the comparisons for *z*F3.

Three additional contextual contrasts were also examined. Relative to singleton versus cluster contexts, the target word *ride* differed significantly from *pried*, *cried* and *tried* on both formants. As well, the target word *rude* differed significantly from *prude* and *crude* on both formants. Relative to syllable stress, the target *rebel* (verb) differed significantly from *rebel* (noun) on *z*F3 but not on *z*F2. Finally, relative to the effect of the particular vowel following consonantal /r/ (i.e. /u/ vs. /a͡ɪ/), there were no significant differences on either variable between the pairs *ride*-*rude*, *pried*-*prude*, *cried*-*crude*.

In addition to the group-level comparisons, the frequency and magnitude of individual speaker departures from group means were examined by constructing plots of each of the 26 speakers' *z*F2 by *z*F3 values for each target word (see also Notes). The observed patterns were consistent with the statistical analyses. These plots also supported prior statistical findings for sex, with no apparent trends in the distributions of symbols for male and female speakers in any of the target words (i.e. there did not appear to be any interaction effects for sex by linguistic context).

Derived variables. Pairwise Wilcoxon-Mann-Whitney comparisons were subsequently conducted across the 12 targets using the *z*F3−F2 and *z*F3/F2 data.

Table 4. *Pairwise comparisons of the z score distributions among the 12 /r/ and /ɝ/ targets for F2 (bottom half of table, not bold) and F3 (top half of table, bold)*[a]

	bird	*burg*	*burr*	*ride*	*rude*	*rebel* (noun)	*rebel* (verb)	*pried*	*cried*	*tried*	*crude*	*prude*
bird	—	**n.s.**	**n.s.**	**n.s.**	**n.s.**	**n.s.**	**0.0089**	**0.0400**	**0.0118**	**0.0028**	**0.0108**	**0.0329**
burg	n.s.	—	**n.s.**	**n.s.**	**n.s.**	**n.s.**	**0.0026**	**0.0113**	**0.0035**	**0.0008**	**0.0034**	**0.0129**
burr	n.s.	n.s.	—	**n.s.**	**n.s.**	**n.s.**	**0.0005**	**0.0007**	**0.0003**	**0.0001**	**0.0006**	**0.0010**
ride	0.0020	0.0042	0.0132	—	**n.s.**	**n.s.**	**0.0022**	**0.0103**	**0.0026**	**0.0009**	**0.0022**	**0.0067**
rude	0.0010	0.0027	0.0052	n.s.	–	**n.s.**	**0.0073**	**0.0246**	**0.0071**	**0.0022**	**0.0071**	**0.0230**
rebel (noun)	0.0025	0.0056	0.0148	n.s.	n.s.	—	**n.s.**	**n.s.**	**n.s.**	**0.0152**	**n.s.**	**n.s.**
rebel (verb)	n.s.	n.s.	n.s.	0.0013	0.0006	0.0018	—	**n.s.**	**n.s.**	**n.s.**	**n.s.**	**n.s.**
pried	n.s.	n.s.	n.s.	0.0005	0.0002	0.0009	n.s.	—	**n.s.**	**n.s.**	**n.s.**	**n.s.**
cried	n.s.	0.0178	0.0063	0.0001	0.0000	0.0002	n.s.	n.s.	—	**n.s.**	**n.s.**	**n.s.**
tried	0.0250	0.0060	0.0034	0.0000	0.0000	0.0001	n.s.	n.s.	n.s.	—	**n.s.**	**n.s.**
crude	n.s.	n.s.	n.s.	0.0010	0.0010	0.0034	n.s.	n.s.	n.s.	0.0258	—	**n.s.**
prude	n.s.	n.s.	n.s.	0.0063	0.0045	0.0142	n.s.	n.s.	0.0056	0.0009	n.s.	—

[a] Cell entries are p values from Wilcoxon-Mann-Whitney tests (n.s. $= p > 0.05$).

Results of comparisons are shown in table 5, with *z*F3−F2 probability values (not bold) included in the bottom half of the table and *z*F3/F2 probabilities (bold) in the top half of the table. For *z*F3−F2, 33 out of 66 (50%) of the comparisons were significant ($p<0.05$), and for *z*F3/F2, 38 out of 66 (58%) of the comparisons were significant. Again, this number of significant findings was considerably greater than the percentage of comparisons (5%) that might be expected to be significant by chance.

