# The acoustic bases for gender identification from children's voices

Theodore L. Perry, Ralph N. Ohde,[a)] and Daniel H. Ashmead
*Department of Hearing and Speech Sciences, Vanderbilt Bill Wilkerson Center for Otolaryngology and Communication Sciences, Vanderbilt University Medical Center, Nashville, Tennessee 37232*

(Received 20 August 1999; revised 28 February 2001; accepted 14 March 2001)

The purpose of this study was to examine the acoustic characteristics of children's speech and voices that account for listeners' ability to identify gender. In Experiment I, vocal recordings and gross physical measurements of 4-, 8-, 12-, and 16-year olds were taken (10 girls and 10 boys per age group). The speech sample consisted of seven nondiphthongal vowels of American English (/æ/ ''had,'' /ɛ/ ''head,'' /i/ ''heed,'' /I/ ''hid,'' /ɑ/ ''hod,'' /ʌ/ ''hud,'' and /u/ ''who'd'') produced in the carrier phrase, ''Say /hVd/ again.'' Fundamental frequency ($f_0$) and formant frequencies ($F1$, $F2$, $F3$) were measured from these syllables. In Experiment II, 20 adults rated the syllables produced by the children in Experiment I based on a six-point gender rating scale. The results from these experiments indicate (1) vowel formant frequencies differentiate gender for children as young as four years of age, while formant frequencies and $f_0$ differentiate gender after 12 years of age, (2) the relationship between gross measures of physical size and vocal characteristics is apparent for at least 12- and 16-year olds, and (3) listeners can identify gender from the speech and voice of children as young as four years of age, and with respect to young children, listeners appear to base their gender ratings on vowel formant frequencies. The findings are discussed in relation to the development of gender identity and its perceptual representation in speech and voice.  [DOI: 10.1121/1.1370525]

PACS numbers: 43.70.Ep, 43.71.Bp [AL]

## I. INTRODUCTION

Differences between male and female speech and voice are an important aspect of gender identity. These differences are in part a result of anatomical structures such as vocal fold size and vocal tract length as well as learned characteristics of vocal production such as intonation contour (Cruttenden, 1986; Crystal, 1982; Ohde and Sharf, 1992). Thus, prosodic features that are overlaid (suprasegmentals) upon sound segments in words, phrases, or sentences and include intonation, stress, duration, and juncture may be important in gender identification. It is well known that a typical $f_0$ for an adult male is around 120 Hz, while a typical $f_0$ for an adult female is around 200 Hz. Moreover, an average adult male will have lower formant frequencies than an average adult female, because of longer vocal tracts in the former than latter speakers (Ladefoged and Broadbent, 1957; Laver and Trudgill, 1979; Peterson and Barney, 1952; Lee, Potamianos, and Narayanan, 1999; Tecumseh Fitch and Giedd, 1999).

The acoustic differences between male and female voices are also influenced by behavioral factors. It has been noted that adult males will often speak with an unnaturally lower vocal pitch and women will often speak with an unnaturally higher vocal pitch in order to conform to stereotypical views of vocal production characteristics (Sachs, Lieberman, and Erickson, 1973). In studies of the perception of maleness or femaleness in a transsexual's voice, it was demonstrated that listeners were more likely to rate a voice as being female sounding with increasing variability in the intonation contour (i.e., the relative rising and falling of $f_0$) of the speech sample (Oates and Dacakis, 1983; Spencer, 1988; Wolfe, Ratusnik, Smith, and Northrop, 1990).

There have been several previous investigations of gender-specific vowel formant frequency characteristics of preadolescent children (Eguchi and Hirsh, 1969; Hasek, Singh, and Murry, 1980; Bennett, 1981; Busby and Plant, 1995; Lee, Potamianos, and Narayanan, 1999). Bennett (1981) analyzed the vocal productions of 7- to 8-year-old children for five vowels in a fixed phonetic context. The vowel resonances of boys were consistently lower than those of the girls, and several measures of body size were related to formant frequencies. The findings were consistent with the conclusion that boys had larger vocal tracts than girls. Busby and Plant (1995) examined the acoustic features of Australian vowels produced by preadolescent children. Five boys and five girls from each of four age groups (5-, 7-, 9-, and 11-year olds) vocalized 11 test words representing the nondiphthong vowels of Australian English. Fundamental frequency, the first three formant frequencies, amplitudes of the first three formants, and vowel duration were measured. The results revealed that there was no difference in $f_0$ between boys and girls within each age group, the values of the first three formant frequencies for the girls were higher than those for the boys, the formant amplitudes for the boys were higher than those for the girls, and there was no consistent variation in vowel duration values across age and/or gender. For $f_0$ and formant frequencies ($F1$,$F2$), Eguchi and Hirsh (1969) reported similar gender results for boys and girls between 11 and 13 years of age. Overall, these findings are consistent with the notion that boys have larger vocal tracts than girls. The specific age at which this distinction occurs is controversial (Lee *et al.*, 1999).

[a)]Electronic mail: ohdexxrn@ctrvax.vanderbilt.edu

The above studies show that there are differences between boys and girls, particularly in resonance characteristics of vowels. Because of these production differences, there has been an interest in determining their role in perceiving gender from children's speech and voice. Weinberg and Bennett (1971) presented adult listeners with recorded samples of spontaneous speech produced by 5- and 6-year-old children. The listeners correctly identified speaker gender for 78% of the boys and 71% of the girls. An acoustic analysis revealed a large overlap in the $f_0$ ranges of the boys and girls. This result suggested that gender identification of children's speech and voice must be based not only on fundamental frequency, but on other acoustic properties as well. Sachs, Lieberman, and Erickson (1973) studied the speech and voice of 26 preadolescent children (4–14 years). Adult listeners made correct gender identifications of 81% of the speakers. Acoustically, the boys had a higher average $f_0$ than the girls, although the boys did have lower resonance frequencies than the girls. Due to the counterintuitive finding that the boys had a higher average $f_0$ than the girls, Sachs *et al.* (1973) suggested that gender identifications were based on the differences in the formant frequencies between the boys and the girls.

Other research has also suggested that, although $f_0$ provides the most dominant cue for gender identification from the voices of adults, other vocal properties may be important, especially in children. For example, listeners are able to identify gender: (1) when $f_0$ is filtered (e.g., Lass, Almerino, Jordan, and Walsh, 1980) or suppressed (e.g., Coleman, 1971, 1973) within the speech signal, (2) when speech samples are whispered (e.g., Bennett and Weinberg, 1979), and (3) when the frequency range of $f_0$ is very similar, like that of preadolescent boys and girls (e.g., Bennett, 1981; Ingrisano, Weismer, and Schuckers, 1980). In light of this evidence, it appears that adults are accustomed to using acoustic properties other than $f_0$ alone as a basis for gender identification. In addition to understanding the development of gender, there is recent information underscoring the importance of the voice in processing phonetic (vowels and consonants) information (Johnson, 1990a, 1990b; Haley and Ohde, 1996). Moreover, it has been claimed that the perceptual representation of gender is auditory based and qualitatively different from phonetic information (Mullennix *et al.*, 1995).

