Identification of Children's Gender and Age by Listeners

Ofer Amir, Merav Engel, Esther Shabtai, and Noam Amir, *Tel-Aviv, Israel*

Summary: Objectives. Voice carries abundant information about the speaker. This acoustic information changes throughout life. Although the ability of identifying audible cues on a speaker's gender and age is considered an intuitive task, little is known about the ability to identify and decipher this perceptual information. Most studies in the field have examined the ability to identify adults' gender and age, thus the purpose of the present study was to evaluate listeners' ability to identify gender and age of children and adolescents.

Methods. A total of 120 children in six age groups, 8, 10, 12, 14, 16, and 18 years, were recorded while producing isolated vowels and fixed sentences. The recordings were evaluated by a group of 38 untrained naive listeners, who were asked to identify the speakers' gender and age.

Results. Listeners were able to identify children's gender at an overall rate of 81.81%. This task was performed more successfully based on recordings of sentences (85.39%) than on isolated vowels (78.22%). Listeners were able to identify the children's age group at an overall rate of 37.16%. This task was also performed more successfully based on recordings of sentences (39.58%) than on isolated vowels (34.71%). Furthermore, when an error of ±1 age group was allowed, correct responses for age identification exceeded 80%.

Conclusions. Listeners have the ability to identify children's gender and age, based on short audio recordings, even before puberty. The success rates in these perceptual tasks are dependent on the child's age and gender.

Key Words: Children–Perception–Voice–Age–Gender–Hebrew.

INTRODUCTION

The human voice carries dynamic acoustic information about the speaker. Listeners are thus able to identify various personal attributes of the speaker, based solely on his/her voice. Such attributes include, for example, emotional state,[1] personality traits,[2] sexual attractiveness,[3] health condition,[4] physical size,[5,6] gender,[7] and age.[8] Whereas most studies on listeners' perception of speakers' characteristics have focused on perception of adults' voice, the present study was aimed at evaluating the ability of listeners to identify children's gender and age, based on their voice.

Physical size is considered difficult to identify based on voice in humans and in animals.[9–11] Gonzalez[12] reviewed a series of studies on the perception of physical characteristics. His data show that, on the one hand, listeners are highly consistent in their judgments of speakers' height and weight, under different acoustic conditions. On the other hand, listeners' estimations of height and weight (although consistent) are mostly inaccurate or mistaken (only 14% of judgments were significantly correlated with the physical measurements). He concluded, "listeners follow vocal stereotypes about the body size of speakers, even though these stereotypes are wrong" (p297). When the perceptual task is set as a binary task, and listeners are required to differentiate between "big" and "small" speakers, a much higher performance rate can be obtained.[13]

Auditory perception of speaker's gender is considered an intuitive task.[9] Studies show that listeners are able to successfully identify the gender of adult speakers, based on their voice. Pitch is considered a primary cue on which listeners rely for gender identification.[7,14] Nonetheless, it was shown that adults' gender could also be identified in recordings that did not provide the F0 cue. For example, listeners were shown to identify speakers' gender based on recordings of voiceless fricatives[5,15] or when F0 was filtered out from the signal.[16] Although many studies have examined perception of gender based on adult and elder voice, only a few studies have focused on the perception of gender in childhood. This might be attributed to the conception that vocal differences between genders appear only after puberty because there are no morphological differences during childhood between the male and female larynges, as well as voice differences.[17] However, the few studies that have examined listeners' ability to identify children's gender from their voice reported on above-chance correct identification rates. Bennett and Weinberg,[18] for example, reported that listeners were able to identify the gender of preadolescent children at 60–80% correct rate. In contrast, another study found that listeners had difficulties identifying gender of young children and that when children were between the ages of 5 and 11, correct gender identification rates varied from 19% to 40%.[19] In addition, it was reported that listeners tended to identify younger voices (ages 5 and 7) as girls and older voices (age 11) as boys.