For *z*F3−F2, the vocalic /ɝ/ forms were not significantly different from each other, but they differed significantly from the word-initial singleton /r/ targets on all 12 comparisons, and from the cluster targets on 10 of the 15 comparisons. None of the word-initial singleton /r/ targets differed from each other, but they differed from the cluster targets on eight of the 20 comparisons (none of the differences involved *prude*, *crude* or *rebel* [verb]). The cluster contexts differed from each other on three of the 10 comparisons (all comparing *prude* with the clusters that included /a͡ɪ/). Relative to singletons versus clusters, *ride* differed significantly from *pried* and *cried* but not from *tried*, and *rude* did not differ from *prude* or *crude*. Relative to syllable stress, *rebel* (verb) did not differ significantly from *rebel* (noun). Finally, relative to effect of the vowel following consonantal /r/, there were no significant differences on the pairs *ride-rude* or *cried-crude*, but *pried* and *prude* were significantly different from each other.

For *z*F3/F2, the vocalic /ɝ/ targets again did not differ significantly from each other, but they did differ significantly from the word-initial singleton /r/ targets on all 12 of the comparisons. The vocalic /ɝ/ targets differed on six of the 15 comparisons with the cluster contexts (all the differences involved *prude* and *crude*). The word-initial singleton /r/ contexts differed from each other on two of the six comparisons (both of which involved *rebel* [verb]) and differed from the cluster contexts on 15 of the 20 comparisons. The cluster contexts differed from each other on three of the 10 comparisons (only on comparisons involving *prude* with clusters that included /a͡ɪ/). Relative to singletons versus clusters, the word *ride* differed significantly from *pried*, *tried* and *cried*, and *rude* differed significantly from *prude* and *crude*. Relative to syllable stress, *rebel* (verb) differed significantly from *rebel* (noun). Finally, relative to the effect of the vowel following consonantal /r/, the pattern was identical to that observed for *z*F3−F2; the pairs *ride-rude* and *cried-crude* were not significantly different, but *pried* and *prude* were significantly different from each other.

Consistent with the analysis of the original formants, plots of individual speaker values for the derived *z*F3−F2 and *z*F3/F2 variables were also constructed to examine the frequency and magnitude of individual speaker departures from group means on each of these analyses. These plots are shown in figure 1 and figure 2, respectively. In both figures, the plots are consistent with the statistical analysis.

Conclusion. In response to Question 2 we conclude that US English rhotics produced by speakers in this age range do differ acoustically across linguistic contexts. Regardless of the analytical approach, four groups of targets patterned together acoustically (i.e. within each group, the targets never differed statistically from one another): (a) vocalic /ɝ/ in *bird*, *burg* and *burr*; (b) word-initial singleton /r/ in *ride*, *rude* and *rebel* (noun); (c) word-initial cluster /r/ in *pried*, *tried* and *cried*; and (d) word-initial cluster /r/ in *prude* and *crude*. Word-initial singleton /r/ in the target *rebel* (verb) did not consistently pattern with any other target. Anecdotally, speakers

Table 5. *Pairwise comparisons of the z score distributions among the 12 targets for F3 − F2 (bottom half of table, not bold) and F3/F2 (top half of table, bold)*[a]

	bird	*burg*	*burr*	*ride*	*rude*	*rebel* (noun)	*rebel* (verb)	*pried*	*cried*	*tried*	*crude*	*prude*
bird	—	**n.s.**	**n.s.**	**0.0004**	**0.0000**	**0.0002**	**0.0130**	**n.s.**	**n.s.**	**n.s.**	**0.0335**	**0.0010**
burg	n.s.	—	**n.s.**	**0.0007**	**0.0000**	**0.0002**	**0.0148**	**n.s.**	**n.s.**	**n.s.**	**0.0466**	**0.0016**
burr	n.s.	n.s.	—	**0.0001**	**0.0000**	**0.0001**	**0.0024**	**n.s.**	**n.s.**	**n.s.**	**0.0091**	**0.0001**
ride	0.0022	0.0028	0.0001	—	**n.s.**	**n.s.**	**n.s.**	**0.0023**	**0.0007**	**0.0022**	**n.s.**	**n.s.**
rude	0.0007	0.0009	0.0001	n.s.	—	**n.s.**	**0.0062**	**0.0001**	**0.0001**	**0.0001**	**0.0160**	**0.0429**
rebel (noun)	0.0002	0.0001	0.0000	n.s.	n.s.	—	**0.0081**	**0.0004**	**0.0002**	**0.0003**	**0.0184**	**n.s.**
rebel (verb)	0.0040	0.0045	0.0002	n.s.	n.s.	n.s.	—	**0.0462**	**0.0186**	**0.0280**	**n.s.**	**n.s.**
pried	n.s.	0.0445	0.0016	0.0489	0.0135	0.0028	n.s.	—	**n.s.**	**n.s.**	**n.s.**	**0.0050**
cried	n.s.	n.s.	0.0168	0.0200	0.0025	0.0026	n.s.	n.s.	—	**n.s.**	**n.s.**	**0.0014**
tried	n.s.	n.s.	0.0064	n.s.	0.0267	0.0056	n.s.	n.s.	n.s.	—	**n.s.**	**0.0029**
crude	0.0118	0.0105	0.0007	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	—	**n.s.**
prude	0.0005	0.0004	0.0000	n.s.	n.s.	n.s.	n.s.	0.0216	0.0132	0.0363	n.s.	—

[a] Cell entries are p values from Wilcoxon-Mann-Whitney tests (n.s. = $p > 0.05$).