Although previous research has provided information on the importance of $f_0$ and formant frequencies in the acoustic and perceptual differentiation of gender, there is no study employing a consistent production/perceptual methodology to examine gender development in production, along with an adults' perception of gender throughout the period ranging from pre- to post-adolescence. For example, studies assessing an adults' perception of gender frequently have employed sentential material (Bennett and Weinberg, 1979; Ingrisano *et al.*, 1980), which could provide listeners with gender-specific prosodic cues other than $f_0$ and formant frequencies (Wolfe *et al.*, 1990). Since formant frequencies may contribute to gender differentiation, there is a potential relationship between vocal tract resonances and indices of body size, even for very young children. Previous findings show such a relationship between measures of body size and formant frequencies, but these findings were limited to 7- to 8-year-old children (Bennett, 1981). Because the perception of gender appears to depend on several acoustic factors, including $f_0$, formant frequencies, and breathiness (Klatt and Klatt, 1990), an evaluation of these properties in development is important for an understanding of both gender and the potential integration of source ($f_0$) and transfer function (formant frequencies) as cues to gender identification. Thus, the general purpose of this research was to examine $f_0$ and formant frequencies to assess their developmental role in the production and perception of gender. The two experiments were designed to answer the following questions.

(1) Do $f_0$ and vowel formant frequencies differentiate gender in 4-, 8-, 12-, and 16-year-old boys and girls?
(2) Is there a developmental relationship between gross measures of physical size and the vocal characteristics of these children?
(3) Is there a perceptual differentiation of gender between 4 and 16 years of age that parallels differences in production?

## II. EXPERIMENT I: ACOUSTIC PROPERTIES OF GIRLS' AND BOYS' SPEECH AND VOICE

Anatomical properties of the vocal tract as well as learned speech characteristics influence the identification of gender from speech and voice. These speech characteristics include changes in mouth opening, lip rounding, and larynx height. The acoustic parameters considered most influential are $f_0$ and vocal tract resonances (Abercrombie, 1967; Eguchi and Hirsh, 1969; Kent, 1976; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Nolan, 1983). The focus of experiment I is on the salience of $f_0$ and formant frequencies as acoustic predictors of gender.

### A. Method

#### *1. Subjects*

Twenty children in each of the age groups, 4-, 8-, 12-, and 16-year olds, participated with 10 girls and 10 boys per age group. The mean (and age range) ages (years:months) were 4:3 (4:1 to 4:8), 8:7 (8:5 to 9:5), 12:4 (12:2 to 12:9), and 16:4 (16:3 to 16:10) years. Age ranges computed for boys and girls separately were very similar, and varied between 1 to 4 months. All children spoke Standard American dialect, and none had a known speech disorder. A hearing screening at 20-dB HL for the octave frequencies between 500 and 4000 Hz was performed on children.

#### *2. Speech sample and recording procedure*

The speech sample consisted of seven nondiphthongal vowels of American English (/æ/ ''had,'' /ɛ/ ''head,'' /i/ ''heed,'' /I/ ''hid,'' /ɑ/ ''hod,'' /ʌ/ ''hud,'' and /u/ ''who'd'') produced in the neutral context of /hVd/. Each /hVd/ syllable was embedded in the carrier phrase, ''Say /hVd/ again.'' The children listened to a tape recording of an adult female vocalizing each syllable seven times within the carrier phrase according to a randomized schedule. There was a 3-s interval between presentations of the syllables, during which the children repeated the phrase last heard. For each child, five of the seven syllables produced for each vowel were chosen as token utterances by the experimenter on the basis of clarity of recording and correctness of the vowel sound. All speech recordings were made at the children's schools in a portable sound-treated booth (noise level approximately 35–40 dBA) using a Tascam reel-to-reel tape recorder (Model 22-2) and a Signet unidirectional electret condenser recording microphone (Model RK 201). The microphone was placed in a microphone stand and positioned 8–10 in. from the speaker's mouth. The recording procedure took approximately 15–20 min for each child to complete.

#### *3. Perceptual verification of vowels*

Perceptual verification tests of the speech samples were conducted to determine that the vowels were consistent with their intended targets. Five phonetically experienced adult native speakers of English with no known speech or hearing disorder identified the 2800 syllables in random order from all experimental syllables ($5\,\text{syllables}\times 7\,\text{vowels}\times 80$ speakers). A criterion of 80% agreement (four out of five listeners) was used to consider the vowels as judged to be the intended targets. One-hundred fourteen of the syllables were judged at 80% agreement, while the remainder of the syllables were judged at 100% agreement. Consequently, all of these syllables were used.

#### *4. Physical measurements*

Body size measurements of the children similar to those obtained by Bennett (1981) were taken. Standing height, sitting height, body weight, and neck circumference were measured with the children's shoes removed. Standing height and body weight were measured using a medical scale, while sitting height and neck circumference were measured by a tape measure. Sitting height was measured as children sat on a stool. The distance from the seat of the stool to the vertex of the head was measured. Neck circumference was measured at the angle of the thyroid cartilage. This point was determined by palpation, and the measurement was taken just above the thyroid (laryngeal) prominence.

#### *5. Formant frequency and fundamental frequency analyses*

The speech samples were analyzed on a microcomputer system (CSL—Computerized Speech Lab, Model 4300, Software Version 5.0, Kay Elemetrics Corp.) at a 20-kHz sampling rate. Each syllable was isolated and removed from the sentence, ''Say /hVd/ again,'' by determining the onset of aspiration in /h/ and the point of the stop-gap in /d/. This was accomplished by producing a spectrogram of the sentence and segmenting the syllable from the phrase. In order to ensure capture of the entire vowel, segmentation was made 5–10 ms into the stop-gap of each syllable. Measurements of formant frequencies of children's speech can be difficult because of their high $f_0$'s and unusual voice types (Kent, 1978). To minimize difficulties in formant frequency measurement, an analysis procedure similar to Sussman, Minifie, Buder, Stoel-Gammon, and Smith (1996) was adopted. $F1$, $F2$, and $F3$ were measured from the spectrogram of each syllable at the duration midpoint of the vowel. Furthermore, LPC (linear predictive coding) analyses were obtained with a 10-msec Triangular window at approximately the same point at which the formant values were determined from the spectrogram by placing the cursor at the midpoint of $F2$ stability. In the LPC analysis, the number of coefficients employed varied between 14 and 20 to derive the best estimate of formant frequencies. Both formant values from the spectrogram and the LPC analysis were recorded, and the specific frequency was determined as the average for a valid criterion measure (Sussman *et al.*, 1996). There was no more than 150-Hz difference between the values recorded from the spectrogram and the LPC analysis.

Fundamental frequency of the vowel was determined with the CSL pitch extraction program. Fundamental frequency is computed as the inverse of the time between glottal impulse markers. In addition to the absolute formant and $f_0$ frequency values, frequency scale factors (K factors) were used to show the percentage relationship of male and female $f_0$ and formant frequencies (Fant, 1973).

#### *6. Statistical analyses*

Since the major emphasis in Experiment I was to measure acoustic differences in children's speech as a function of gender, these data were analyzed using repeated measures analysis of variance (ANOVA), simple correlations, and multiple regression analysis (Kirk, 1982). All *post-hoc* tests were made using Tukey HSD comparisons at $p<0.05$. Separate analyses were conducted on data for $f_0$, $F1$, $F2$, and $F3$ as well as measures of physical size.

Several vowels were included in these analyses to represent the comprehensive range of productions in the oral cavity. Based on previous research (Bennett, 1981), gender differences for specific formants were likely to interact with specific vowel productions. However, the overall gender effects were of greatest interest, and the interactions of gender and vowel type entail complexities that do not have a clear theoretical basis. Therefore, the statistical interactions involving a vowel are reported, but not examined in detail.

