Developmental and gender-related trends of intra-talker variability in consonant production

Rachel Romeo,[a] Valerie Hazan,[b] and Michèle Pettinato
Speech, Hearing and Phonetic Sciences, UCL, Chandler House, 2 Wakefield Street, London WC1E 1PF, United Kingdom

This study investigates the effect of age and gender on the internal structure, cross-category distance, and discriminability of phonemic categories for two contrasts varying in fricative place of articulation (/s/-/ʃ/) and stop voicing (/b/-/p/) in word-initial tokens spoken by adults and normally developing children aged 9–14 yr. Vast between- and within-talker variability was observed with 16% of speakers exhibiting some degree of overlap between phonemic categories—a possible contribution to the range of talker intelligibility found in the literature. Females of all ages produced farther and thus more discriminable categories than males, although gender-marking for fricative between-category distance did not emerge until approximately 11 yr of age. Children produced farther yet also much more dispersed categories than adults with increasing discriminability with age, such that by age 13, children's categories were no less discriminable than those of adults. However, children's ages did not predict category distance or dispersion, indicating that convergence on adult-like category structure must occur later in adolescence.

© 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4824160]

I. INTRODUCTION

Studies of the effect of cross-talker variations in intelligibility have sought to identify acoustic-phonetic correlates of intelligibility, linking talker means for various acoustic-phonetic measures with intelligibility rates for these same talkers (Bradlow *et al*., 1996; Hazan and Markham, 2004; van Dommelen and Hazan, 2012). Such investigations are often based on the assumption that intelligibility is strongly linked to the distance between phoneme categories with measurements by means possibly neglecting patterns seen in token-by-token analysis (Hitchcock and Koenig, 2013). Moreover, listeners are also sensitive to the internal phoneme category structure of individual talkers (e.g., Allen *et al*., 2003; Clayards *et al*., 2008; Sumner, 2011; Nygaard and Pisoni, 1998; Theodore *et al*., 2010; Theodore *et al*., 2013), and this within-category phonetic variation has been shown to affect lexical access (McMurray *et al*., 2002). Newman, Clouse, and Burnham (2001) explored the contrast between voiceless sibilant fricatives /s/ and /ʃ/, measuring intra-talker variability by the spectral distance between these phonemic categories and degree of overlap between category distributions. After determining that talkers did differ in category distinctness, they found that listeners were slower to identify initial consonants in tokens spoken by more internally variable talkers with category overlap affecting perception above and beyond small between-category distances. Hazan and Baker (2011) also examined measures of between-category distance and within-category dispersion in

a corpus of 40 adult talkers of British English in two different contrasts: /s/-/ʃ/ and /p/-/b/. Degrees of variability for the fricative and stop contrasts were not correlated, suggesting that intra-talker phonemic variability is contrast-and/or cue-specific rather than a general talker characteristic.

The goal of the present study is to characterize the extent of acoustic variability both within talkers across these same two phonemic contrasts: (/s/-/ʃ/ and /p/-/b/) and between talkers demographically across both age and gender. Stops and fricatives are the two largest feature (articulation manner) categories of English consonants (Ladefoged, 2005), and while the /s/-/ʃ/ contrast is distinguished by place of articulation, the /b/-/p/ contrast is distinguished by voicing; thus these contrasts allow comparison of two critical phonetic features primarily marked by differences in spectral (/s/-/ʃ/) and temporal (/p/-/b/) patterns. Both contrasts have been previously shown to vary both within talkers and systematically between talkers, although not necessarily following the same demographic patterns (Newman *et al*., 2001; Allen *et al*., 2003; Theodore *et al*., 2009).

The production of the contrast between /s/ and /ʃ/ in both adults and children has been investigated in a number of recent studies. These studies differ, to a certain extent, in the measures used to characterize fricative spectra and in the analysis methods used to obtain these measures. The first four moments of the spectral energy distribution have often been used for the analysis of fricative spectra (Forrest *et al*., 1988): The first moment represents the center of gravity (centroid or spectral mean) of the fricative; the second moment, the amount by which the spectrum energy deviates from the centroid, thus giving a measure of standard deviation; the third reflects energy above and below the centroid and therefore gives a measure of skewness; finally the fourth moment reflects the peakedness of the spectrum and

[a] Current affiliation: Ph.D. program in Speech and Hearing Bioscience and Technology, Harvard University and Massachusetts Institute of Technology.
[b] Author to whom correspondence should be addressed. Electronic mail: v.hazan@ucl.ac.uk

therefore gives a measure of kurtosis. These four moments, as well as additional acoustic cues such as spectral transitions (especially the second formant) into the subsequent vowel, have been found contrastive between /s/ and /ʃ/ (McMurray and Jongman, 2011) and transitions are especially prevalent in some languages (e.g., Japanese: Li, 2012). However, in English, the first moment—the centroid—has been found to be the primary acoustic cue for differentiating between the two voiceless sibilants with a higher centroid obtained for /s/ than /ʃ/ (Nittrouer, 1995; Jongman et al., 2000; Fox and Nissen, 2005; Li et al., 2009; Haley et al., 2010), and thus will be the measure used here. In earlier studies, estimates of spectral moments were typically made using discrete Fourier transforms (e.g., Nittrouer, 1995; Jongman et al., 2000; Fox and Nissen, 2005; Li et al., 2009; Maniwa et al., 2009; Li, 2012). However, it has been argued that spectral estimates made using DFTs exhibit large variance due to the random nature of frication noise (Shadle, 2012; Koenig et al., 2013). A more precise spectral estimation can be obtained using multitaper spectra (Blacklock, 2004), and the advantages of using such an analysis have been outlined by Reidy (2013). It should be noted, though, that calculations of fricative centroids using both methods have been found to be strongly correlated (Reidy and Beckman, 2012).

In this study, we will be investigating the effect of gender and age on internal category structure for /s/ and /ʃ/. Although Newman et al. (2001) and Hazan and Baker (2011) noted acoustic differences between male and female voices, neither study systematically analyzes gender effects for this contrast. Women are generally reported to produce higher fricative centroids than men (Jongman et al., 2000; Flipsen et al., 1999; Fox and Nissen, 2005). These spectral differences may be partly related to anatomical differences between adult males and females (Fitch and Giedd, 1999; Fuchs and Toda, 2010) but are also affected by sociophonetic factors in the realization of /s/ especially with women and girls using more fronted articulations leading to higher spectral peaks for /s/ (Stuart-Smith, 2007; Fuchs and Toda, 2010; Holliday et al., 2010). Holliday et al. (2010) found evidence of gender-marking for /s/ in Ohio-based girls as young as 3 yr old; these gender differences were unlikely to be due to physiological factors as they were found for American but not Japanese children of the same age. Age effects have also been investigated for sibilant contrasts. Comparisons of fricatives produced by adults and children have shown evidence of category overlap in English-speaking children but not adults (Li et al., 2009), although some overlap was seen in the distributions for a small number of adults in Holliday et al. (2010) using a peakERB measure (spectral peak calculated on an auditory scale). Furthermore the age at which children achieve adult-like production of this contrast is unclear. McGowan and Nittrouer (1988) found that 7-yr-olds produce sibilants that differ from adult models. In children with normal hearing aged between 2 and 9 yr, peak location for /ʃ/ decreased with increasing age but not peak location for /s/ (Fox and Nissen, 2005; Todd et al., 2011). The greater effect of age on /ʃ/ production reflects similar findings by Fox and Nissen (2005) with an older age range (6–14 yr). Measures of

internal category structure for fricatives produced by children have also been evaluated. Greater within-category variability in both duration and spectral characteristics was found in fricatives produced by children aged between 3 and 9 yr (Munson, 2004). Also, using measures of peakERB and category compactness, Holliday et al. (2010) derived a measure of the "robustness" of the /s/-/ʃ/ contrast for adults and children aged 2–5 yr, representing the accuracy rate with which a fitted logistic regression model could predict the fricative target. Data showed an increase in contrast robustness with age and greater robustness in girls; this was interpreted as a result of gender-marking.