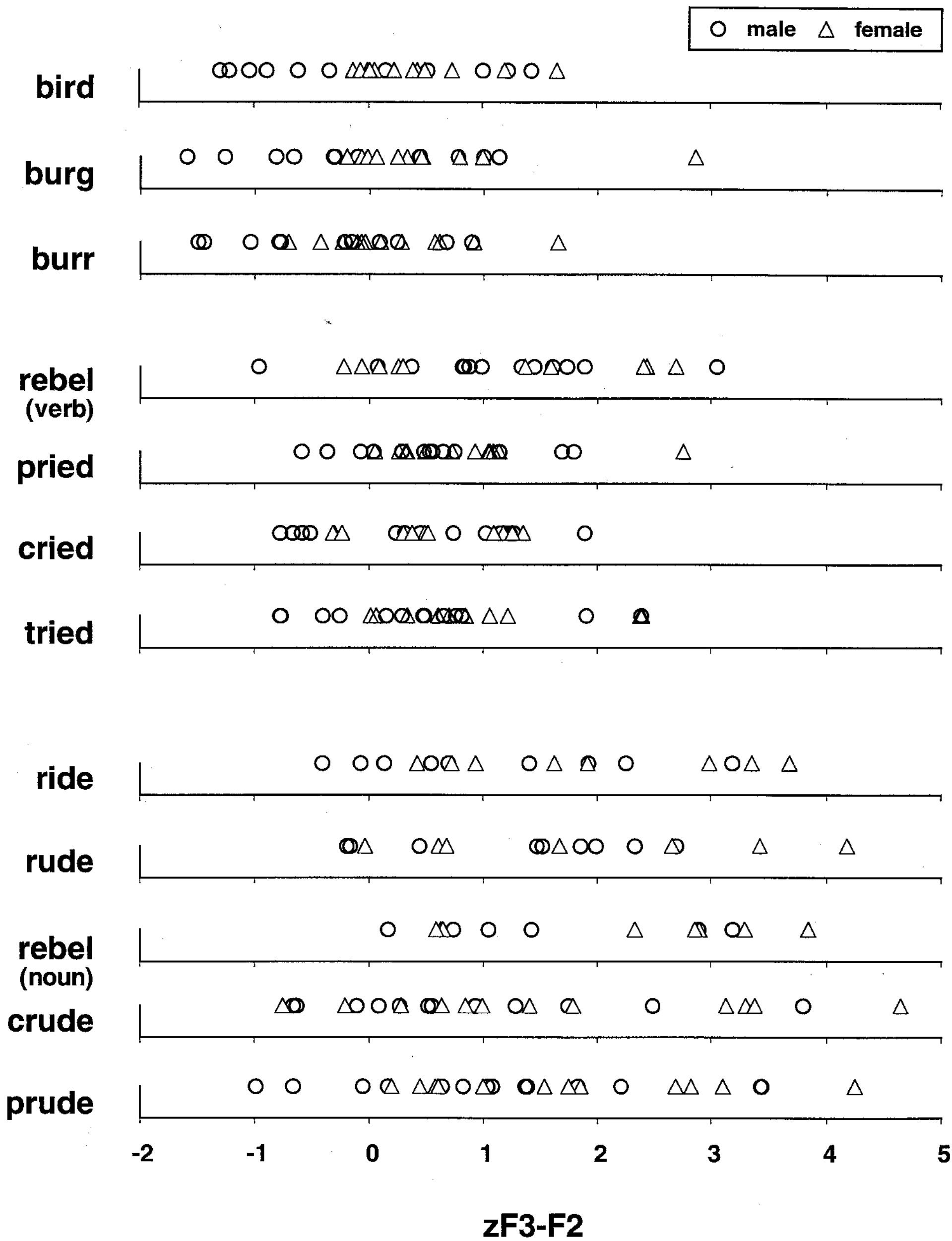

Figure 1. *Normalized F3 − F2 data for /r/ and /ɝ/ in 12 words produced by 26 typically speaking adolescents.*

appeared to have difficulty with *rebel* (verb), with some productions perceived to be articulated as [ɝbɛl].