### B. Results and discussion

#### *1. Reliability measures*

Interjudge reliability was determined by having a second examiner sophisticated in acoustic analyses measure 10% of the speech sample. All four age groups were represented in these analyses. An equal number of samples from boys and girls were analyzed. The speakers and vowels were randomly selected. Interjudge reliability was high for both $f_0$ and formant frequency measures. Correlations of the two experimenters were computed for the two measures. The mean and range of the $f_0$ correlations across the speaker age groups were 0.91, and 0.86–0.99, respectively. The mean and range of the formant frequency correlations were 0.93, and 0.87–0.97, respectively.

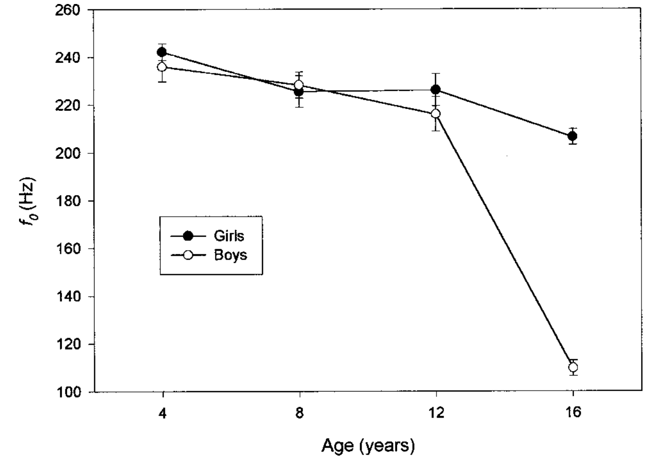


FIG. 1. Mean fundamental frequency ($f_0$) across the CVC syllables for 4-, 8-, 12-, and 16-year-old speakers. Error bars show $\pm 1$ standard error.

### 2. Fundamental frequency

The $f_0$ means averaged across vowel type are illustrated in Fig. 1. There were gender differences between the $f_0$ values of girls and boys only in the 16-year-old age group. In an Age (4) $\times$ Gender (2) $\times$ Vowel (7) analysis of variance, there were main effects of age [$F(3,72)=86.96$; $p<0.0001$], gender [$F(1,72)=50.44$; $p<0.0001$], vowel [$F(6,432)=10.27$; $p<0.0001$], and a significant interaction of age$\times$gender [$F(3,72)=35.72$; $p<0.0001$]. Within each age group, the only significant gender difference was for the 16-year olds.

The finding that $f_0$ was similar for girls and boys at 4, 8, and 12 years is consistent with previous research (e.g., Bennett, 1983; Busby and Plant, 1995; Kent, 1976), although Hasek *et al.* (1980) found gender differences in $f_0$ for 7- to 10-year-old children. Nevertheless, the data from the current study suggest that there is little or no difference in $f_0$ between boys and girls 12 years old or younger. Therefore, it is unlikely that $f_0$ provides sufficient information to differentiate speaker gender in young children. Because most acoustic studies of children's speech have shown little difference in $f_0$ between boys and girls under 13 years of age, it appears that vocal fold size increases modestly throughout middle childhood (Kent, 1976; Hollien, Green, and Massey, 1994; Kent and Vorperian, 1995).

K factors were calculated for these children to determine the pattern of gender differences in $f_0$ and formant frequencies using the formula $\text{Kn}\% = 100[(\text{Fn female}/\text{Fn male}) - 1]$ where Fn is either $f_0$, $F1$, $F2$, or $F3$ (Fant, 1973; Bennett, 1981; Nordstrom, 1997). These values represent the percentage difference between the vocal and speech characteristics of the boys and girls. K factors were calculated for mean $f_0$, $F1$, $F2$, and $F3$ values within each age group for the boys and girls referred to as K0, K1, K2, and K3, respectively.

The extent of gender differences in $f_0$ within the younger age groups was very small (K0: 2.6% for 4-year olds, $-1.2\%$ for 8-year olds, and 4.7% for 12-year olds). The negative value for the 8-year olds reflects a slightly lower mean $f_0$ for the girls than the boys. However, the difference between $f_0$ of the 16-year-old boys and girls was substantial (88.2%).

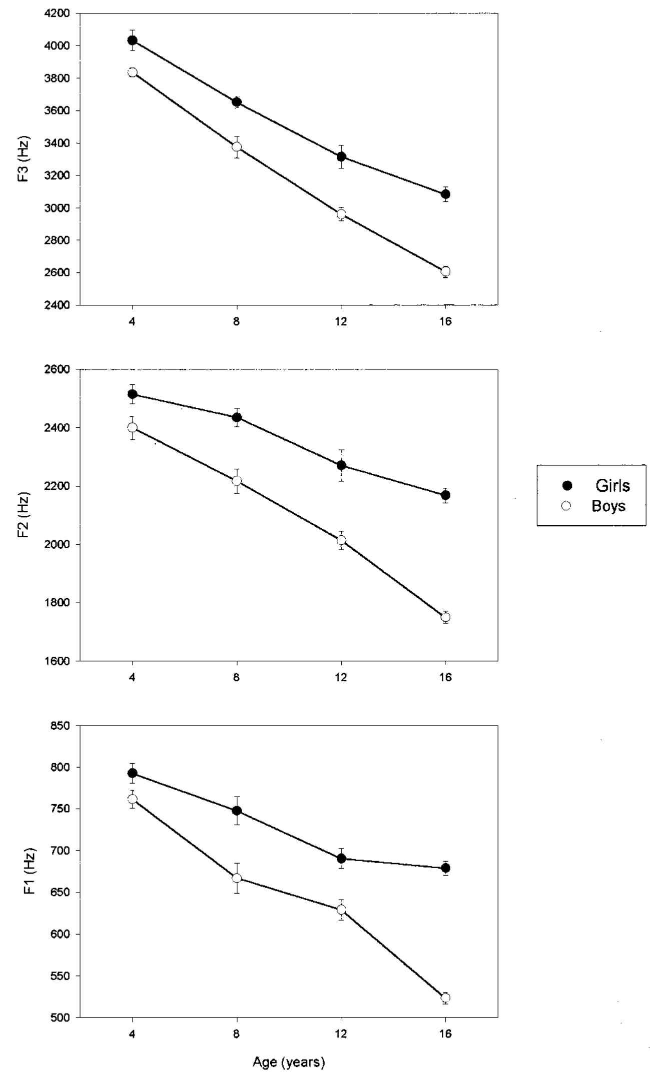


FIG. 2. Mean $F1$, $F2$, and $F3$ values across the CVC syllables for 4-, 8-, 12-, and 16-year-old speakers. Error bars show $\pm 1$ standard error.

### 3. Vowel formant frequency

The above analyses suggest that $f_0$ does not provide the information necessary to differentiate gender until after 12 years of age. However, gender differences may exist in the vowel formant frequencies of these children. The three panels of Fig. 2 illustrate the $F1$, $F2$, and $F3$ means collapsed across the vowels. Separate Age(4)$\times$Gender(2)$\times$Vowel(7) analyses of variance were performed on the $F1$, $F2$, and $F3$ values. The results of the analysis for $F1$ values showed main effects of age [$F(3,72)=69.74$; $p<0.0001$], gender [$F(1,72)=84.45$; $p<0.0001$], and vowel [$F(6,432)=1008.19$; $p<0.0001$]. A significant age$\times$gender interaction was found [$F(3,72)=8.87$; $p<0.0001$].[1] The analysis for $F2$ values showed the main effects of age [$F(3,72)=73.15$; $p<0.0001$], gender [$F(1,72)=98.20$; $p<0.0001$], and vowel [$F(6,432)=980.69$; $p<0.0001$], and a significant age$\times$gender [$F(3,72)=6.16$; $p<0.001$] interaction.[2] The analysis for $F3$ values showed the main effects of age [$F(3,72)=175.70$; $p<0.0001$], gender [$F(1,72)=84.18$; $p$

<0.0001], and vowel [$F(6,432)=167.94$; $p<0.0001$] and a significant age×gender [$F(3,72)=2.78$; $p<0.05$] interaction.