For the stop voicing contrast, analyses will focus on the measure of voice onset time (VOT), the duration between burst release, and the onset of voicing periodicity. Although the stop voicing contrast is marked by a multiplicity of acoustic pattern differences (Lisker, 1986), VOT is recognized as the primary cue marking the contrast (Lisker, 1978). For the voicing contrast, there is also evidence of listeners' sensitivity to the internal structure of phoneme categories: Work on the effect of specific talker characteristics (e.g., VOT range used) on the perception of the stop voicing contrast suggested that listeners adapt both phoneme boundaries and internal category structure to talkers' characteristic productions (Theodore et al., 2010; Theodore et al., 2013). An effect of gender has been shown in the production of the voicing contrast with women producing longer VOTs for /p/ only (Whiteside and Irving, 1998) or longer VOTs for both stops (Whiteside et al., 2004), although some studies find no gender difference for VOT (Morris et al., 2008). Whiteside and Marshall (2001) suggested that these differences in gender effects across studies carried out with different speaker populations may be linked to sociophonetic factors and their effect on allophonic variation. These gender differences in VOT production have been shown to emerge between the ages of 9 and 11 yr with longer VOTs for voiceless stops in girls than boys (Whiteside and Marshall, 2001). Women also typically display larger between-category distance and overall discriminability for stops (Whiteside and Irving, 1998), although gender disparities in dispersion, if they exist, are rarely if ever reported. Thus by comparing male and female talkers within the different age groups on these acoustic-phonetic dimensions, we can better understand how these two demographics might interact throughout speech development. As regards to age effects on the production of the voicing contrast, children typically produce longer VOTs overall (Ohde, 1985), converging to adult-like stop VOTs by age 11 (Whiteside and Marshall, 2001). Furthermore, young children typically exhibit smaller distances between contrasting phonemes before overcompensating as adolescents (Nissen and Fox, 2005; Fox and Nissen, 2005) and exhibit greater category dispersion than adults (Whiteside et al., 2003).

Many of the child studies cited in the preceding text have examined development in the production of these contrasts over the first decade of life. This study will focus on older children, and the main research objective is to establish whether adult-like profiles in terms of phoneme category structure and contrast discriminability are achieved by age

14. In their broader study of speech production in children, Lee *et al.* (1999) found that the greater within-speaker variation found in younger children diminished from around age 12 and reached adult levels around the age of 14 for girls and 15 for boys. Within the age range of 9–14, our expectations are therefore that there will be an effect of age for measures of dispersion and an increase in discriminability. A second objective is to examine whether there is an effect of gender on within-talker variability. The previous literature would lead us to predict greater between-category distance and therefore discriminability for adult female speakers, but the age at which this gender effect will appear—if it has not already—is unclear and is worthy of investigation.

II. METHOD

A. Participants

Speech was recorded from native Standard Southern British English (SSBE) speakers with no reported history of speech or hearing impairment. There were 40 adults (20 male, aged 18–29 yr, mean: 20.8 yr; 20 female, aged 20–28, mean: 23.5 yr) and 73 children aged 9–14 yr (34 male, mean: 12.0 yr; 39 female, mean: 12.6 yr), intended to include the age range at which children achieve adult-like production in the contrasts of interest based on previous literature. Additionally, the following subjects were entirely excluded: Two male children due to many unnatural speech tokens (shouting, whispering, singing, etc.), one male child for producing the tokens out of context, and one male and female child each for technical failure.

B. Materials

Words were elicited via a picture-naming task rather than by repetition of a heard token (as was used in Newman *et al.*, 2001), to approximate a more natural, uninfluenced speaking environment. Eight near minimal pairs, four containing word-onset phonemes /p/ or /b/, and four with /s/ or /ʃ/ were included in the acoustic analyses. The pairs were: Beach-peach, bee-pea, bill-pill, bin-pin, sea-sheep, seat-sheet, cell-shell, and sack-shack. This subset of 16 nouns was analyzed for both children and adults. The adult speech tokens were from the LUCID corpus of British English (Baker and Hazan, 2010); the 16 nouns selected were a subset of 36 keywords included in the recording session. Child tokens were from the child version of the LUCID corpus which is currently being collected; in the child recordings, only the set of 16 keywords included in the analysis were recorded.

C. Procedure

Participants sat in a sound-treated room. The pictures were presented on screen via DMDX software (Forster and Forster, 2003). Participants' utterances were recorded from a Beyerdynamic DT297PV microphone headset directly to the disk at a sampling rate of 22 050 Hz and were automatically saved into separate audio files by DMDX.

As each picture appeared on the computer screen, participants named it aloud within the frame "I can see a (noun)" to approximate a natural linguistic context and elicit citation style speech, which would be closer to a clear than conversational speech style (Smiljanic and Bradlow, 2009). Each picture was presented eight times in a randomized order yielding 128 tokens per talker: 64 per contrast and 32 per phoneme category. After recording, a phonetically trained listener excluded tokens judged to be misarticulations (e.g., frontal misarticulations of /s/, depalatization of /ʃ/, etc.), as well as unrepresentative tokens due to technical malfunction and/or disruptive background noise, missing, incomplete, or wrong keywords, rapid corrections resulting in atypical speaking rates, or any other solitary outlying production for a speaker. For the 40 adults, the picture-naming task yielded 5120 total tokens of which 194 (3.8%) were excluded, and for the 73 children, this task yielded 9344 total tokens of which 266 tokens (2.8%) were excluded.

This resulted in an average loss of 3.6 tokens per talker (3%) and 101 tokens per phoneme (3%).

D. Analysis

Word-initial consonant onsets and offsets were annotated for each token in PRAAT (Boersma and Weenink, 2012). For keywords with initial /p/-/b/, markers were placed at the burst release onset (at the initial rise in amplitude) and at the beginning of the first full periodic cycle indicating the onset of voicing (Ladefoged, 2003). A PRAAT script was then used to calculate the VOT for each token.

For keywords with initial /s/-/ʃ/, markers were placed at the start of the frication portion (after the periodicity of the previous vowel had ended and noise amplitude began to rise) and at the end of the frication portion (beginning of the first full periodic cycle of subsequent vowel), excluding portions of mixed excitation. For each token, the R scripts described in Reidy (2013) were used to band-pass filter the audio file (with low and high cut off frequencies of 300 and 20 000 Hz), then compute multitaper spectra (Thomson, 1982) using eight tapers for the middle 50% of the fricative. Because frication energy is spectrally dynamic over time and peaks tend to emerge more in the middle to end (Behrens and Blumstein, 1988), this medial measurement avoids the effects of frication onset variation and subsequent vowel formant transitions (Meyer and Ptok, 2011). No pre-emphasis was applied. The four spectral moments were then obtained. Because all four moments were correlated (all Pearson's $r > |0.07|$, all $p < 0.001$) and frication centroids are often the most distinguishing measure of /s/ and /ʃ/ (e.g., Jongman *et al.*, 2000), only this measure was used to quantify category means and dispersions (standard deviations of all token centroids from one speaker). To evaluate the comparability of centroid measures obtained using multitaper and DFT spectra, a PRAAT script was used to calculate the same four spectral moments using a Hanning window, using the same bandwidth and duration settings. All four moments were highly correlated (all $r > 0.87$, all $p < 0.01$) with the highest correlations for centroid measures ($r > 0.99$, $p < 0.001$), extending the findings of Reidy and Beckman (2012) to talkers of British English accent and to children.

To quantify talkers' discriminability, three additional measures were derived for each phoneme contrast for each

J. Acoust. Soc. Am., Vol. 134, No. 5, November 2013

Romeo *et al.*: Intra-talker variability   3783

talker. *Within-category dispersion* equaled the mean of the standard deviations of the frication centroids for all /s/ and /ʃ/ tokens and of the stop VOTs for all /p/ and /b/ tokens. *Between-category distance* equaled the difference between the mean centroids for the fricative contrast and between the mean VOTs for the stop contrast. The magnitude of this disparity was further calculated as the difference between the minimum /s/ and maximum /ʃ/ centroids and between the minimum /p/ and maximum /b/ VOTs; if positive, this measures the magnitude of distance between contrasting category extremes and if negative, the magnitude of category overlap. Finally, an overall measure of phoneme *discriminability d(a)* corresponded to sensitivity from signal detection theory—because phoneme category variances were not equal, $d(a)$ equals the difference between the mean values of the frication centroids or stop VOTs (distance), divided by the square root of the mean of the variances (roughly dispersion). When these variances are equal, $d(a)$ reduces to the common d' (Simpson and Fitter, 1973).

III. RESULTS

The 73 children were divided into three age subgroups (9–10 yr, $n = 24$; 11–12 yr, $n = 16$; and 13–14 yr, $n = 33$). As these three groups only differed for two individual phonemic, contrast, or interaction measures [remaining $F(2,67) < 2.18$, $p > 0.12$], the subgroups were collapsed for most comparisons. Neither did splitting children into more similarly sized subgroups ($n = 24$–25) uncover significant age group effects [all $F(2,67) < 1.18$, all $p > 0.19$]. Additionally, children's chronological age (in months) did not correlate with any but one of these measures (remaining $r < 0.14$, $p > 0.23$). Therefore separate analyses for subgroups are only presented when measures significantly differ by age group or where chronological age is a significant predictor for a measure.

A. Fricatives

Figure 1 shows the distribution of /s/ and /ʃ/ tokens' centroids as a function of age group (adult, child) and gender, displaying apparent differences in distributions. The singular effects of age and gender on centroid values and on the various phoneme category structure measures were further explored.