The findings are consistent with four specific contextual effects. First, the presence or place of a stop consonant following /ɝ/ apparently did not alter the formant frequencies for this target allophone. Second, characteristics of vowels following initial singleton /r/ were not associated with formant differences, whereas vowels

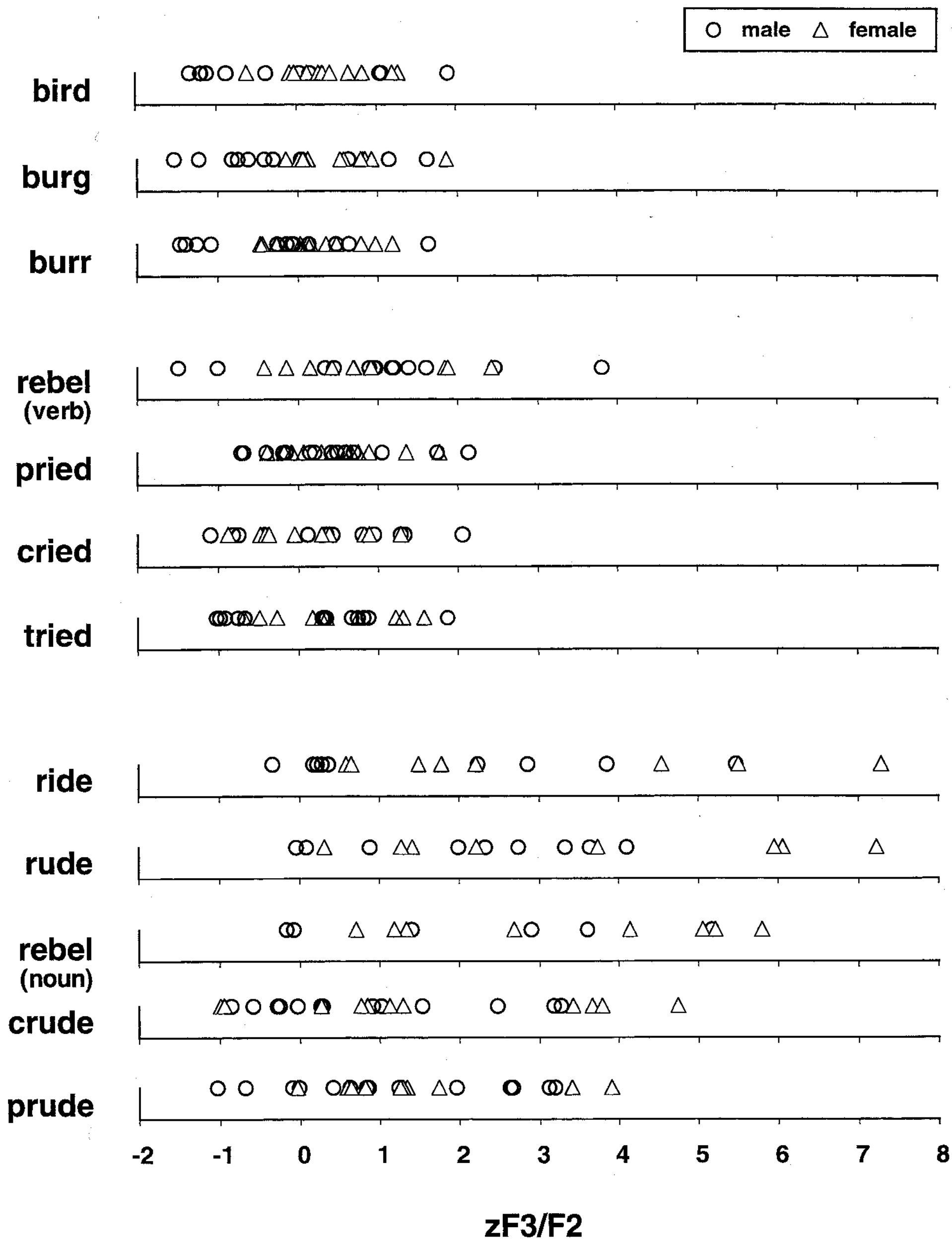

Figure 2. *Normalized F3/F2 data for /r/ and /ɝ/ in 12 words produced by 26 typically speaking adolescents.*

specifically following the initial /pr/ cluster were associated with significant formant differences on the rhotic consonant. Third, singleton /r/ was associated with significantly lower formant frequencies when compared with the corresponding cluster context (e.g. *ride* vs. *cried*, *rude* vs. *prude*), although there did not appear to be significant differences associated with the specific place of articulation of the stop consonant in the cluster contexts. Finally, in the present data, potential effects of syllable stress on word-initial singleton /r/ were inconsistent.