According to Tukey tests, $F1$ values for the boys were significantly lower than those for the girls within each age group except the 4-year olds. The same pattern held for the $F2$ values. $F3$ values for the boys were significantly lower than for the girls within each age group (Tukey HSD, $p<0.05$). These results suggest that there may be adequate information available from $F1$, $F2$, and $F3$ values to differentiate gender of the 8-, 12-, and 16-year olds. Differences between $F1$ and $F2$ values for the 4-year olds were not significant. However, since $F3$ differences between 4-year-old boys and girls were significantly different, it is conceivable that this information may also be sufficient to differentiate gender.

Computation of K factors in this study revealed that the percentage of differences between boys and girls in $F1$ (K1: 2.6%, 11.4%, 9.1%, and 27.5% for 4-, 8-, 12-, and 16-year olds, respectively), $F2$ (K2: 4.6%, 10.3%, 12.5%, and 23.8% for 4-, 8-, 12-, and 16-year olds, respectively) and $F3$ (K3: 5.2%, 8.3%, 12.0%, and 18.4% for 4-, 8-, 12-, and 16-year olds, respectively) values within each age group remained very similar, and these formant frequencies were lower for the boys than for the girls. Busby and Plant (1995) also showed that vowel formant frequencies were consistently lower for boys than girls, and that $F1$, $F2$, and $F3$ decreased in value with age. Though Bennett (1981) only examined 7- and 8-year olds, she found that vowel formant frequencies for the boys were, on average, 10% lower than those for the girls. Similar data from the 8-year olds in the present study revealed that the boys' formant frequencies were approximately 9% lower than those for the girls. Thus, gender differences in formant frequencies increase with age and are apparent, even in very young children. If gender identification from young children's speech is possible, then listeners may utilize this difference in formant frequencies.

### 4. Relationship between physical size and formant frequencies

Bennett (1981) suggested that gender differences in formant frequencies of children may be attributable to vocal tract size, and in particular the pharynx (Kent and Vorperian, 1995). Bennett does not define vocal tract size, but it is reasonable to assume that the length and width of the vocal tract would vary as a function of gross changes in body size. For purposes of this paper, size will include the length and width of the vocal tract. Because it was not possible to obtain exact measurements of vocal tract size, various measurements of physical size were taken. Findings for height, weight, sitting height, and neck circumference are contained in Table I. The difference between the physical size of boys and girls increased with age but became readily apparent only at 16 years of age. On average, the 4-, 8-, and 12-year-old boys were 2 cm taller, 2 kg heavier, and had neck circumferences 2 cm larger than the girls. Separate analyses of variance (age×gender) were carried out for each of the four physical size measures. As expected, all four analyses showed significant main effects of age and gender, as well as significant interactions of age and gender. Since these results were highly predicted based on anatomical development (Kent and Vorperian, 1995), the details of these analyses are not reported. These analyses were run principally to provide the error terms used for Tukey tests of gender differences at each age level. The pattern of the Tukey tests was simple. At age 16 years, boys were significantly larger than girls on all four measures. At age 12 years, boys had significantly larger neck circumferences than girls. None of the other comparisons of boys versus girls within age groups were significant. The measurements employed in the current research are similar to Bennett's (1981) measurements for 7- and 8-year olds (mean age=7 years 11 months), though her measurements showed that the boys were slightly taller (7 cm) and heavier (4 kg) than the girls in her study.

TABLE I. Means and standard deviations (SD) for measures of physical size for 4-, 8-, 12-, and 16-year-old males and females (H=height, W=weight, SH=sitting height, and NC=neck circumference).

| Age group | H (cm) | W (kg) | SH (cm) | NC (cm) |
|---|---|---|---|---|
| 4 years | | | | |
| Male | | | | |
| mean | 106.8 | 17.8 | 60.1 | 26.1 |
| SD | 2.8 | 1.2 | 2.6 | 1.2 |
| Female | | | | |
| mean | 106.9 | 17.9 | 58.6 | 24.9 |
| SD | 5.1 | 2.5 | 3.6 | 1.1 |
| 8 years | | | | |
| Male | | | | |
| mean | 134.9 | 34.4 | 74.3 | 29.0 |
| SD | 8.0 | 9.1 | 4.9 | 1.8 |
| Female | | | | |
| mean | 132.8 | 33.8 | 73.7 | 27.9 |
| SD | 4.7 | 8.2 | 2.9 | 2.1 |
| 12 years | | | | |
| Male | | | | |
| mean | 158.4 | 49.7 | 83.9 | 32.2 |
| SD | 5.9 | 5.8 | 3.9 | 1.6 |
| Female | | | | |
| mean | 155.5 | 44.1 | 84.7 | 29.4 |
| SD | 6.1 | 4.6 | 3.5 | 1.5 |
| 16 years | | | | |
| Male | | | | |
| mean | 179.8 | 79.1 | 96.1 | 38.0 |
| SD | 4.8 | 10.2 | 2.7 | 2.9 |
| Female | | | | |
| mean | 164.7 | 60.8 | 90.1 | 31.4 |
| SD | 6.1 | 7.6 | 2.5 | 1.2 |

Figure 3 shows scatter plots of the relationship between body weight and the formant measures. The measures of body size were highly correlated with one another and all showed similar relationships with formant values, so only body weight is shown for illustrative purposes. The formant values are averaged across all speech tokens for each individual child speaker. The symbols in Fig. 3 indicate the age and gender of each child. Linear regression lines are shown separately for the boys and girls (each regression was computed on all 40 children of a given gender, that is, on children from all four age groups). There are two points to emphasize with respect to Fig. 3. First, physical size is strongly

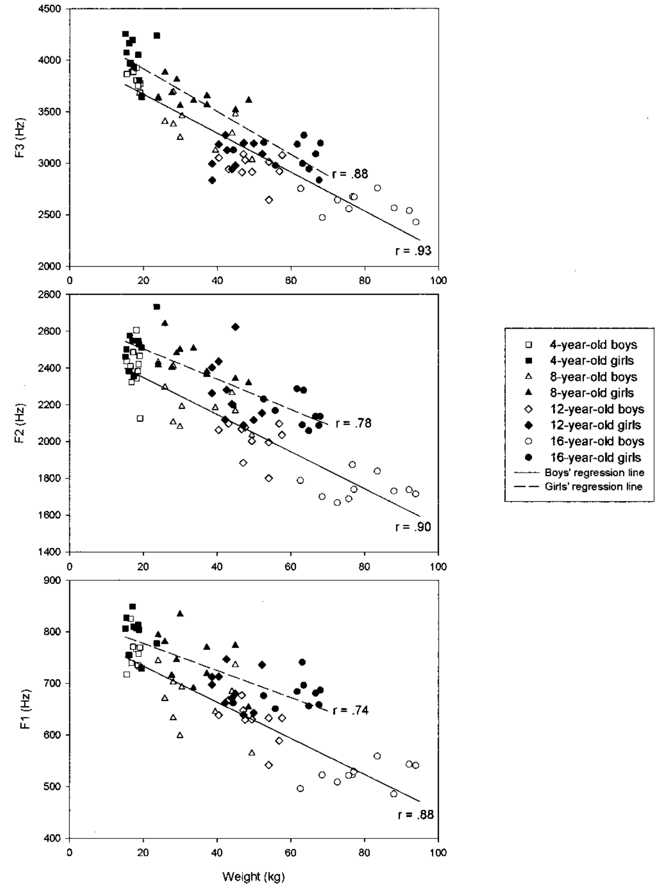


FIG. 3. Average formant frequencies versus body weight for individual children.

correlated with formant frequencies, as would be expected if vocal tract size grows along with grosser measures of size such as body weight. All the correlations indicated in Fig. 3 were statistically significant, and the slope values for boys and girls were similar, as can be seen in the figure. The second point about Fig. 3 is that there are gender differences in formant frequencies even when physical size is taken into account. The girls' formant frequencies tended to be higher than the boys,' especially for $F1$ and $F2$.