As opposed to the difficulties that listeners show in perceiving speaker's physical size, perception of speaker's age was shown to be more reliable. In different studies, listeners were able to estimate adult speakers' age above chance levels, using different methodologies, even when judgments were based on limited auditory signal. Ptacek and Sander,[20] for example, reported that listeners were able to successfully assign voices into two distinct age categories: older than 65 and younger than 35 years. Correct assignment rates were 78% for prolonged /a/ vowels, 87% for backward speech, and 99% for

Accepted for publication June 13, 2011.
From the Department of Communication Disorders, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel.
Address correspondence and reprint requests to Ofer Amir, Department of Communication Disorders, Sheba Medical Center, Tel-Aviv University, Tel-Hashomer 52621, Israel. E-mail: oferamir@post.tau.ac.il
Journal of Voice, Vol. 26, No. 3, pp. 313-321
0892-1997/$36.00

doi:10.1016/j.jvoice.2011.06.001

forward speech. In a different study,[21] a large group of adult speakers between the ages of 20 and 89 years were recorded, and listeners were able to accurately identify speakers' age, with a significant high correlation ($r = 0.88$) between chronological and perceived age. These results were later supported by another study in which listeners were required to make age estimations of 80 adult males between the ages of 40 and 80 years.[22] In that study, a correlation of 0.77 was obtained between the chronological and perceived age. Neiman and Applegate[23] asked a group of listeners to assign recordings of 36 adult speakers into seven age categories and examined the percentage of correct correspondence between the chronological age and perceived age groups. They reported on an overall correct assignment rate of 80.25% and that values ranged between 61.26% and 90.32% for male voices and between 73.72% and 90.57% for female voices. These findings demonstrate that listeners can estimate adults' age based on their voice and that direct estimation of age is more difficult than assigning voices into age categories.

Data on listeners' perception of children's age are scarce. We were able to identify only two studies in which this topic was examined. In one study,[19] it was reported that experienced listeners' agreement on the perception of children's age was better than that of naive listeners. In the second study,[24] listeners' agreement, as reflected by the standard deviation (SD) of the perceived age, was larger for voice samples recorded from adults than those from children. Nonetheless, correct age estimations were not reported in both studies. The purpose of the present study was, therefore, to evaluate listeners' ability to identify gender and estimate age of children and adolescents, based on their voice.

METHODS

Speakers

After obtaining approval of our institutional ethics committee, and parental written consent, 120 children were recruited for this study. Children were divided into six age groups: 8, 10, 12, 14, 16, and 18 years. Each group was composed of 10 boys and 10 girls, with an age range of ±6 months. All children were healthy native Hebrew speakers, with no reported history of speech, voice, or hearing disorders. Children who had singing or voice training, as well as children who were sick or dysphonic on the day of the recordings, were not included. Table 1 summarizes the children's age distribution and physical characteristics.

Recordings

Each child was recorded in a quiet room sustaining the isolated vowels /a/ and /i/ six times for 3–5 seconds and uttering two repetitions of a voiced sentence. The voiced sentences were /dani ba im aba laavoda/ and /gad azaɣ lanu bayam/, which translate into English as "Danny came to work with dad" and "Gad helped us on the beach," respectively ("Danny" and "Gad" are both common Hebrew names). Recordings were conducted using a Sennheiser PC20 headset microphone (Sennheiser Communications, Hanover, Germany), located 5 cm from the corner of the child's mouth and connected directly to a computer sound card. Before the recording, each child was instructed briefly about the recordings, and the different tasks were presented visually using a card. The children were instructed to read the cards silently first and then recorded while saying the sentences or vowels as naturally as possible, without reading it. No vocal demonstration was given to the children to reduce a possible bias effect. Sampling rate for recording was 48 kHz (16 bit). To reduce a possible order effect, the different tasks were performed in a random order that was changed between speakers.

From the recordings of each child, the third production of the vowel /a/ and the second recording of the voiced sentence were used for the perceptual task, yielding a total of 120 vowels and 120 sentences.