Discussion and conclusion

Speaker normalization

We interpret findings as support for the feasibility of using standard normalization procedures such as *z* scores to accommodate individual differences in acoustic approaches to measurement in speech-genetics studies. The use of standard scores procedures to adjust for physical differences across age and sex is not without precedent. Farrer and Meaney (1985), for example, examined physical deterioration in patients with Huntington's disease using *z* scores on a range of anthropometric measures. Garn, Smith and LaVelle (1984) used *z* scores created from head and face measurements to identify pattern profiles in individuals with dysmorphogenesis syndromes. Similar procedures have also been applied in efforts to more clearly define physical phenotypes in genetic studies (e.g. Cartwright, Kula and Wright, 1999). Most relevant to the present study, Lobanov (1971) used a similar normalization procedure for classifying Russian vowels.

Context effects

The array of findings indicating the need to control for context effects in consonant /r/ data are notably consistent with prior findings summarized in table 1 and warrant comment.

The failure to obtain a difference among the vocalic /ɝ/ targets is consistent with Lehiste (1964), who noted that formant frequencies of /ɝ/ were unaffected by the presence or absence of a following consonant. Hagiwara (1995) did not specifically address the question, but post-hoc analysis using his raw token data suggested no significant difference on either F2 or F3 frequencies between the target words *bert* and *turk*.

Differences observed between vocalic /ɝ/ and word-initial singleton /r/ are most consistent with the findings of those of Hagiwara (1995) who noted significantly lower values in the latter context for F2 but not for F3 (see table 3 to confirm the directions of differences). Lehiste (1964) reported significantly lower frequencies for singleton /r/ for both formants. Differences associated with sex of the speaker may be a source of these findings because both Hagiwara's study and the current study included female speakers, whereas Lehiste's study included only males. Hagiwara reported different patterns on F3 for his male and female speakers; F3 values were lower in word-initial contexts for the male speakers (consistent with Lehiste), but F3 values were higher for the word-initial contexts for the female speakers (and there was no difference between F3 values for vocalic /ɝ/ and word-initial singleton /r/ when the data were pooled across sex). Thus, although not confirmed in the present study, these other findings suggest a possible interaction between sex and linguistic context. It should be noted that sample sizes in Lehiste ($n=5$) and Hagiwara ($n=15$) were smaller than those in the present study, and both studies assessed adults rather than adolescents.

Word-initial singleton versus cluster contexts were not specifically examined in any of the studies in table 1. Chaney (1988) did present mean data on each of her four control speakers for F2 in singleton and cluster contexts. In each comparison, values for the cluster context were higher, which is consistent with the finding of significantly higher values for the cluster contexts in the current study. Espy-Wilson (1992) included clusters in her protocol, but did not compare them with singletons or provide data that could be used to construct a comparison. Guenther *et al.* (1999) presented graphical data for their seven speakers on medial singleton and cluster /r/ productions. Visual examination of the figures in that report suggested a tendency for higher F3 frequencies in clusters compared to singletons (recall that F2 was not examined by Guenther *et al.*).

Comparative findings do not appear to be available on the acoustics of rhotics in stressed versus unstressed syllables. Espy-Wilson (1992) included target words with /r/ in unstressed, prestressed and poststressed positions but did not address findings at this level. It is possible that the inconsistent findings for syllable stress effects obtained in the current study may in part reflect the previously discussed problem of excessive token loss in the target word *rebel* (noun).

The failure to observe effects of a change in the following vowel on the formant frequencies of word-initial singleton /r/ is consistent with the findings of Hagiwara (1995), who included nine different vowel contexts in his analysis, including the present /u/ (*rude*) context, but not /ɑɪ/ (*ride*). Chaney (1988), however, reported a significant main effect for vowels on singleton /r/ for both F2 and F3. It may be noteworthy that the significant findings of Chaney were obtained from the speech of preschool-aged children. The possibility of age effects on the interaction of vowels with preceding segments is supported by the findings of Nittrouer, Studdert-Kennedy and McGowan (1989), who reported greater vowel effects on preceding fricatives in children (ages 3–7 years) compared to adults. The significant effects of vowels on the word-initial cluster contexts in the current study are consistent with Chaney's findings.

Extended discussion of the articulatory correlates of the context findings obtained in the present and prior studies is beyond the scope of this paper. The interested reader is referred to Espy-Wilson *et al.* (2000), Guenther *et al.* (1999) and Westbury *et al.* (1998) for recent detailed discussions.