This gender difference in formant frequencies was further analyzed with stepwise multiple regression. The logic of each regression analysis, which was done separately for $F1$, $F2$, and $F3$, was to enter predictors of the formant frequency in the following order: physical size measures (body weight, neck circumference, sitting height, standing height); age group (4, 8, 12, or 16 years); and gender (boy, girl). If gender has a unique effect on formant frequencies, over and above physical size and age, then the addition of gender as a predictor on the last step should produce a significant increase in the $R^2$ value. The gender-related increase in $R^2$ was significant for all three formants: [$F1$—$R^2$ increment$=0.054$, $F(1,73)=21.052$; $p<0.0001$], [$F2$—$R^2$ increment$=0.088$, $F(1,73)=35.484$; $p<0.0001$], and [$F3$—$R^2$ increment $=0.049$, $F(1,73)=43.645$; $p<0.0001$]. The overall $R^2$ values with all predictors in the model were 0.902, 0.905, and 0.918 for $F1$, $F2$, and $F3$, respectively, so the variance in formant frequencies was accounted for quite well by these predictors. In summary, formant frequencies were strongly correlated with body size measures for both boys and girls, but there were gender differences in formant frequencies, even after the body size measures used in this study were taken into account. These remaining gender differences could reflect differences in vocal tract size between boys and girls that were not captured by the body size measures taken, but the difference could also reflect variations in articulatory postures of lip rounding and jaw opening for boys and girls (Lindblom and Sundberg, 1971; Fant, 1973; Bennett and Weinberg, 1979; Bennett, 1981).

## III. EXPERIMENT II: PERCEPTION OF GIRLS' AND BOYS' SPEECH AND VOICE

The acoustic data from Experiment I showed that $f_0$ was only slightly different between the 4-, 8-, and 12-year-old boys and girls, while $f_0$ was significantly lower for the 16-year-old boys than for the 16-year-old girls. However, all formant frequency values were lower for the 8-, 12-, and 16-year-old boys than for the girls. As for the 4-year olds, there were only small gender differences in $F1$ and $F2$ values, but $F3$ was lower for the boys than for the girls. Because acoustic gender differences were observed in all age groups, it was of interest to determine if adult listeners could differentiate gender from the CVC syllables, and to see whether this identification varied with speaker age. Since the magnitude of acoustic differences as a function of gender differed across age groups, it was important to employ a scaling technique, which would be a more sensitive identification procedure than a two category response choice.

### A. Method

#### *1. Subjects and listening procedure*

The subjects for this experiment were 20 undergraduate students (equal numbers of men and women) with no known speech or hearing disorders by self-report. All subjects spoke general American dialect.

Perceptual tests were performed in an IAC booth over headphones (TDH-49) at a comfortable listening level (75-dB SPL). Before beginning a series of trial blocks, listeners were informed by the experimenter that they would hear speech sounds from either 4-, 8-, 12-, or 16-year-old boys and girls and that they would make gender ratings of each vocalization based on a six-point scale described below. Though listeners were not told the age of the speakers to be heard, they were informed that all voices in a single session were from the same age group. On-line data collection as well as the generation of random orderings of the vowel sounds within each trial block was performed by a computer. Listeners were not given any feedback as to the actual gender of the speakers.

Listeners used a six-point rating scale to indicate gender. Ratings were entered via a six-button response box. The categories were as follows:

1 positively a female;

2 appeared to be a female;

3 unsure, may have been a female;

4 unsure, may have been a male;

5 appeared to be a male;

6 positively a male.

The rating scale numbers were written on the response box, and a copy of the rating scale was always within view. This scale allowed listeners to make judgments based upon their certainty of gender, unlike previous studies that permitted only binary male and female responses (Bennett and Weinberg, 1979; Ingrisano *et al.*, 1980; Sachs *et al.*, 1973; Weinberg and Bennett, 1971).

The events in a trial consisted of a 300-ms warning light, a 300-ms delay, the stimulus, and a response light indicating that it was time to make a rating. There was a four-second interval between trials. Subjects were presented five trial blocks per session, with all five trial blocks consisting of stimuli from only one speaker age group. There were four listening sessions, with the order of speaker age group counterbalanced (each listener had a different one of the 24 possible orders). No more than one session was conducted per day, and there were no more than ten days between sessions.

The stimuli consisted of the CVC syllables from Experiment I. Each of the seven vowel sounds was produced five times by the 20 speakers within each of the four age groups (4-, 8-, 12-, and 16-year olds), yielding a total of 2800 syllables. There were 700 stimuli for each of the four age groups. These 700 stimuli were divided into 5 trial blocks per speaker group. Within each age group, a trial block consisted of the 140 utterances produced by 20 children, each saying 7 vowel sounds once. For reliability purposes, four of the utterances within each vowel sound were presented twice. Thus, each of the 5 trial blocks consisted of 168 trials, 140 trials corresponding to seven vowels from each of 20 children in a single age group plus 28 repetition trials.

### 2. Reliability

In order to determine the test–retest reliability, one syllable per speaker was chosen at random for each of the seven vowel sounds within each speaker age group for a total of 560 (80 speakers$\times$7 vowels) reliability tokens for each listener. Kappa (Cohen, 1969; Fleiss, 1981) was used to determine the chance-corrected reliability of the gender ratings for the two presentations of the same stimuli for the listeners within each trial block. Two categories were used in this analysis. A rating of ''1,'' ''2,'' or ''3'' was considered as ''female sounding,'' and a rating of ''4,'' ''5,'' or ''6'' was considered as ''male sounding.'' The kappa value was 0.64, which according to Fleiss (1981) reflects good agreement.

## B. Results and discussion

### 1. Gender ratings

Previous data suggested that female listeners were better than male listeners in determining gender from children's speech and voice (Ingrisano *et al.*, 1980). Preliminary analyses of the data in which ratings were collapsed categorically as female (''1,'' ''2,'' ''3'') and male (''4,'' ''5,'' ''6'') indicated no differences in the rating responses by listener gender. Thus, listener gender was excluded as a variable in further analyses.