1. Phoneme discrimination

Mean centroid values and standard deviation as a function of age group and gender are presented in Table I. A mixed-design analysis of variance (ANOVA) with phoneme as a
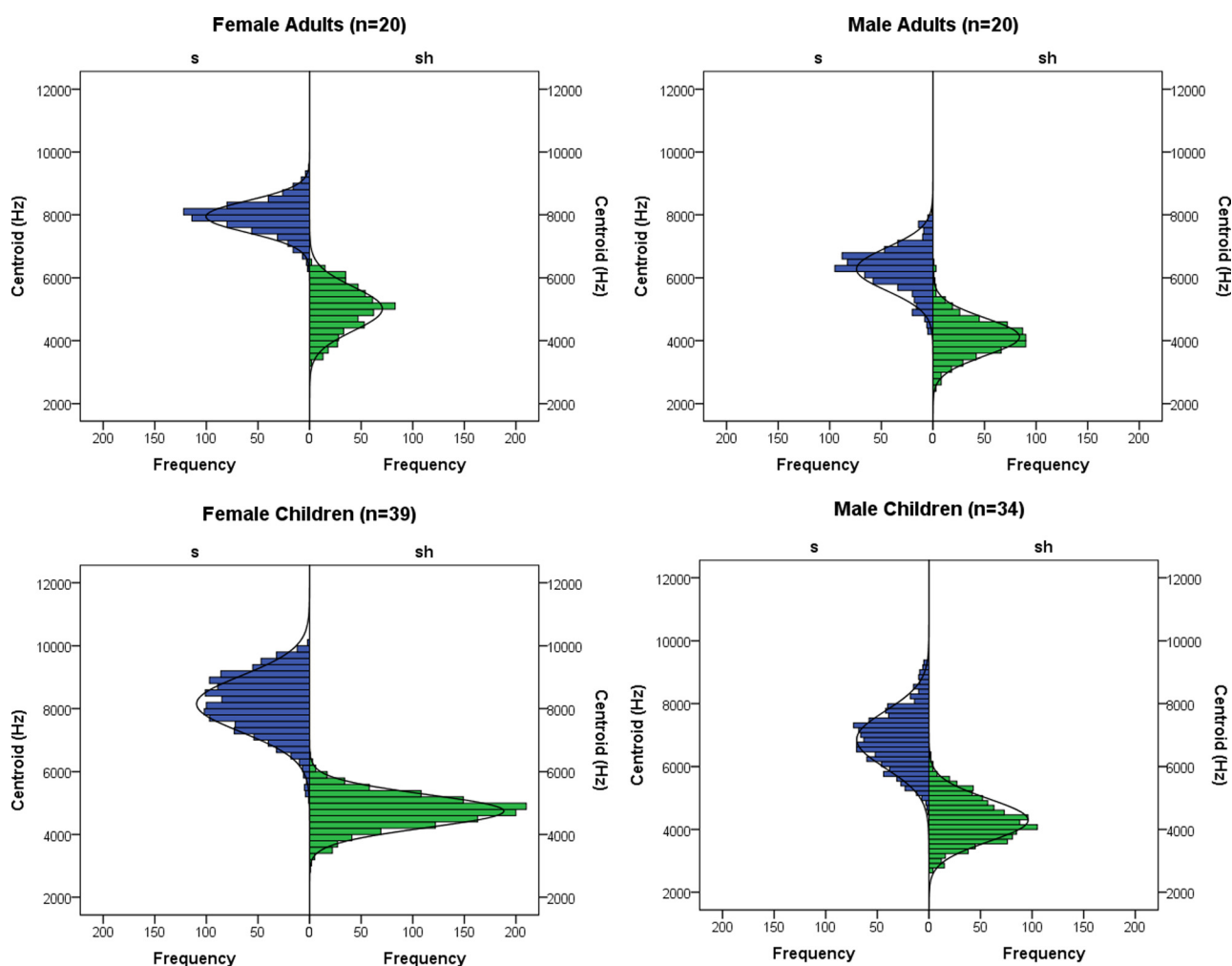

FIG. 1. (Color online) Histograms of fricative centroids (in Hz) for productions of /s/ (labeled "s") and /ʃ/ (labeled "sh") tokens as a function of gender and age group. Children were aged 9–14 years.

Romeo *et al.*: Intra-talker variability

TABLE I. Mean fricative centroids and stop VOTs for child and adult age groups. Standard deviations are in parentheses.

			Fricative centroid (Hz)		Stop VOT (ms)	
			/s/	/ʃ/	/b/	/p/
Children	9–10 yrs (n = 24)	Male (n = 13)	7339 (724)	4446 (596)	8.3 (2.2)	66.9 (17.1)
		Female (n = 11)	7675 (732)	4869 (281)	8.0 (1.1)	76.6 (11.1)
	11–12 yrs (n = 16)	Male (n = 7)	6657 (767)	4217 (551)	10.5 (2.0)	69.3 (26.2)
		Female (n = 9)	8619 (312)	4757 (309)	10.6 (1.8)	76.5 (17.4)
	13–14 yrs (n = 33)	Male (n = 14)	6430 (599)	4196 (485)	12.0 (4.7)	74.7 (20.9)
		Female (n = 19)	8192 (769)	4663 (457)	10.6 (3.6)	80.6 (15.3)
	All children (n = 73)	Male (n = 34)	6818 (788)	4299 (537)	10.3 (3.8)	70.6 (20.4)
		Female (n = 39)	8135 (751)	4745 (381)	9.9 (2.9)	78.5 (14.6)
Adults (n = 40)		Male (n = 20)	6277 (632)	4124 (482)	12.1 (2.4)	57.2 (13.7)
		Female (n = 20)	7946 (367)	5016 (649)	12.9 (4.0)	70.0 (13.9)

bi-level within-subjects factor, both age group and gender as bi-level between-subjects factors, and centroids (first moment) as the dependent variable revealed a main effect of phoneme [$F(1,109) = 1380.42$, $p < 0.001$, effect size $\eta^2 = 0.927$], indicating that centroid values sufficiently distinguished intended /s/ and /ʃ/ productions, with /s/ centroids having higher energy concentrations (mean: 7298 Hz) than /ʃ/ centroids (mean: 4545 Hz). A second ANOVA with standard deviations of the centroids (square root of the second moment) as the dependent variable revealed that talkers' token-by-token /s/ categories were more dispersed than their /ʃ/ categories [$F(1,109) = 12.56$, $p < 0.01$, $\eta^2 = 0.103$, mean /s/ $\sigma = 405$ Hz, mean /ʃ/ $\sigma = 351$ Hz].

However, within individual talkers, the dispersions between the two categories are marginally correlated ($r = 0.174$, $p < 0.07$), justifying an averaged within-talker dispersion value for simplicity.

2. Effect of gender

The preceding mixed-design ANOVA for centroids also revealed a main effect of gender [$F(1,109) = 135.63$, $p < 0.001$, $\eta^2 = 0.554$], indicating that females overall produced higher centroids (6462 Hz) than males (5381 Hz). Furthermore, a gender*phoneme interaction [$F(1,109) = 30.98$, $p < 0.001$, $\eta^2 = 0.22$] and *post hoc* univariate ANOVAs (gender as independent variable, mean category centroids separately as dependent variables) indicated that females produced higher centroids for both /s/ [$F(1,111) = 118.27$, $p < 0.001$, female: 8077 Hz, male: 6622 Hz] and for /ʃ/ [$F(1,111) = 39.24$, $p < 0.001$, female: 4835 Hz, male: 4232 Hz], yet the effect size is much larger for /s/ (1455 Hz; $\eta^2 = 0.516$) than for /ʃ/ (603 Hz; $\eta^2 = 0.261$).

To investigate category distinction, measures of between-category distance, within-category dispersion, and overall discriminability were calculated for each talker as described in the preceding text (see Fig. 2). These measures were separately used as the dependent variables of a multivariate ANOVA with gender and age group as independent variables. This yielded a main effect of gender for between-category distance [$F(1,109) = 30.98$, $p < 0.001$, $\eta^2 = 0.221$], indicating that females' phoneme categories were farther apart than males' categories (male: 2341 Hz, female: 3166 Hz). However, a gender*age interaction [$F(1,105) = 5.42$, $p < 0.01$, $\eta^2 = 0.13$]

reveals that this gender difference does not emerge until approximately age 11 with 9- and 10-yr-olds exhibiting no male-female difference in category distance [$t(22) = 0.26$, $p > 0.7$].

Despite no main effect of gender on dispersion [$F(1,109) = 0.5$, $p > 0.4$], a gender*phoneme interaction [$F(1,109) = 8.50$, $p < 0.01$, $\eta^2 = 0.072$] indicates that females have a much higher standard deviation of their /s/ centroids (422 Hz) than their /ʃ/ centroids (322 Hz), while males show little difference in dispersion between the two categories (/s/ = 387, /ʃ/ = 379 Hz), driving the overall /s/-/ʃ/ dispersion difference mentioned previously.