Conclusion

The findings of this study indicate that comparative acoustic normalization of the rhotic sounds produced by multiple-age speakers in the adolescent period can be accomplished using z scores to control for age and sex differences likely due to differences in vocal tract dimensions. Also, consistent with previous studies of US English rhotics, consonant /r/ production by children in this age range is clearly associated with linguistic context. These two findings suggest that for speech-genetics studies involving rhotic distortions, the reference data found in the Appendix (derived from Lee *et al.*, 1999) may be used to determine the relative status (i.e. normalcy) of productions of stressed vocalic /ɝ/ from speakers aged 5 years to young adulthood. The procedure is to compute a z score for each speaker in question, which in turn would be referenced to some appropriate standard of typical speech production (i.e. a predetermined cutoff value). Owing to the observed phonetic context effects in the present study, such information cannot be generalized to other English rhotics. A companion paper (Shriberg, Flipsen, Karlsson and McSweeny, 2001) uses this approach to develop an acoustic marker for residual /ɝ/ distortions.

Acknowledgements

Thanks to the following colleagues for their valuable assistance at different stages of this project: Chad Allen, Diane Austin, Kate Bunton, Frederic Gruber, Robert Hagiwara, Paul Milenkovic, Rachel Riely, Kristin Tjaden, David Wilson, and clinicians in the Madison Metropolitan School District. Special thanks to Sungbok Lee for providing the individual token data used to construct the Appendix and to Eugene Buder and Karen Forrest for their insightful comments on previous versions of this paper. Preparation of this paper was supported by grants from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, DC00496 and DC02746.

Note

1. For the interested reader, individual speaker data for the 26 speakers in the current study, including the *z*F2 by *z*F3 plots, are archived in Technical Report No. 8 at the Phonology Project website (http://www.waisman.wisc.edu/phonology/).

References

Angelocci, A. A., Kopp, G. A. and Holbrook, A., 1964, The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys. *Journal of Speech and Hearing Disorders*, **29**, 156–170.

Bishop, D. V. M., North, T. and Donlan, C., 1996, Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, **37**, 391–403.

Bond, Z. S. and Wilson, H. F., 1980, /s/ plus stop clusters in children's speech. *Phonetica*, **37**, 149–158.

Cartwright, A. R., Kula, K. and Wright, J. T., 1999, Craniofacial features associated with amelogenesis imperfecta. *Journal of Craniofacial Genetics and Developmental Biology*, **19**, 148–156.

Chaney, C., 1988, Acoustic analysis of correct and misarticulated semivowels. *Journal of Speech and Hearing Research*, **31**, 275–287.

Conover, W. J. and Iman, R. L., 1981, Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, **35**, 124–133.

Cucchiarini, C., 1996, Assessing transcription agreement: Methodological aspects. *Clinical Linguistics and Phonetics*, **10**, 131–155.

Dalston, R. M., 1975, Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, **57**, 462–469.

Dixon, W. E., Jr, Matheny, A. P. and Mohr, S. R., 1995, Heredity and environment in phoneme articulation: Hereditary and environmental contributions to articulation proficiency. *Acta geneticae medicae et gemellologiae*, **44**, 63–73.

Espy-Wilson, C. Y., 1992, Acoustic measures for linguistic features distinguishing the semivowels /w, j, r, l/ in American English. *Journal of the Acoustical Society of America*, **92**, 736–757.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S. and Alwan, A., 2000, Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America*, **108**, 343–356.

Farrer, L. A. and Meaney, F. J., 1985, An anthropmetric assessment of the Huntington's disease patients and families. *American Journal of Physical Anthropology*, **67**, 185–194.

Felsenfeld, S., McGue, M. and Broen, P. A., 1995, Familial aggregation of phonological disorders: Results from a 28-year follow-up. *Journal of Speech and Hearing Research*, **38**, 1091–1107.

Felsenfeld, S. and Plomin, R., 1997, Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. *Journal of Speech, Language, and Hearing Research*, **40**, 778–791.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P. and Pembrey, M. E., 1998, Localization of a gene implicated in a severe speech and language disorder. *Nature Genetics*, **18**, 168–170.

Flipsen, P., Jr, Tjaden, K., Weismer, G., and Karlsson, H., 1996, Acoustic analysis protocol (Tech. Rep. No. 4). Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

Flipsen, P., Jr, Shriberg, L. D., Weismer, G., Karlsson, H. and McSweeny, J, 1998, Acoustic reference data for /s/ in adolescents (Tech. Rep. No. 8). Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

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.

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.