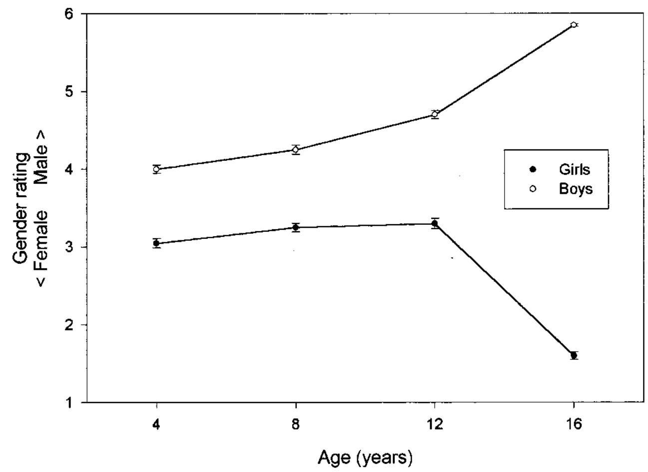


FIG. 4. Mean gender ratings ($1=$girl; $6=$boy) for 4-, 8-, 12-, and 16-year-old speakers. Error bars show $\pm 1$ standard error.

Based on the six-point gender rating scale, listeners' mean ratings are shown in Fig. 4. In a Speaker Age (4) $\times$Speaker Gender (2)$\times$Vowel (7) repeated measures analysis of variance, there were main effects of speaker age [$F(3,57)=8.70$; $p<0.001$], speaker gender [$F(1,19)=683.69$; $p<0.0001$], and vowel [$F(6,114)=57.88$; $p<0.0001$], and a significant speaker age$\times$speaker gender [$F(3,57)=328.61$; $p<0.0001$] interaction.[3] As illustrated, gender ratings were significantly different between boys' and girls' speech and voice within each age group (Tukey HSD, $p<0.05$). However, the gender ratings were clearly more similar at the three younger speaker ages than at 16 years.

To obtain an overall measure of whether gender was rated appropriately, the six-point rating scale was collapsed into a dichotomous scale, where 1, 2, or 3 was ''female'' and 4, 5, or 6 was ''male.'' The rating of each utterance by each listener was then scored as correct or not, based on whether the collapsed rating corresponded to the gender of the speaker. The percentages of utterances rated for which a correct gender rating was made were as follows: 4-year-olds (boys 67%, girls 62%); 8-year olds (boys 74%, girls 56%); 12-year-olds (boys 82%, girls 56%); and 16-year olds (boys 99.7%, girls 95%). Each of these percentages was based on 7,000 ratings, so that even the modest values of 56% for the 8- and 12-year-old girls exceeded the 50% chance level according to binomial tests. Nevertheless, listeners were better able to assign a correct gender to utterances spoken by boys than by girls in the 4-, 8-, and 12-year age groups.

Listeners in this study responded to CVC syllables based on a six-point gender rating scale. Overall, results indicated that listeners were able to rate vocalizations as either male of female sounding, even for the youngest children. Similarly, Weinberg and Bennett (1971) reported an identification rate of 78% for boys and 71% for girls in 5- and 6-year olds. Ingrisano *et al.* (1980) reported a gender identification rate of 71% for 4- and 5-year olds, and Bennett and Weinberg (1979) reported an identification rate of approximately 65%

TABLE II. Simple correlations between $f_0$, $F1$, $F2$, $F3$ and listener gender ratings for 4-, 8-, 12-, and 16-year-old boys and girls.

| Age group | $f_0$ | $F1$ | $F2$ | $F3$ |
|---|---|---|---|---|
| 4 years | $-0.60^a$ | $-0.47$ | $-0.33$ | $-0.82^a$ |
| 8 years | $-0.52^a$ | $-0.81^a$ | $-0.84^a$ | $-0.86^a$ |
| 12 years | $-0.56^a$ | $-0.89^a$ | $-0.78^a$ | $-0.83^a$ |
| 16 years | $-0.98^a$ | $-0.96^a$ | $-0.95^a$ | $-0.90^a$ |

[a]$p<0.05$.

for 6- and 7-year olds. Overall, listeners in this study correctly identified the speaker gender of 74% of the utterances averaged across the four age groups. This is comparable to the 81% identification rate reported by Sachs *et al.* (1973) for children between 4 and 14 years of age. Differences in gender identification between the current study and previous research may be due to the kind of speech material that was used. The current research used the CVC syllable, whereas all previous studies employed sentences. The sentence material may have contained additional prosodic cues to gender identification.

### *2. Relationship between voice/speech characteristics and gender ratings*

To investigate the relationship between acoustic measures and gender ratings, simple correlations were performed. Table II shows the simple correlation coefficients between the average listener gender rating across all utterances for each child speaker, and $f_0$, $F1$, $F2$, and $F3$. Note that a negative correlation means that a higher frequency was associated with ''girl-ness'' (since the gender rating scale had lower values for girls and higher values for boys). This table suggests that listeners gave considerable emphasis to $F3$ for the 4-year olds, to all three formants for the 8- and 12-year olds, and to $f_0$ as well as all the formants for the 16-year olds.

It has been shown that $f_0$ provides the most salient cue to gender perception from the voices of adults (e.g., Coleman, 1971, 1973; Lass, Hughes, Bowyer, Waters, and Bourne, 1976; Schwartz, 1968). However, even in the absence of $f_0$, listeners are able to make correct gender identifications from adults' speech (e.g., Lass *et al.*, 1976; Schwartz, 1968). As for children's voices, this study as well as others (e.g., Bennett, 1983; Bennett and Weinberg, 1979; Busby and Plant, 1995; Kent, 1976; Weinberg and Bennett, 1971) indicate that there is little or no difference in $f_0$ between boys and girls under the age of 12. Since $f_0$ probably does not play a large role in the determination of gender from children's voices in the four to 12 year age range, the contribution of vocal tract resonance characteristics to gender identification appears important.

## IV. GENERAL DISCUSSION

The main objective of this study was to examine children's acoustic characteristics that account for listeners' ability to identify gender from the voice and speech. To accomplish this, vocal recordings of 4-, 8-, 12-, and 16-year-old boys and girls were made, and acoustic analyses of these recordings were performed. These speech tokens were then used in a gender perceptual rating task. Results from these acoustic and perceptual experiments have provided information about the developmental bases for listeners' ability to identify gender from speech and voice.

### A. Acoustic correlates of gender

Adult listeners were able to determine gender from the speech and voice of children as young as four years of age. Since $f_0$ is very similar for young girls and boys between 4- and 12-years-old, it is unlikely that this property was utilized by the listeners. Instead, listeners must have attended to other acoustic correlates of gender such as formant frequencies.

The average formant frequency ($F1$, $F2$, $F3$) functions revealed consistently lower values for boys than girls. Vocal tract length is one determinant of formant frequencies. Several anatomical studies have provided support for gender differences in vocal tract length. Fant (1973) found differences in both oral and pharyngeal dimensions for male female vocal tracts, but the distinction was largest for the pharynx. In a cephalometric study, King (1952) found that the pharyngeal portion extending from the hard palate to the hyoid bone was longer for boys than girls ranging from one to 16 years. The difference in vocal tract length values for boys and girls ranged from 2% to 13%, with the smaller value for the 1-year olds and the larger value for the 16-year olds. Based on these anatomical data, it would be predicted that differences in formant frequencies of boys and girls would increase as a function of age. The results of the current study clearly support this prediction. Across $F1$, $F2$, and $F3$, gender differences were greatest for the 16-year-old group and smallest for the 4-year-old group.