Listeners

Thirty-eight native Hebrew-speaking women with a mean age of 30.46 years (SD = 6.36) volunteered for the listening task. All listeners had no professional background in speech or voice training. Schoolteachers, singing teachers, and speech-language pathologists were not included in the listeners group because they were regarded as experienced listeners. All

TABLE 1.
Mean Chronological Age, Height, Weight, and Neck Circumference for Each Age Group and Gender (Standard Deviation in Parentheses)

Age Group (y)	Gender	Age (y)	Height (cm)	Weight (kg)	Neck (cm)
8	Boys	7.94 (0.34)	127.85 (5.15)	28.56 (6.81)	27.91 (2.13)
	Girls	8.09 (0.28)	125.08 (5.30)	27.18 (5.41)	26.62 (1.39)
10	Boys	9.99 (0.35)	139.05 (4.54)	35.03 (6.74)	30.05 (3.06)
	Girls	9.93 (0.24)	136.23 (8.93)	34.33 (10.66)	27.70 (1.95)
12	Boys	11.95 (0.25)	150.90 (4.90)	44.34 (6.70)	30.87 (2.40)
	Girls	11.97 (0.35)	152.45 (6.47)	44.96 (10.98)	29.59 (1.62)
14	Boys	14.02 (0.33)	166.45 (8.29)	50.70 (8.10)	33.16 (1.54)
	Girls	14.07 (0.19)	160.10 (5.96)	48.77 (7.27)	30.23 (1.63)
16	Boys	15.87 (0.36)	173.55 (7.19)	60.08 (8.17)	34.51 (2.06)
	Girls	15.98 (0.28)	161.05 (4.93)	56.37 (7.03)	30.93 (1.23)
18	Boys	17.80 (0.14)	177.80 (5.17)	73.87 (15.09)	36.79 (3.27)
	Girls	17.79 (0.28)	159.70 (5.35)	62.86 (12.47)	32.87 (2.57)

listeners were healthy with no reported hearing, speech, or voice problems.

Procedure

Recordings were divided into vowels and sentences and presented to the listeners using an ad hoc *MATLAB* program (MathWorks, Natick, MA) through Sony MDR-XD300 headphones (Sony, Tokyo, Japan). After a brief explanation, a procedural training task was performed, which included four recordings. During this training task, listeners could adjust the playback intensity level. During the main listening task, each recording was presented once and then the listener was presented with two consecutive dialog boxes. In one box, she assigned a gender to the recording, and on the second, she assigned the recording into one of the six possible age groups.

In addition to the target stimuli, each listener judged 12 vowels and 12 sentences that were presented twice for intrajudge reliability evaluation. Consequently, each listener provided 528 responses ([120 vowels + 12 reliability items + 120 sentences + 12 reliability items] × 2 questions), thus a total of 20,064 responses were recorded throughout the study. Intrajudge agreement was higher for sentences than for vowels. For the gender question, 87.93% of the sentences were judged consistently by the same listener ($\kappa = 0.75$, $P < 0.01$) and 83.12% of the vowels were judged consistently ($\kappa = 0.66$, $P < 0.01$). For the age question, 55.3% of the sentences were judged consistently (weighted $\kappa = 0.68$, $P < 0.01$) and 39.04% of the vowels were judged consistently (weighted $\kappa = 0.50$, $P < 0.01$). Moreover, Spearman correlation coefficients for the comparison between the first and second age estimations were 0.85 ($P < 0.01$) for sentences and 0.69 ($P < 0.01$) for vowels. Interjudge correlation was evaluated for the different age groups, genders, and stimuli. For the gender question, kappa correlation coefficients ranged between 0.30 and 0.76 ($P < 0.001$). For the age question, Kendall correlation coefficients ranged between 0.26 and 0.81 ($P < 0.001$).

To reduce a possible order effect, listeners were divided into four subgroups for which the order of the questions (gender/age) and stimuli type (vowel/sentence) were alternated. No time limitation was imposed on the task, but each recording was played once. The complete listening task lasted approximately 50 minutes, including a 5-minute break.

RESULTS

The relation between biological and perceived gender and between chronological and perceived age were evaluated using two approaches: (1) correlation between actual and perceived values and (2) percentage of correct identification. Results are presented, first, for gender identification, followed by the results for age identification.

Correlation between biological and perceived gender

For all stimuli, an overall kappa correlation coefficient of 0.64 ($P < 0.001$) was obtained between the biological and perceived gender. A higher correlation coefficient value was obtained for sentences in comparison to vowels (0.71 and 0.56, $P < 0.001$, respectively). The error of identifying a female as a male occurred more frequently (1003/4560 cases; 21.9%) than the error of identifying a male as a female (656/4560 cases; 14.4%). The difference between the occurrences of the two possible errors was found statistically significant, using the McNemar test ($P < 0.001$). This result was consistent for the sentence and for vowel stimuli, analyzed separately, as well as for all stimuli collapsed together.