When large category dispersions such as these are not separated by similarly large between-category distances, this can result in intra-talker category overlap. While much of the visible overlap in Fig. 1 simply results from averaging across talkers (the mean /s/ for one talker could coincide with the mean /ʃ/ for another talker), overlapping categories were seen in the individual distributions of 11 talkers (10% of the sample). Interestingly, all but two overlapping talkers were male, and all but one were children, although they were scattered relatively evenly across the child age range with no correlation between specific age and overlap ($p > 0.7$). Their magnitude of overlap ranged from 29 to 1457 Hz, comprising 3% to 41% of tokens. Combined with the female tendency to produce further fricative categories, this nearly all-male overlap gave rise to a gender effect on the magnitude of distance between category extremes [$F(1,109) = 30.01$, $p < 0.001$, $\eta^2 = 0.22$] with males averaging 671 Hz between /s/-/ʃ/ distribution extremes and females averaging 1569 Hz.

Finally, because between-category distance but not within-category dispersion varied between genders, this created a gender difference in overall category discriminability [$F(1,109) = 25.62$, $p < 0.001$, $\eta^2 = 0.19$], such that males' sibilant categories were less discriminable [mean $d(a)$: 6.28] than females' categories [mean $d(a) = 8.77$].

3. Effect of age

The previously described mixed-design ANOVA with centroid means as the dependent variable revealed only a marginally significant main effect of age group [$F(1,109) = 3.01$, $p < 0.1$].
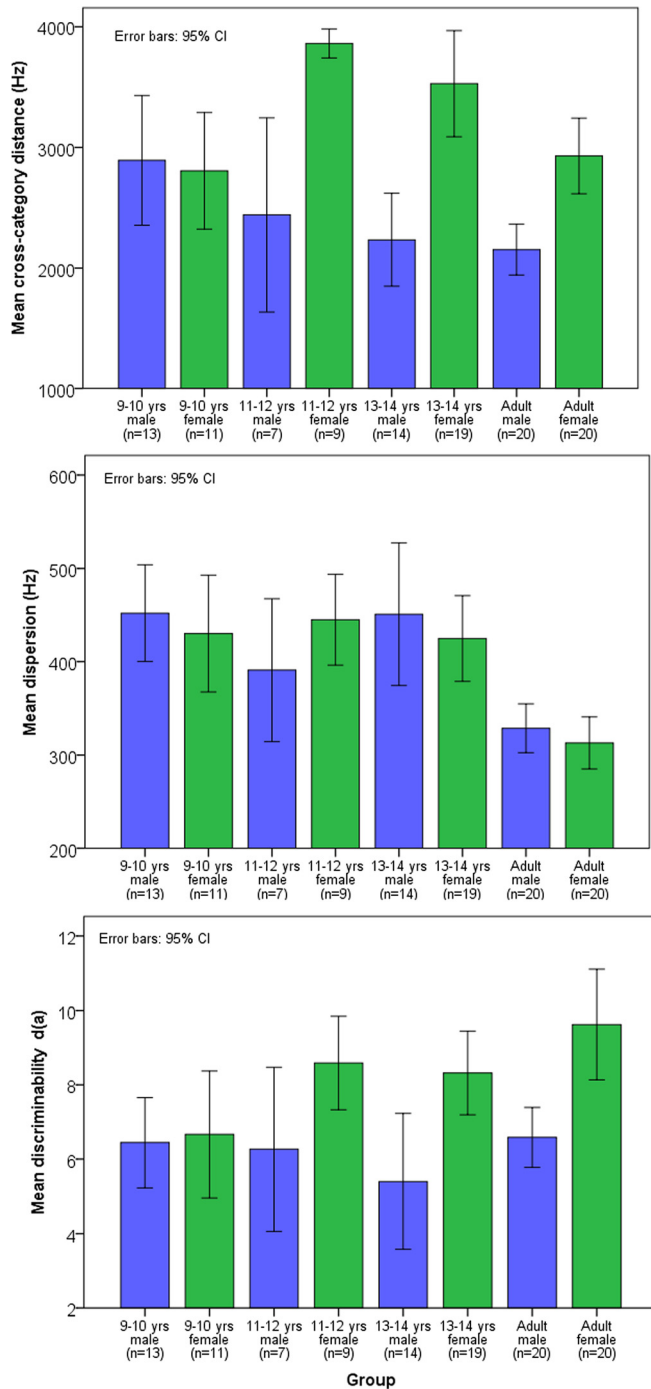
FIG. 2. (Color online) Distance between /s/-/ʃ/ categories, average dispersion within both categories, and /s/-/ʃ/ category discriminability $d(a)$ as a function of gender and age subgroups.

However, a group*phoneme interaction [$F(1,109) = 8.18$, $p < 0.01$, $\eta^2 = 0.07$] and *post hoc* univariate ANOVAs (gender as independent variable, mean category centroids separately as dependent variables) indicated that children produced higher /s/ centroids than adults [$F(1,111) = 4.51$, $p < 0.05$, $\eta^2 = 0.04$; child mean: 7530 Hz, adult mean: 7111 Hz], yet the groups exhibit no difference in /ʃ/ centroids [$F(1,111) > 0.09$, $p > 0.7$, child: 4535 Hz, adult: 4570 Hz], thus indicating that overall age differences are primarily due to children's higher /s/ categories.

Again we refer to the above-mentioned multivariate ANOVA with independent variables of gender and age group and dependent variables of between-category distance, within-category dispersion, and overall discriminability calculated per subject. Children overall had not yet decreased their between-category distance to adult-like values [$F(1,109) = 8.18$, $p < 0.01$, $\eta^2 = 0.07$, adult: 2541 Hz, children: 2965 Hz] with all individual child age subgroups exhibiting at least marginally farther categories than adults (all $t > 3.13$, all $p < 0.08$). Looking closer, a gender*age group interaction effect that disappears as the children get older [9–10 yr olds: $F(60) = 6.16$, $p < 0.02$, 11–12 yr olds: $F(52) = 3.56$, $p < 0.07$, 13–14 yr olds: $F(69) = 2.49$, $p > 0.12$] reveals that males very gradually decrease in between-category distance between age 9 and adulthood, while females dramatically increase their between-category distances between ages 10 and 11 before gradually decreasing. This is due exclusively to a significant rise in these young females' average /s/ centroid between the 9–10 yr old group and the 11–12 yr olds [$t(18) = 3.58$, $p < 0.01$], which yields a gender difference in /s/ centroids and between-category distance that persists throughout adulthood. Surprisingly though, within each gender group, children's precise ages (in months) do not predict their between-category distance [multiple regression coefficient $t(72) = -0.66$, $p > 0.9$], and the significant difference obtained between children and adults suggests that a meaningful decrease in between-category distance must occur between ages 14 and 18.

In the aforementioned mixed-design ANOVA with centroid standard deviations as dependent variable, a main effect of age group [$F(1,109) = 46.22$, $p < 0.001$, $\eta^2 = 0.30$] indicates that children on average have more dispersed categories (435 Hz) than adults (321 Hz). Furthermore, a phoneme*age group interaction [$F(1,109) = 14.96$, $p < 0.001$, $\eta^2 = 0.12$] and *post hoc* univariate ANOVAs (age group as independent variable, mean category standard deviations separately as dependent variables) revealed that this age difference in dispersion is much larger for /s/ categories [$F(111) = 49.70$, $p < 0.001$, $\eta^2 = 0.31$, child $\sigma = 492$ Hz, adult $\sigma = 318$ Hz, difference = 174 Hz] than for /ʃ/ categories [$F(111) = 6.44$, $p < 0.02$, $\eta^2 = 0.06$, child $\sigma = 378$ Hz, adult $\sigma = 323$ Hz, difference = 55 Hz]. This adult-child difference in within-category dispersion also held for each child age subgroup (all $t > 27$, all $p < 0.001$), yet, within each gender group, children's precise ages in months did not predict their within-category dispersion [multiple regression coefficient $t(72) = 0.66$, $p > 0.5$], indicating that another meaningful shift in category structure must occur between ages 14 and 18.

From a multivariate ANOVA with independent variables of gender and age group, and dependent variable of overall discriminability, it at first appears that children's large category distances were not sufficient to counteract their larger dispersions, yielding slightly less discriminable categories than adults [$F(1,109) = 5.55$, $p < 0.05$, $\eta^2 = 0.05$, adult $d(a)$: 8.10, child $d(a)$: 6.94]. However, closer examination reveals that this small effect is due exclusively to the very low discriminability of the youngest (9–10 yr old) female children; this differs from that of female adults

[$t(29) = 2.65$, $p < 0.05$] and that of all other female children [$t(37) = 2.19$, $p < 0.05$]. These young girls had yet to undergo the extreme increase in /s/ centroids, and when they are excluded from the child group, only a marginal difference in overall discriminability between children and adults remains [$F(98) = 3.34$, $p > 0.07$]. Finally, within genders, children's precise ages in months once more do not predict their overall discriminability [multiple regression coefficient $t(72) = 0.69$, $p > 0.4$]. Again, this suggests that between the ages of 14 and 18, children's categories must grow closer together but also much less dispersed, preserving their general adult-like category discriminability.