The computation of K factors in this study revealed that the percentage of differences between boys and girls in $F1$, $F2$, and $F3$ values within each age group remained very similar, and these formant frequencies were lower for the boys than for the girls. Busby and Plant (1955) also showed that vowel formant frequencies were consistently lower for boys than girls, and that $F1$, $F2$, and $F3$ decreased in value with age. The formant frequency K factors were computed as an average across vowel contexts. An examination of specific vowels showed that for 8-, 12-, and 16-year olds, average formant frequencies were always lower for boys than girls. Previous studies employing K-factor analyses of adult formant frequencies have shown that sexual distinctions are vowel and formant dependent (Fant, 1973). For example, Fant determined that male/female differences were largest for $F2$ and $F3$ of the front vowels, and $F1$ of [æ]. These differences were smaller for $F1$ and $F2$ of back vowels. Thus, certain vowels may have a larger impact of gender differentiation than other vowels.

One obvious implication of the current findings showing lower vocal tract resonances for young boys than girls is that sexual differences exist in vocal tract length. Although direct estimates of vocal tract size are difficult to obtain for children, at least one previous study estimated more gross measures of vocal tract dimensions (Bennett, 1981). Bennett measured standing height, sitting height, weight, and neck circumference in 7- to 8-year-old boys and girls. Simple correlations between each of the measures of body size and the various vowel formants revealed that 93% of these correlations were significant. All of the correlations were negative, indicating that an increase in a physical dimension was associated with a decrease in formant frequency. These same physical dimensions were measured in the current study for a population of children similar in age. The majority of the correlations were negative, and the correlations for body weight were in the same range as those reported by Bennett. Moreover, these findings are consistent with the K-factor analyses revealing that the average formant frequency difference between 7- to 8-year-old boys and girls was 10% and 9% for the Bennett and current studies, respectively. Thus, the acoustic analyses and measures of gross vocal tract dimensions were comparable across studies.

In a recent study (Tecumseh Fitch and Giedd, 1999), magnetic resonance imaging (MRI) was used to quantify the vocal tract morphology of subjects between 2 and 25 years of age. The MRI results revealed positive correlations between vocal tract length and either body height or weight. As illustrated in Fig. 3 of the current study, the relationship between body weight and formant frequency change was similar for boys and girls. Thus, an increase in weight resulted in lower formant frequencies for both boys and girls. Although it would be reasonable to predict that vocal tract length contributed to lower formant frequencies of boys compared to girls at a given age level in the current research, the MRI data only partially support this prediction. Tecumseh Fitch and Giedd found that vocal tract length was significantly longer in boys than girls after 10 years of age. These findings indicate that vocal tract length contributes to differences in formant frequencies between boys and girls after age 10. However, the results also suggest that formant frequency differences between boys and girls before age 10 come from anatomical and/or articulatory sources other than vocal tract length.

### B. Perceptual correlates of gender

It is plausible that listeners in this study based their gender rating of the younger children on information from vowel formant frequencies. The third formant values accounted for the highest percentage of the total variance in listener rating behavior for the 4- and 8-year olds, while $F1$ accounted for the highest percentage of the total variance in listener rating behavior for the 12-year olds. Across all age groups, boys tended to have lower vowel formant frequencies than girls. Bennett and Weinberg (1979) and Ingrisano *et al.* (1980) found similar results for 4- to 7-year olds. We conclude that vowel formant frequencies most likely played a primary role in listener rating behavior of the younger children.

It is generally assumed that the perception of gender from speech and voice is dependent upon several acoustic correlates of sound such as $f_0$ and formant frequencies (Mullennix *et al.*, 1995). Even though gender is characterized by various acoustic properties, our perceptual system is able to adapt to and compensate for different types of acoustic information (e.g., in the absence of $f_0$), which provide cues as to speaker gender. In a series of experiments, Mullennix *et al.* (1995) investigated the perceptual representation of gender from the voice. In one experiment, listeners were presented synthetic vowel sounds designed to range from male to female voices and were asked to rate each stimulus on a six-point scale similar to the one used in the current study. There was a gradual identification function between male and female voices, as well as good discrimination between voices within each gender category. They concluded that the perceptual representation of gender from the voice is auditory based, not categorical. Though the current study did not directly examine the perceptual mode (i.e., categorical versus auditory perception) for identifying gender from speech and voice, the results indicate that listeners were attending to various $f_0$ and formant acoustic cues in order to make their gender ratings.

Since the focus of the current research was on the role of formant frequencies and $f_0$ in gender identification, a CVC production unit was used to minimize other prosodic cues to gender. However, gender-specific prosodic differences have been shown to provide listeners with cues as to speaker gender (e.g., Spencer, 1988; Wolfe *et al.*, 1990). For example, typical female speech patterns generally show greater tonal variability (i.e., more upward intonations) than male speech patterns (Crystal, 1975). Wolfe *et al.* (1990) found when listeners were presented voices with similar $f_0$ ranges, they tended to identify voices that were less monotonous (i.e., more $f_0$ variability) as female sounding. There is a strong possibility that the results from several studies that examined the gender perception of children's speech were influenced by this type of prosodic effect, because listener responses were based on sentence stimuli (e.g., Bennett and Weinberg, 1979; Ingrisano *et al.*, 1980). Nevertheless, it is unlikely that prosodic effects influenced listener ratings in the current study. The CVC syllables in this study were of very short duration (mean duration was 200 ms), and stressed with typical rhythmic and intonational patterns.

## V. CONCLUSIONS

In the present study we showed clear gender differences in both children's vowel productions, and adults' perception of children's speech and voice. The acoustic as well as the perceptual data from this study clarify the role of $f_0$ and formant frequency properties in the production and perceptual differentiation of gender across a large age range of child speakers. Moreover, the current research also provides directions for future research. Although a number of production studies, including the present, have examined the role of $f_0$ and formant frequency in gender identification, there is a need for research on the role of voice quality and particularly breathiness in sex differentiation. In the perception of gender, future research should examine the contribution of the auditory mode of perception through appropriate discrimination conditions of sound continua. As in the current research, the inclusion of a rating scale in conjunction with the discrimination paradigm may provide additional information on the mode of gender identification (Mullennix *et al.*, 1995). Further perceptual studies should examine variations in prosody as cues to gender perception.

The conclusions supported by the results of this research are as follows.

(1) Vowel formant frequencies differentiate gender for very young children, whereas formant frequencies and $f_0$ differentiate gender after 12 years of age.
(2) Adults can perceive the gender of children as young as four years of age, and the magnitude of the difference between boy and girl perceptual ratings is large by age 16. Adults' perception of gender from children's speech and voice is strongly related to formant frequencies for children aged 4 to 12 years, whereas $f_0$ plays a key role by 16 years.

## ACKNOWLEDGMENTS

This article is based on a Ph.D. dissertation completed at Vanderbilt University by the first author under the direction of the second and third authors. The research was supported by NIH Grant Nos. DC00464 and DC00523, and a Vanderbilt University Research Council grant. We express our sincere gratitude to the children who participated in this study, to their parents, and to the adult listeners. The authors express their appreciation to William Caul, Jeffery Franks, Joseph Lappin, and Maureen Powers for their discussions, interest, and advice about this work. The authors thank Anders Lofqvist, Eugene Buder, and two anonymous reviewers for comments on this manuscript.

[1]Other significant interactions for $F1$ were as follows: vowel×age [$F(18,432)=7.66$; $p<0.0001$], vowel×gender [$F(6,432)=18.81$; $p<0.0001$], and vowel×age×gender [$F(18,432)=2.29$; $p<0.01$].

[2]Other significant interactions for $F2$ and $F3$ were as follows: $F2$—vowel×age [$F(18,432)=13.21$; $p<0.0001$], vowel×gender [$F(6,432)=4.84$; $p<0.001$], and vowel×age×gender [$F(18,432)=3.65$; $p<0.01$]; $F3$—vowel×age [$F(18,432)=3.10$; $p<0.0001$].