Garn, S. M., Smith, B. H. and LaVelle, M., 1984, Application of pattern profile analysis to malformations of the head and face. *Radiology*, **150**, 683–690.

Gibbon, F. E., 2000, Features of impaired tongue control in children with articulation/phonological disorders. Paper presented at the International Clinical Phonetics and Linguistics Association Conference, August, Edinburgh, Scotland.

Gruber, F. A., 1997, Normalization of developmental phonological disorder: A survival analysis. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Gruber, F. A., 1999, Variability and sequential order of consonant normalization in children with speech delay. *Journal of Speech, Language, and Hearing Research*, **42**, 460–472.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M. and Perkell, J. S., 1999, Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, **105**, 2854–2865.

Hagiwara, R. E., 1995, Acoustic realizations of American /r/ as produced by women and men (UCLA Working Papers in Phonetics No. 90). Los Angeles, CA: University of California.

Hewlett, N., 1988, Acoustic properties of /k/ and /t/ in normal and phonologically disordered speech. *Clinical Linguistics and Phonetics*, **2**, 29–48.

Hillenbrand, J., Getty, L. A., Clark, M. J. and Wheeler, K., 1995, Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97**, 3099–3111.

Hoffman, P. R., Stager, S., and Daniloff, R. G., 1983, Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders*, **48**, 210–215.

Huer, M. B., 1989, Acoustic tracking of articulation errors: /r/. *Journal of Speech and Hearing Disorders*, **54**, 530–534.

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

Karlsson, H. B., Shriberg, L. D., Flipsen, P., and McSweeny, J. L., 2001, Toward an acoustic marker for residual /s/ distortions. Manuscript submitted for publication.

Kent, R. D., 1996, Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, **5(3)**, 7–23.

Kent, R. D., and Vorperian, H. K., 1995, *Development of the craniofacial-oral-laryngeal anatomy* (San Diego, CA: Singular Publishing Group, Inc.).

Klein, R. P., 1971, Acoustic analysis of the acquisition of acceptable r in American English. *Child Development*, **42**, 543–550.

Lahey, M. and Edwards, J., 1995, Specific Language Impairment: preliminary investigation of factors associated with family history and with patterns of language performance. *Journal of Speech and Hearing Research*, **38**, 643–657.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Levy, E. R., Hodgson, S., Fox, M., Jeremiah, S., Povey, S., Jamison, D. C., Green, E. D., Vargha-Khadem, F. and Monaco, A. P., 2000, The SPCH1 region on human 7q31: Genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *American Journal of Human Genetics*, **67**, 357–368.

Lee, S., Potamianos, A. and Narayanan, S., 1999, Acoustics of children's speech: Developmental changes in temporal and spectral parameters. *Journal of the Acoustical Society of America*, **105**, 1455–1468.

Lehiste, I., 1964, Acoustical characteristics of selected English consonants. *International Journal of American Linguistics*, **30(3)**, 1–197.

Lewis, B. A., Cox, N. J. and Byard, P. J., 1993, Segregation analysis of speech and language disorders. *Behaviour Genetics*, **23**, 291–297.

Lewis, B. A. and Freebairn, L., 1993, A clinical tool for evaluating the familial basis of speech and language disorders. *American Journal of Speech-Language Pathology*, **2(2)**, 38–43.

Lewis, B. A. and Shriberg, L. D., 1994, Life span interrelationships among speech, prosody-voice, and nontraditional phonological measures. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association. November, New Orleans, LA.

Lewis, B. A. and Thompson, L. A., 1992, A study of developmental speech and language disorders in twins. *Journal of Speech and Hearing Research*, **35**, 1086–1094.

Lindau, M., 1985, The story of /r/. In: V. A. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged* (New York: Academic Press Inc.), pp. 157–168.

Lobanov, B. M., 1971, Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, **49**, 606–608.

Matheny, A. P. and Bruggemann, C. E., 1973, Children's speech: Hereditary components and sex differences. *Folia phoniatrica*, **25**, 442–449.

Maxwell, E. M. and Weismer, G., 1982, The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics*, **3**, 29–43.

McSweeny, J. L. and Shriberg, L. D., 1995, Segmental and suprasegmental transcription reliability (Tech. Rep. No. 2). Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

Milenkovic, P., 1996, CSpeech (Version 4) [Computer program]. Madison, WI: 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.

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.

Olive, J. P., Greenwood, A. and Coleman, J., 1993, *Acoustics of American English speech. A dynamic approach* (New York: Springer-Verlag).

Parlour, S. F., 1991, Using self-report to identify adults with a history of developmental articulation disorder: a validation. Paper presented at the annual meeting of the Behavior Genetics Association, June, St. Louis, MO.