B. Stops

Figure 3 shows the VOT distributions for /b/ and /p/ tokens split across age groups and genders, displaying apparent differences in distributions. The singular effects of age and gender on VOT values and on the various phoneme category structure measures were further explored.

1. Phoneme discrimination

Mean VOT values and standard deviation as a function of age group and gender are presented in Table I. A mixed-design ANOVA with phoneme as a bi-level within-subjects factor, both age group and gender as bi-level between-subjects factors, and VOT as the dependent variable revealed a main effect of phoneme [$F(1,109) = 1367.64$, $p < 0.001$, $\eta^2 = 0.93$], indicating, as expected, that VOT values sufficiently distinguished intended /b/ and /p/ productions, with longer VOTs for /p/ (69 ms) than /b/ (11 ms). Additionally, the same ANOVA with standard deviations as the dependent variable revealed that talkers' token-by-token /p/ categories were more dispersed than their /b/ categories [$F(1,109) = 632.56$, $p < 0.001$, $\eta^2 = 0.85$, /p/ $\sigma = 4$ ms, /b/ $\sigma = 17$ ms]. This is unsurprising given that unvoiced stops are less constrained in the length of aspiration before voicing, while voiced stops are much more limited in range, including a minimum VOT of 0 ms (there was no evidence of pre-voicing in any tokens). However, within individual talkers, the two category dispersions are highly correlated ($r = 0.284$, $p < 0.01$), again justifying an averaged overall within-talker dispersion value for simplicity.

2. Effect of gender

The preceding mixed-design ANOVA for VOTs also revealed a main effect of gender [$F(1,109) = 9.48$, $p < 0.01$, $\eta^2 = 0.08$], indicating that females overall produced longer
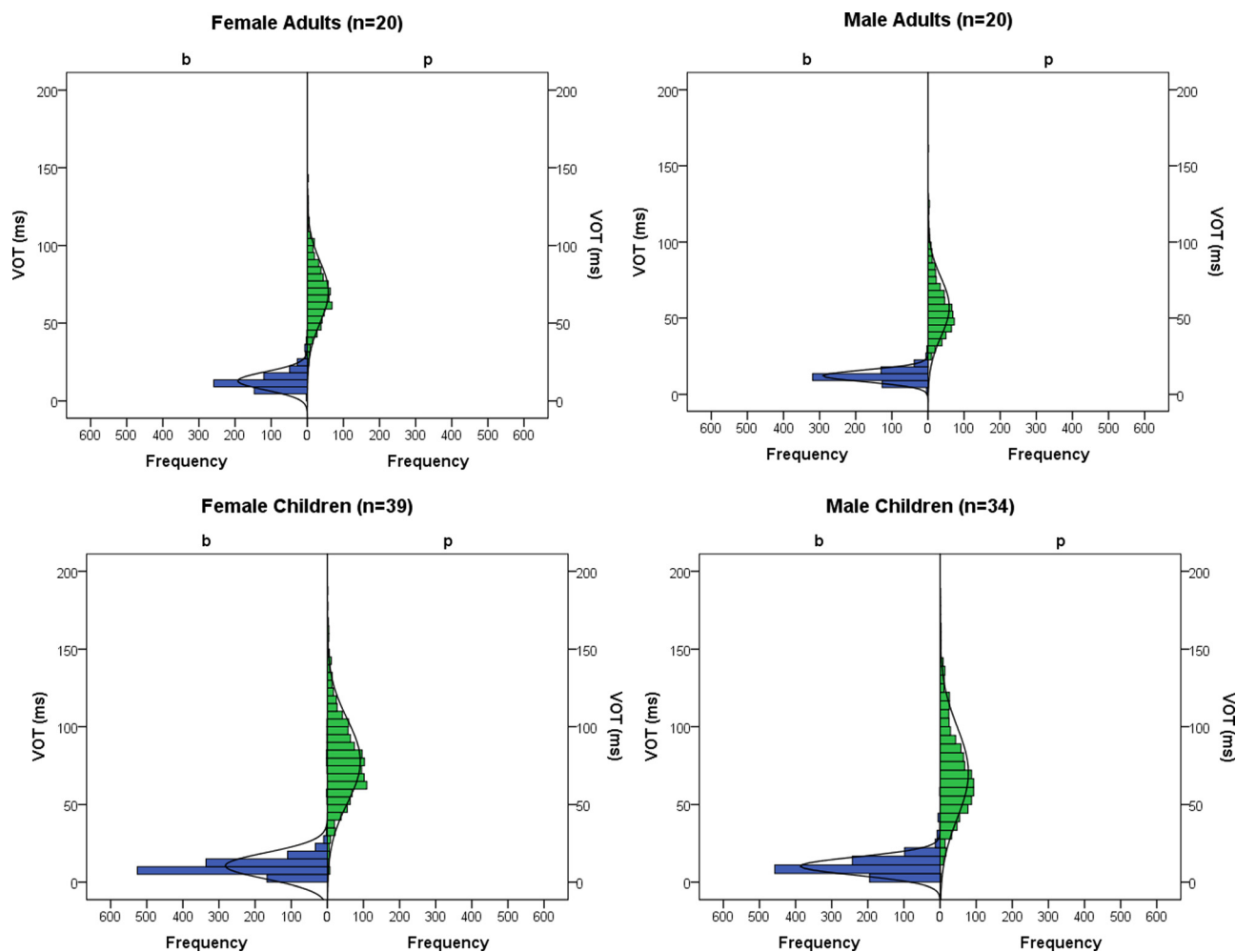

FIG. 3. (Color online) Histograms of VOT durations for productions of /b/ (labeled "b") and /p/ (labeled "p") tokens as a factor of gender and age group. Children were aged 9–14 years.

J. Acoust. Soc. Am., Vol. 134, No. 5, November 2013

Romeo et al.: Intra-talker variability   3787

VOTs (43 ms) than males (38 ms). Furthermore, a gender* phoneme interaction $[F(1,109) = 10.668, \quad p < 0.001, \eta^2 = 0.09]$ and *post hoc* univariate ANOVAs (gender as independent variable, mean category VOTs separately as dependent variables) indicated that females produced longer /p/ VOTs $[F(1,111) = 9.72, \quad p < 0.01, \quad \eta^2 = 0.08$; female $= 74$ ms, male $= 64$ ms], yet no differences in /b/ VOTs $[F(1,111) = 0.01, \text{ p} > 0.0.9, \text{ both genders 11 ms}]$, indicating that any gender differences are exclusively due to /p/ VOTs.

A multivariate ANOVA was conducted with independent factors of gender and age group and dependent measures of between-category distance, within-category dispersion, and overall discriminability, a summary of which is displayed in Fig. 4. As with fricatives, males had smaller between-category distances than females $[F(1,109) = 10.67, p < 0.01, \quad \eta^2 = 0.09, \text{ male: 53 ms, female: 63 ms}]$, and genders showed no difference in within-category dispersion $[F (109) = 0.79, \quad p > 0.3, \text{ male: 10 ms, female: 11 ms}]$. Together, this yields the slight category overlap seen in Fig. 3, much of which results from averaging across talkers. However, as for fricatives, category overlap was seen in the individual distributions for 11 talkers, and again this group was largely comprised of male children (7 male children, 2 female children, and 1 each of male and female adults). The magnitude of overlap ranged from 2 to 32 ms and comprised 3% to 47% of an individual's tokens. Finally, because between-category distance but not within-category dispersion varied between genders, this created a gender difference in overall discriminability $[F(1,109) = 7.28, \quad p < 0.01, \eta^2 = 0.06]$ such that males' labial stop categories are less discriminable [mean $d(a) = 4.49$] than females' categories [mean $d(a) = 5.15$].