[3]Other significant interactions were as follows: vowel×speaker age [*Frs(18,342)=11.60*, $p<0.0001$], vowel×speaker gender [$F(6,114)=28.58$, $p<0.0001$], and vowel×speakerage×speakergender [$F(18,432)=5.80$, $p<0.0001$].

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Edinburgh University Press, Edinburgh).

Bennett, S. (**1981**). ''Vowel formant frequency characteristics of preadolescent males and females,'' J. Acoust. Soc. Am. **69**, 321–238.

Bennett, S. (**1983**). ''A 3-year longitudinal study of school-aged children's fundamental frequencies,'' J. Speech Hear. Res. **26**, 137–142.

Bennett, S., and Weinberg, B. (**1979**). ''Sexual characteristics of preadolescent children's voices,'' J. Acoust. Soc. Am. **66**, 989–1000.

Busby, P., and Plant, G. (**1995**). ''Formant frequency values of vowels produced by preadolescent boys and girls,'' J. Acoust. Soc. Am. **97**, 2603–2606.

Cohen, J. (**1969**). *Statistical Power Analysis for the Behavioral Sciences* (Academic, New York).

Coleman, R. (**1971**). ''Male and female voice quality and its relationship to vowel formant frequencies,'' J. Speech Hear. Res. **14**, 565–577.

Coleman, R. (**1973**). ''Speaker identification in the absence of inter-subject differences in glottal source characteristics,'' J. Acoust. Soc. Am. **53**, 1741–1743.

Cruttenden, A. (**1986**). *Intonation* (Cambridge, University Press, Cambridge).

Crystal, D. (**1975**). *The English Tone of Voice* (St. Martin's Press, New York).

Crystal, D. (**1982**). *Clinical Linguistics* (Springer-Verlag, Vienna).

Eguchi, S., and Hirsh, I. (**1969**). ''Development of speech sounds in children,'' Acta Oto-Laryngol. Suppl. **257**, 2–51.

Fant, G. (**1973**). ''A note on vocal tract size factors and nonuniform F-pattern scalings,'' in *Speech Sounds and Features* (MIT Press, Cambridge, MA), pp. 84–93.

Fleiss, J. (**1981**). *Statistical Methods for Rates and Proportions* (Wiley, New York).

Haley, K. L., and Ohde, R. N. (**1996**). ''Stimulus uncertainty and speaker normalization processes in the perception of nasal consonants,'' Phonetica **53**, 185–199.

Hasek, C., Singh, S., and Murry, T. (**1980**). ''Acoustic attributes of preadolescent voices,'' J. Acoust. Soc. Am. **68**, 1262–1265.

Hollien, H., Green, R., and Massey, K. (**1994**). ''Longitudinal research on adolescent voice change in males,'' J. Acoust. Soc. Am. **96**, 2646–2654.

Ingrisano, D., Weismer, G., and Schuckers, G. (**1980**). ''Sex identification of preschool children's voices,'' Folia Phoniatr. **32**, 61–69.

Johnson, K. (**1990a**). ''The role of perceived speaker identity in $f_0$ normalization of vowels,'' J. Acoust. Soc. Am. **88**, 642–654.

Johnson, K. (**1990b**). ''Contrast and normalization in vowel perception,'' J. Phonetics **18**, 229–254.

Kent, R. (**1976**). ''Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies,'' J. Speech Hear. Res. **19**, 421–447.

Kent, R. D. (**1978**). ''Imitation of synthesized vowels by preschool children,'' J. Acoust. Soc. Am. **63**, 1193–1198.

Kent, R., and Vorperian, H. (**1995**). *Development of the Craniofacial–oral–laryngeal Anatomy* (Singular Publishing, San Diego).

King, E. (**1952**). ''A roentgenographic study of pharyngeal growth,'' Angle Orthod. **22**, 23–37.

Kirk, R. (**1982**). *Experimental Design: Procedures for the Behavioral Sciences* (Wadsworth, Belmont, CA).

Klatt, D., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Ladefoged, P., and Broadbent, D. (**1957**). ''Information conveyed by vowels,'' J. Acoust. Soc. Am. **29**, 98–104.

Lass, N., Almerino, C., Jordan, L., and Walsh, J. (**1980**). ''The effect of filtered speech on speaker race and sex identifications,'' J. Phonetics **8**, 101–112.

Lass, N., Hughes, K., Bowyer, M., Waters, L., and Bourne, V. (**1976**). ''Speaker sex identification from voiced, whispered, and filtered isolated vowels,'' J. Acoust. Soc. Am. **59**, 675–678.

Laver, J., and Trudgill, P. (**1979**). ''Phonetic and linguistic markers in speech,'' in *Social Markers in Speech*, edited by K. Scherer and H. Giles (Cambridge University Press, Cambridge), pp. 1–32.

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.

Liberman, A., Cooper, F., Shankweiler, D., and Studdert-Kennedy, M. (**1967**). ''Perception of speech code,'' Psychol. Rev. **74**, 431–461.

Lindblom, B., and Sundberg, J. (**1971**). ''Acoustical consequences of lip, tongue, jaw, and larynx movement,'' J. Acoust. Soc. Am. **50**, 1166–1179.

Mullennix, J., Johnson, K., Topcu-Durgun, M., and Farnsworth, L. (**1995**). ''The perceptual representation of voice gender,'' J. Acoust. Soc. Am. **98**, 3080–3095.

Nolan, F. (**1983**). *The Phonetic Bases of Speaker Recognition* (Cambridge University Press, Cambridge).

Nordstrom, P. (**1997**). ''Female and infant vocal tracts simulated from male area function,'' J. Phonetics **5**, 81–92.

Oates, J., and Dacakis, G. (**1983**). ''Speech pathology considerations in the management of transsexualism—A review,'' Br. J. Comm. Dis. **18**, 139–151.

Ohde, R., and Sharf, D. (**1992**). *Phonetic Analysis of Normal and Abnormal Speech* (McMillan, Columbus, OH).

Peterson, G., and Barney, H. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Sachs, J., Lieberman, P., and Erickson, D. (**1973**). ''Anatomical and cultural determinants of male and female speech,'' in *Language Attitudes*, edited by R. W. Shuy and R. W. Fasold (Georgetown University Press, Washington, DC), pp. 74–83.

Schwartz, M. (**1968**). ''Identification of speaker sex from isolated, voiceless fricatives,'' J. Acoust. Soc. Am. **43**, 1178.

Spencer, L. (**1988**). ''Speech characteristics of male-to-female transsexuals: A perceptual and acoustic study,'' Folia Phoniatr. **40**, 31–42.

Sussman, H., Minifie, F., Buder, E., Stoel-Gammon, C., and Smith, J. (**1996**). ''Consonant-vowel interdependencies in babbling and early

words: Preliminary examination of a locus equation approach,'' J. Speech Hear. Res. **39**, 424–433.
Tecumseh Fitch, W., 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.
Weinberg, B., and Bennett, S. (**1971**). ''Speaker sex recognition of 5- and 6-year-old children's voices,'' J. Acoust. Soc. Am. **50**, 1210–1213.
Wolfe, V., Ratusnik, D., Smith, F., and Northrop, G. (**1990**). ''Intonation and fundamental frequency in male-to-female transsexuals,'' J. Speech-Hear. Res. **55**, 43–50.