Peterson, G. E. and Barney, H. L., 1952, Control methods used in the study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175–184.

Rice, M. L., Haney, K. R. and Wexler, K., 1998, Family histories of children with SLI who show extended optional infinitives. *Journal of Speech, Language, and Hearing Research*, **41**, 419–432.

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

Shriberg, L. D., Aram, D. M. and Kwiatkowski, J., 1997, Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, **40**, 313–337.

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, MD: Paul H. Brookes), pp. 73–117.

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., Gruber, F. A. and Kwiatkowski, J., 1994, Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research*, **37**, 1151–1177.

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

Shriberg, L. D. and Lof, G. L., 1991, Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, **5**, 225–279.

Spitz, R. V., Tallal, P., Flax, J. and Benasich, A. A., 1997, Look who's talking: A prospective study of familial transmission of language impairments. *Journal of Speech, Language, and Hearing Research*, **40**, 990–1001.

Stockman, I. J., 1996, Phonological development and disorders in African-American children. In: A. G. Kamhi, K. E. Pollock and J. L. Harris (eds) *Communication development and disorders in African-American children: Research, assessment and intervention* (Baltimore, MD: Paul H. Brookes), pp. 117–153.

Syrdal, A. K. and Gopal, H. S., 1986, A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, **79**, 1086–1100.

Tallal, P., Ross, R. and Curtiss, S., 1989, Familial aggregation in Specific Language Impairment. *Journal of Speech and Hearing Disorders*, **54**, 167–173.

Tomblin, J. B., 1989, Familial concentration of developmental language impairment. *Journal of Speech and Hearing Disorders*, **54**, 287–295.

Tomblin, J. B. and Buckwalter, P. R., 1998, Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research*, **41**, 188–199.

Tomblin, J. B., Freese, P. R. and Records, N. L., 1992, Diagnosing Specific Language Impairment in adults for the purpose of pedigree analysis. *Journal of Speech and Hearing Research*, **35**, 832–843.

Van Der Lely, H. K. J. and Stollwerck, L., 1996, A grammatical specific language impairment in children: an autosomal dominant inheritance? *Brain and Language*, **52**, 484–504.

Weismer, G., 1984, Acoustic analysis strategies for the refinement of phonological analysis [Monograph]. *ASHA*, **22**, 30–52.

Weismer, G., Martin, R., Kent, R. D. and Kent, J. F., 1992, Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *Journal of the Acoustical Society of America*, **91**, 1085–1098.

Westbury, J. R., Hashi, M. and Lindstrom, M. J., 1998, Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, **26**, 203–226.

Appendix

Reference data adapted from Lee et al. (1999) *for /ɝ/ in* bird[a]

	Males			Females		
Age	n[b]	F3−F2[c]	F3/F2[c]	n[b]	F3−F2[c]	F3/F2[c]
5 years	26	797 (343)	1.47 (0.21)	20	643 (210)	1.38 (0.14)
6 years	15	567 (152)	1.37 (0.12)	25	644 (346)	1.35 (0.19)
7 years	19	616 (138)	1.38 (0.09)	32	749 (323)	1.44 (0.22)
8 years	38	517 (175)	1.31 (0.12)	19	669 (497)	1.43 (0.46)
9 years	33	527 (145)	1.34 (0.10)	37	541 (119)	1.31 (0.08)
10 years	40	527 (169)	1.32 (0.11)	24	531 (221)	1.31 (0.13)

11 years	40	474 (170)	1.31 (0.12)	31	523 (107)	1.32 (0.07)
12 years	38	477 (160)	1.32 (0.12)	34	585 (363)	1.36 (0.22)
13 years	23	533 (116)	1.39 (0.09)	20	494 (113)	1.30 (0.07)
14 years	15	390 (130)	1.27 (0.10)	17	476 (143)	1.29 (0.08)
15 years	16	418 (137)	1.31 (0.11)	18	363 (101)	1.23 (0.07)
16 years	14	400 (86)	1.30 (0.07)	19	428 (112)	1.27 (0.08)
17 years	10	420 (115)	1.33 (0.10)	14	414 (116)	1.27 (0.09)
18 years	13	343 (122)	1.27 (0.09)	18	489 (86)	1.32 (0.08)
Adult	30	347 (102)	1.28 (0.09)	39	404 (114)	1.28 (0.10)

[a] Based on individual token values for the formant data reported in Lee *et al*. (1999) for the word *bird*. Individual token data were provided by S. Lee.
[b] To be consistent with previous data presented by Lee *et al*., sample size is based on number of tokens (not speakers). Each cell entry is derived from data for 6–23 speakers.
[c] Cell entries are age-group means (and standard deviations).