3. Effect of age

The previously described mixed-design ANOVA with dependent VOT means revealed a main effect of age group $[F(1,109) = 10.67, \quad p < 0.001, \quad \eta^2 = 0.09]$, indicating that children produced overall longer mean VOTs (42 ms) than adults (38 ms). Furthermore, a phoneme*group interaction $[F(1,109) = 18.54, p < 0.001, \eta^2 = 0.15]$ and *post hoc* univariate ANOVAs (gender as independent variable, mean category VOTs separately as dependent variables) indicate that children produced longer /p/ VOTs $[F(1,111) = 11.41, p < 0.001, \eta^2 = 0.09, \text{ child: 75 ms, adult: 64 ms}]$ yet *shorter* /b/ VOTs $[F(1,111) = 14.63, \quad p < 0.001, \quad \eta^2 = 0.17, \text{ child: 10 ms, adult: 13 ms}]$. Repeating this analysis with child age subgroups, 9–10 yr olds produced shorter /b/ VOTs (8 ms) than the older children (11 ms) $[F(1,67) = 7.29, p < 0.001, \eta^2 = 0.18]$, suggesting some lengthening of /b/ VOTs throughout adolescence. Because of the youngest children's shorter /b/ VOTs, it is unlikely that age effects for /p/ VOTs are simply due to children using a slower speech rate. Examination of individual child tokens revealed that in instances of much lengthened words, this longer duration was due to a significant elongation of the vowel rather than of the initial consonant. Furthermore, in /b/-/p/ VOT continua with constant vowel length, the phoneme boundary is minimally affected by speech rate (Jongman, 1986).
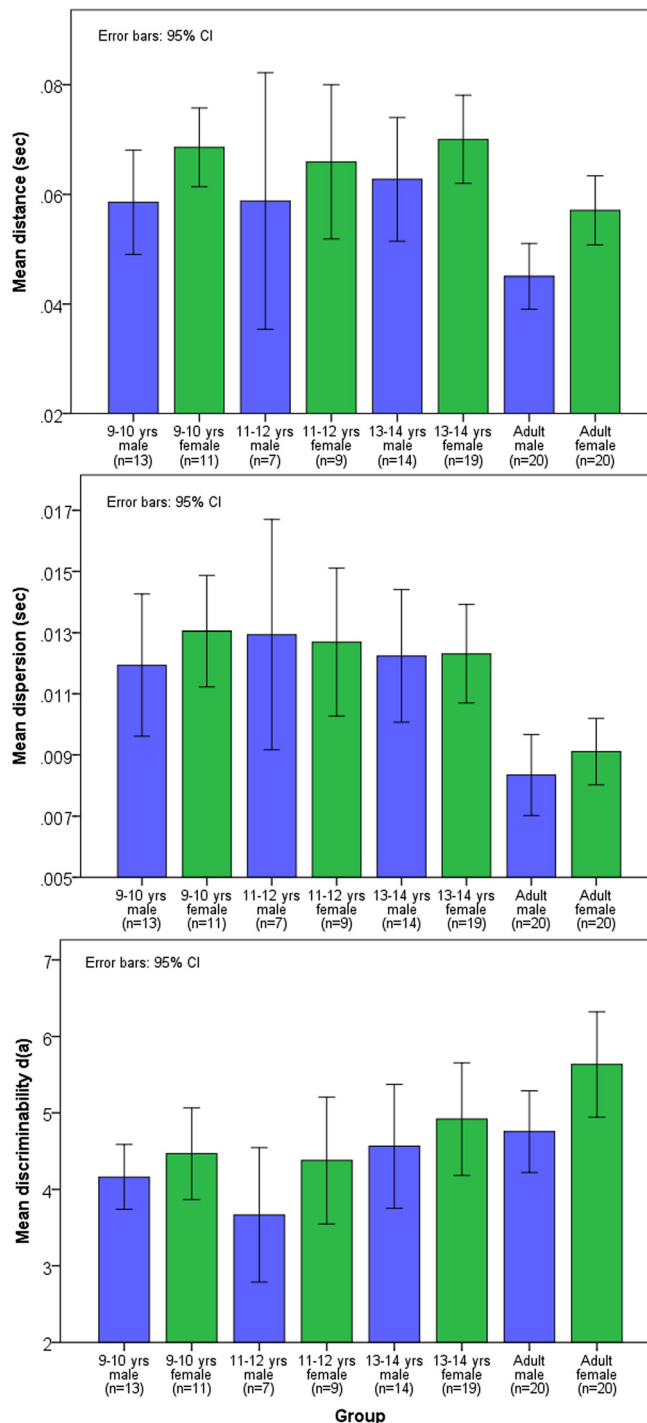

FIG. 4. (Color online) Distance between /b/-/p/ categories, average dispersion within both categories, and /b/-/p/ category discriminability $d(a)$ as a function of gender and age subgroups.

Again we refer to the above-mentioned multivariate ANOVA with independent variables of gender and age group and dependent measures of between-category distance, within-category dispersion, and overall discriminability calculated per subject. Due to their shorter /b/ and longer /p/ VOTs, children had greater between-category distance than adults $[F(1,109) = 18.54, p < 0.001, \eta^2 = 0.15, \text{ adult: 51 ms, child: 65 ms}]$, yet there was no decreasing pattern in distance across children's precise ages in months [multiple regression coefficient $t(1,70) = 0.31, p > 0.7$] nor differences

between age subgroups [$F(2,67) = 0.34$, $p > 0.7$], suggesting that children must acquire adult-like category closeness at some age between 14 and 18.

In the aforementioned mixed-design ANOVA with dependent VOT standard deviations, a main effect of age group [$F(1,109) = 35.96$, $p < 0.001$, $\eta^2 = 0.25$] indicates that children on average have more dispersed categories (12 ms) than adults (9 ms), and a phoneme*age group interaction [$F(1,109) = 35.62$, $p < 0.001$, $\eta^2 = 0.25$] revealed that this is due mostly to differences in /p/ category dispersions (children = 21 ms, adults = 14 ms, difference = 7 ms), with no age difference in /b/ category dispersions (both ages 4 ms). What is more, children did not approach an adult-like narrowness across the ages (108–179 months) studied here [multiple regression coefficient $t(1,70) = 0.84$, $p > 0.4$], again pointing toward an acquisition of adult-like category narrowness between ages 14 and 18.

As with fricatives, a multivariate ANOVA with independent variables of gender and age group and dependent variable of overall discriminability shows that children's stop category distance was not quite sufficient to counteract this extreme category dispersion, yielding lower overall category discriminability $d(a)$ for children [$F(1,109) = 14.50$, $p < 0.01$, $\eta^2 = 0.08$, adult $d(a)$: 5.20, child $d(a)$: 4.45]. However, this was also the only measure to yield a marginally significant regression toward adult values (within genders) between 9 yr 0 months and 14 yr 11 months [$t(1,70) = 1.74$, $p < 0.09$], such that by the age of 13–14, these children produce no less discriminable stop categories than adults [$F(1,69) = 1.91$, $p > 0.1$]. Thus even though stop distance and dispersion are not adult-like by this age, the oldest children balance them to attain an adult-like discriminability of the contrast with children younger than 13 steadily approaching adult-like discriminability. Interestingly though, there was no age*gender effect for any phoneme or contrast measures, indicating that male and female children develop stop articulations similarly as they age into adulthood yet with no distinct trajectory in distance nor dispersion.

C. Relationship between contrasts

Standardized Z-scores of between-category distance, within-category dispersion, overall category discriminability, and presence of overlap were calculated for each contrast and talker; then two-tailed Pearson's product-moment correlations were computed.

For the fricative contrast, distance and dispersion were marginally correlated overall ($r = 0.17$, $p < 0.08$), yet were not correlated within any of the four talker subgroups (all $|r| < 0.16$, all $p > 0.33$), suggesting that any overall correlation is due to demographic trends rather than individual differences. For the stop contrast, distance and dispersion were much more strongly correlated overall ($r = 0.64$, $p < 0.001$) as they were for all children and male adults (male children: $r = 0.61$, $p < 0.001$; female children: $r = 0.52$, $p < 0.001$; male adults: $r = 0.78$, $p < 0.001$) but not so for female adults ($r = 0.31$, $p > 0.1$). Because all correlations are positive, this indicates that as /p/ VOT increases and thus distance between stop categories increases, so does dispersion within

the /p/ category. However, this relationship grows stronger with age for males, yet weakens with age for females. For adults, these are the same results found by Hazan and Baker (2011), and this extends these relationships to children. However, it is plausible that because /p/ VOT distributions are often right-skewed, as are all upper-boundless voiceless aspirated stops (Koenig, 2001), the distance-dispersion relationship could be an artifact of /p/ means being pulled higher as the skew and dispersion grow. Yet when between-category distance is instead measured by category medians, all of these relationships still hold (overall $r = 0.62$, $p < 0.001$; male children: $r = 0.62$, $p < 0.001$; female children: $r = 0.46$, $p < 0.001$; male adults: $r = 0.75$, $p < 0.001$; female adults: $r = 0.31$, $p > 0.1$). This indicates that for stops, between-category distance and within-category dispersion increase together, perhaps to balance discriminability; however, for fricatives, distance and dispersion vary independently between demographically similar individuals, suggesting that the relationship between these measures varies between these contrasts.

Finally, to investigate whether phoneme variability is a consistent characteristic of the talker, correlations between these measures were also calculated between fricative and stop contrasts. Over all talkers, between-category distance ($r = 0.30$, $p < 0.001$), within-category dispersion ($r = 0.47$, $p < 0.001$), and overall discriminability ($r = 0.26$, $p < 0.01$) were correlated across contrasts, appearing as if these measures might indicate a general pattern that individuals exhibit across different types of phonemic contrasts. However, when talkers are split into the four gender*age subgroups, none of these measures were correlated within individuals (all $|r| < 0.28$, all $p > 0.12$), except $d(a)$ for male children ($r = 0.45$, $p < 0.01$) and marginally for their dispersion ($r = 0.30$, $p < 0.10$), suggesting that the overall correlations can be attributed to demographic trends rather than individual differences and, more broadly, that talkers appear to exhibit contrast-specific levels of phonemic distinction rather than general articulatory traits. Yet looking more closely, the correlations seen in male children were largely driven by the presence of overlapping categories—of the people with any category overlap (18 subjects, 16%), most only exhibited overlap in one contrast (14 subjects, 78%), but the few who exhibited overlap in both contrasts (4 subjects, 22%) were exclusively male children. In fact, when the magnitude of overlap is standardized, both male and female children exhibited nearly significant correlations between contrasts in the magnitude of their category overlap or distance (male: $r = 0.33$, $p < 0.06$; female: $r = 0.32$, $p < 0.6$). In all, this suggests that children, especially males, might possess a general trend in their level of phonemic category distinction, yet by the time they reach adulthood, they stop generally applying these trends, resulting in contrast-specific levels of phoneme distinction.

IV. DISCUSSION

This study explored the emergence of adult-like phoneme category structure and contrast discriminability in male and female children aged 9–14. As well as considering

distance between categories, emphasis was placed on the development of within-phoneme dispersion and overall category discriminability. Because within-talker variation in children has been found to approach adult-like values within this age range, it was hypothesized that there should be an age effect for measures of category dispersion and increased discriminability. In line with predictions, children had greater within-category dispersion than adults, specifically within /s/ and /p/; this variability did not tail off in the oldest group nor was there an effect of children's linear ages. Notably the speech stimuli only contained front vowel contexts; had a full English vowel set been included, one would suspect to find even greater dispersion. This suggests that a significant shift toward narrower adult-like dispersions is still to take place between the ages of 14 and 18. This was an unexpected finding given that most literature for stops finds that children typically reach adult-like within-category dispersion by age 11 (Whiteside *et al.*, 2003) or even by age 8 (Eguchi and Hirsh, 1969). Although little work has been done on the spectral dispersion of fricatives between the ages of 9–14, children's greater within-category variability is consistent with that of Munson (2004) for younger children (ages 3–8), and with Lee *et al.* (1999) in terms of *temporal* fricative measures up to age 14. These discrepancies could in part be due to measurements—Whiteside *et al.* (2003) measured stop variability by the coefficient of variance, expressing the standard deviation as a function of the mean to control for the higher standard deviations that naturally accompany larger means. However, even using this measurement, the child participants here still have significantly more variable categories for all four phonemes (all $F > 6.47$, $p < 0.05$). We suspect that at ages 11–14, children might still be mastering the motor control required to produce adult-like speech consistency for all types of phonemes (Murdoch *et al.*, 2012; Walsh and Smith, 2002).

While the 9–10 yr old group had lower overall discriminability than adults, the oldest groups' categories were not only more dispersed but also much farther apart, compensating to produce discriminabilities equal to adults. Interestingly, because there were no discernable age-related trends in either the shortening of distance or the narrowing of dispersion between ages 9 and 14, this indicates that further development must take place after age 14, at which point each child learns to balance these two features independently and individually to achieve adult-like discriminability by their teenage years. Redford and Gildersleeve-Neumann (2009) examined preschool children's different speaking styles and observed a lack of vowel reduction when speaking casually. They proposed that because children were not expert speakers, they were unable to use efficient and reduced movements, resulting in an inability to take the articulatory shortcuts typically associated with adult casual speech. Something similar may be occurring here—because adolescents are less practiced speakers, their production is still more variable as evidenced by greater category dispersion. To maintain phonemic distinctions, they must increase category distances. However, as their motor-speech control continues to increase from ages 14 to 18 (Murdoch *et al.*, 2012; Walsh and Smith, 2002), they are expected to gain the ability to produce more consistent articulatory targets, thus decreasing category dispersion, which could engender a complementary decrease in category distance. Little research exists on the continued development of category structure between the ages of 14–18, and the present study highlights that this age range may provide further insights into the link between motor abilities and phonetic categories.

A second aim of the study was to investigate the effect of gender. Previous literature has shown an effect of gender on category distance and discriminability, although the age at which this becomes evident is not clear (Hazan and Baker, 2011; Fox and Nissen, 2005; Whiteside and Irving, 1998). This study therefore also sought to document the use of potential sociophonetic cues in this age group. Again in line with predictions, males overall had smaller distances between categories, and because they did not differ in dispersion from females, this resulted in overall lower discriminability scores for both adult and child males. For the fricative contrast, this supports previous findings of females producing more fronted /s/ articulations, resulting in higher in spectral peaks (Stuart-Smith, 2007; Fuchs and Toda, 2010) and higher fricative centroid values (Jongman *et al.*, 2000; Flipsen *et al.*, 1999; Fox and Nissen, 2005). This is not surprising because it is known from the sociophonetic literature that gender-marking is particularly strong for /s/ production (Fuchs and Toda, 2010; Heffernan, 2004; Stuart-Smith, 2007), and listeners have been shown to perceive ambiguous /s/-/ʃ/ tokens differently depending on whether they were accompanied by a male or female video (Strand, 1999, 2000).

Also of interest was the age at which this gender-marking for the alveolar fricative first appeared. For the group of SSBE speakers in this study, a marked shift appeared around age 11: while 9–10 yr old females produced higher /s/ and /ʃ/ centroids than their male counterparts, the categories of each were the same distance apart. From age 11, female children strikingly increased their centroid values for /s/—higher even than female adults' values—thus creating a between-category distance larger than males' that lasted through adulthood. Because other studies have found evidence for gender-marking for /s/ in younger children and even toddlers (e.g., Holliday *et al.*, 2010), such distinctions may become accentuated later in adolescence; however, it is difficult to compare findings directly, as methodologies and dialects vary.

Stops may exhibit a less clear patterning for gender, with studies reporting longer VOTs for /p/ only in women (Whiteside and Irving, 1998), longer VOTs for both stops (Whiteside *et al.*, 2004), or no gender difference for VOT (Morris *et al.*, 2008). Here for the stop contrast, female children had overall greater distances, which were achieved through longer /p/ VOTs. Because dispersion was the same as for male children, females produced greater category discriminability although without a clear developmental trend.

Finally, an important finding is the presence of category overlap within talkers. Though overlap would be more expected in casual speech, the present experimental setup elicited fairly formal speech, and still 18 (16%) of the speakers exhibited some amount of overlap in at least one

contrast—this was generally more pronounced for fricatives (at most 41% of a talker's tokens, average 12%) than for stops (at most 21% of a talker's tokens, average 7%), likely due to the clear temporal boundary in citation style stop voicing (Lisker and Abramson, 1967) for which there is no sibilant counterpart. Though the vast majority of these speakers were male children, again showing an effect of gender, overlap was also present in three adults—one male for fricatives, one male for stops, and one female for stops—therefore motor control cannot provide a complete explanation for less distinctive categories.

At present, there is little literature on the prevalence of consonant category overlap; however, it is generally considered quite rare in normal populations—Newman *et al.* (2001) only found 1 in 20 adults with sibilant overlap, while others have found minimal to no overlap within individual adult talkers (Haley *et al.*, 2010) and children aged 7–11 (Mandulak, 2009); however, these studies involved American dialects and smaller samples. Phoneme category overlap is much more common in individuals with abnormal speech, such as those with aphasia, apraxia, ALS, or cleft palates who have more physical motor trouble producing distinctions, showing that overlapping categories might be a sign of disordered speech (Haley *et al.*, 2010). With this thought, Haley and colleagues argue that the significant internal overlap that Newman *et al.* (2001) found in one normally speaking talker was likely a result of measurement error. Despite the robust spectral calculations here, using both multitaper spectra and windowing, we find even more individuals with category overlap, indicating that talkers with considerably non-discrete categories do indeed exist in the typical adult population—at least for SSBE speakers. Newman *et al.* further found that the more internally variable talkers—especially those with overlapping categories—were less readily intelligible to listeners. Along with the strong gender effect seen here, this suggests the frequent finding that women are more intelligible than men (Bradlow *et al.*, 1996; Hazan and Markham, 2004) may in part be due to females' greater category distance. The causes of this advantage are still poorly understood; by examining the interaction between dispersion and distance, the present study offers the beginning of an explanation of the acoustics behind the phenomenon.

Conflicting results between studies indicate acoustic-phonetic sources of talker intelligibility are likely to be multi-dimensional with possible trade-off effects between cues. These are likely to be driven by environmental, contextual, and listener factors, which intertwine to determine a talker's perspicuity (Hazan and Baker, 2011). What is more, simply describing categories through mean values cues is not sufficient either—developmental patterns might be overlooked without additional measures such as category range and discreteness (see Hitchcock and Koenig, 2013 for further discussion). Therefore a holistic approach to phoneme discriminability, including both internal and external category structure, is necessary to fully characterize the wide range of between- and within-talker variability and how this changes through the lifespan and ultimately affects intelligibility.

Allen, J. S., Miller, J. L., and DeSteno, D. (**2003**). "Individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **113**, 544–552.

Baker, R., and Hazan, V. (**2010**). "LUCID: A corpus of spontaneous and read clear speech in British English," in *Proceedings of DiSS-LPSS Joint Workshop*, pp. 3–6.

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.

Blacklock, O. S. (**2004**). "Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods," Ph.D. dissertation, University of Southampton.

Boersma, P. and Weenink, D. (**2012**). "Praat: Doing phonetics by computer (version 5.3.07) [computer program]," http://www.praat.org/ (Last viewed March 5, 2012).

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech. 1. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**, 804–809.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Otolaryngol. Suppl. **257**, 1–51.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H., and McSweeny, J. (**1999**). "Acoustic characteristics of /s/ in adolescents," J. Speech Lang. Hear. Res. **42**, 663–677.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–124.

Forster, K. I., and Forster, J. C. (**2003**). "DMDX: A windows display program with millisecond accuracy," Behav. Res. Methods Instrum. Comput. **35**, 116–124.

Fox, R. A., and Nissen, S. L. (**2005**). "Sex-related acoustic changes in voiceless English fricatives," J. Speech Lang. Hear. Res. **48**, 753–765.

Fuchs, S., and Toda, M. (**2010**). "Do differences in male versus female /s/ reflect biological or sociophonetic factors?," in *Turbulent Sounds. An Interdisciplinary Guide*, edited by S. Fuchs, M. Toda, and M. Zygis (Mouton de Gruyter, Berlin), pp. 281–302.

Haley, K. L., Seelinger, E., Mandulak, K. C., and Zajac, D. J. (**2010**). "Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach," J. Phonetics **38**, 548–554.

Hazan, V., and Baker, R. (**2011**). "Is consonant perception linked to within-category dispersion or across-category distance?," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 839–842.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Heffernan, K. (**2004**). "Evidence from HNR that /s/ is a social marker of gender," Toronto Work. Pap. Linguist. **23**, 71–84.

Hitchcock, E., and Koenig, L. L., (**2013**). "The effects of data reduction in determining the schedule of voicing acquisition in young children," J. Speech Lang. Hear. Res. **56**, 441–457.

Holliday, J. J., Beckman, M. E., and Mays, C. (**2010**). "Did you say susi or shushi? Measuring the emergence of robust fricative contrasts in English- and Japanese-acquiring children," in *Proceedings of INTERSPEECH 2010*, pp. 1886–1889.

Jongman, A. (**1986**). "Effects of speaking rate on the perception of syllable-initial stop consonants," J. Acoust. Soc. Am. **79**, S9.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

J. Acoust. Soc. Am., Vol. 134, No. 5, November 2013

Romeo *et al.*: Intra-talker variability   3791

Koenig, L. L. (**2001**). "Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends," J Speech Lang. Hear. Res. **44**, 1058–1068.

Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (**2013**). "Towards improved spectral measures of /s/: Results from adolescents," J Speech Lang. Hear. Res. **56**, 1175–1189.

Ladefoged, P. (**2003**). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques* (Blackwell Publishing, Oxford, UK), pp. 138–143.

Ladefoged, P. (**2005**). *Vowels and Consonants: An Introduction to the Sounds of Languages*, 2nd ed. (Blackwell Publishing, Oxford, UK), pp. 50–56.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Li, F. (**2012**). "Language-specific developmental differences in speech production: A cross-language acoustic study," Child Dev. **83**, 1303–1315.

Li, F., Edwards, J., and Beckman, M. E. (**2009**). "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," J. Phonetics **37**, 111–124.

Lisker, L. (**1978**). "In qualified defense of VOT," Lang. Speech **21**, 375–383.

Lisker, L. (**1986**). " 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**, 3–11.

Lisker, L., and Abramson, A. S. (**1967**). "Some effects of context on voice onset time in English stops," Lang. Speech **10**, 1–28.

Mandulak, K. C. (**2009**). "Production and perception of the voiceless sibilant fricatives in typically developing children with applications for children with cleft palate," unpublished Ph.D. dissertation, University of North Carolina at Chapel Hill.

Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**, 3962–3973.

McGowan, R. S., and Nittrouer, S. (**1988**). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/," J. Acoust. Soc. Am. **83**, 229–236.

McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**, 219–246.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2002**). "Gradient effects of within-category phonetic variation on lexical access," Cognition **86**, B33–B42.

Meyer, S., and Ptok, M. (**2011**). "The objective evaluation of sibilants," Laryngorhinootologie **90**, 753–759.

Morris, R. J., McCrea, C. R., and Herring, K. D. (**2008**). "Voice onset time differences between adult males and females: Isolated syllables," J. Phonetics **36**, 308–317.

Munson, B. (**2004**). "Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean," J. Speech Lang. Hear. Res. **47**, 58–69.

Murdoch, B. E., Cheng, H. Y., and Goozée, J. V. (**2012**). "Developmental changes in the variability of tongue and lip movements during speech from childhood to adulthood: An EMA study," Clin. Ling. Phon. **26**, 216–231.

Newman, R. S., Clouse, S. A., and Burnham, J. L. (**2001**). "The perceptual consequences of within-talker variability in fricative production," J. Acoust. Soc. Am. **109**, 1181–1196.

Nissen, S. L., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**, 2570–2578.

Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**, 520–530.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception." Perc. Psych. **60**, 355–376.

Ohde, R. N. (**1985**). "Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children," J. Acoust. Soc. Am. **78**, 1554–1561.

Redford, M. A., and Gildersleeve-Neumann, C. E. (**2009**). "The development of distinct speaking styles in preschool children," J Speech Lang. Hear. Res. **52**, 1434–1448.

Reidy, P. F. (**2013**). "An introduction to random processes for the spectral analysis of speech data," in *Ohio State University Working Papers in Linguistics*, pp. 67–116.

Reidy, P. F., and Beckman, M. E. (**2012**). "The effect of spectral estimator on common spectral measures for sibilant fricatives," in *Proceedings of INTERSPEECH 2012*, pp. 1516–1519.

Shadle, C. H. (**2012**). "The acoustics and aerodynamics of fricatives," in *The Oxford Handbook of Laboratory Phonology*, edited by A. Cohn, C. Fougeron, and M. K. Hoffman (Oxford University Press, Oxford, UK), pp. 511–526.

Simpson, A. J., and Fitter, M. J. (**1973**). "What is the best index of detectability?," Psychol. Bull. **80**, 481–488.

Smiljanic, R. and Bradlow, A. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Linguist. Compass. **3**, 236–264.

Strand, E. (**1999**). "Uncovering the role of gender stereotypes in speech perception," J. Lang. Soc. Psych. **18**, 86–99.

Strand, E. (**2000**). "Gender stereotype effects in speech processing," unpublished dissertation. Ohio State University.

Stuart-Smith, J. (**2007**). "Empirical evidence for gendered speech production: /s/ in Glaswegian," in *Laboratory Phonology 9. Phonology and Phonetics*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, New York), pp. 65–86.

Sumner, M. (**2011**). "The role of variation in the perception of accented speech," Cognition **119**, 131–136.

Theodore, R. M., Miller, J. L., and DeSteno, D. (**2009**). "Individual talker differences in voice-onset-time: Contextual influences," J. Acoust. Soc. Am. **125**, 3974–3982.

Theodore, R. M., Miller, J. L., and DeSteno, D. (**2010**). "Characteristics of listener sensitivity to talker-specific phonetic detail," J. Acoust. Soc. Am. **128**, 2090–2099.

Theodore, R. M., Myers, E. B., and Lomibao, J. (**2013**). "Listeners' sensitivity to talker differences in voice-onset-time: Phonetic boundaries and internal category structure," POMA **19**, 060248.

Thomson, D. J. (**1982**). "Spectrum estimation and harmonic analysis," Proc. IEEE **70**, 1055–1096.

Todd, A. E., Edwards, J. R., and Litovsky, R. Y. (**2011**). "Production of contrast between sibilant fricatives by children with cochlear implants," J. Acoust. Soc. Am. **130**, 3969–3979.

van Dommelen, W., and Hazan, V. (**2012**). "Impact of talker variability on word recognition in non-native listeners," J. Acoust. Soc. Am. **132**, 1690–1699.

Walsh B., and Smith, A. (**2002**). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," J. Speech Lang. Hear. Res. **45**, 1119–1133.

Whiteside, S. P., Dobbin, R., and Henry, L. (**2003**). "Patterns of variability in voice onset time: A developmental study of motor speech skills in humans," Neurosci. Lett. **347**, 29–32.

Whiteside, S. P., Henry, L., and Dobbin, R. (**2004**). "Sex differences in voice onset time: A developmental study of phonetic context effects in British English," J. Acoust. Soc. Am. **116**, 1179–1183.

Whiteside, S. P., and Irving, C. (**1998**). "Speakers' sex differences in voice onset time: A study of isolated word production," Percept. Mot. Skill **86**, 651–654.

Whiteside, S. P., and Marshall, J. (**2001**). "Developmental trends in voice onset time: Some evidence for sex differences." Phonetica **58**, 196–210.