Percentage of correct identification: gender

Overall, listeners correctly identified children's gender based on their voice in 81.81% of presentations (7461/9120 cases). Table 2 presents a summary of correct response rate of gender identification for the sentences and vowels obtained from children and adolescents in this study.

Data show that in the younger age groups (ages 8–12), girls were identified correctly more often than boys, but in the older age groups (ages 14–18), boys were identified correctly more often. Consequently, overall identification rate was higher for boys than for girls (85.61% vs 78.00%, respectively). In general, gender identification was better for sentence recordings than that for vowels (85.39% vs 78.22%, respectively). This advantage of sentences over vowels was observed more consistently in the younger age groups than in the older groups.

An analysis of variance with a mixed model revealed a main effect for gender [$F(1,37) = 99.16$, $P < 0.001$], age group [$F(5,185) = 71.30$, $P < 0.001$], and stimuli [$F(1,37) = 88.06$, $P < 0.001$]. A significant gender × age group interaction was

TABLE 2.
Summary of Correct Response Rate for Gender Identification for Boys and Girls and for Sentences and Vowels

		By Gender			By Stimuli		
Age Group (y)	Overall	Boys	Girls	Difference	Sentence	Vowel	Difference
8	68.82	66.32	71.32	−5.00**	76.71	60.92	15.79*
10	78.68	74.08	83.29	−9.21**	86.05	71.32	14.73*
12	79.07	74.74	83.29	−8.55**	83.82	74.21	9.61*
14	90.20	98.82	81.58	17.24**	89.34	91.05	−1.71^{NS}
16	87.30	99.87	74.74	25.13**	89.74	84.87	4.87*
18	86.84	99.87	73.82	26.05**	86.71	86.97	−0.26^{NS}

Abbreviation: NS, not significant.
*$P < 0.05$, **$P < 0.01$.

found [$F(5,185) = 82.37$, $P < 0.001$]. To inspect this interaction, a test of effect slices was conducted. It revealed significant gender differences for all age groups (*corrected* $P < 0.01$). This analysis confirmed that in the three younger age groups, girls were identified correctly at higher rates than boys, whereas in the three older age groups, boys were identified correctly more than girls.

A significant stimuli × age group interaction was found [$F(5,185) = 15.81$, $P < 0.001$]. A test of effect slices was conducted and revealed a significant advantage for sentences over vowels in the three younger age groups and in the 16-year-old group (*corrected* $P < 0.01$).

A significant gender × stimuli interaction was found [$F(1,37) = 28.19$, $P < 0.001$]. This interaction was further inspected by a test of effect slices. It revealed that although sentences were identified correctly more often than vowels in most cases, the numerical difference between the stimuli was larger for boys than for girls (11.23% vs 3.11%, respectively). Concurrently, the numerical difference between genders (ie, boys identified correctly more often than girls) was larger for sentences than for vowels (11.67% vs 3.55%, respectively).

Figure 1 illustrates correct gender identification rates obtained for sentences (Figure 1A) and vowels (Figure 1B) for boys and girls in the six age groups.

Correlation between chronological and perceived age

For all stimuli, an overall Pearson correlation coefficient value of $r = 0.85$ ($P < 0.001$) was obtained between chronological and perceived age. Similar values were found for sentences ($r = 0.87$, $P < 0.001$) and vowels ($r = 0.83$, $P < 0.001$). Figure 2 presents a scatter plot of chronological versus perceived age, for boys and girls, obtained for sentences (Figure 2A) and vowels (Figure 2B), as well as correlation coefficients.

In addition to the strong correlation between chronological and perceived age, this figure demonstrates that listeners' responses varied mostly around ±1 age group (2 years). The only exception for that was the 18-year-old female group, which was more scattered. Furthermore, older girls (aged 14 years or older) were typically identified as younger than their chronological age.

Percentage of correct identification: age

Overall, listeners assigned children to their age group accurately in 3389 of 9120 cases (37.16%). Table 3 presents a summary of correct age identification rates for the recordings of sentences and vowels obtained from the children in this study.

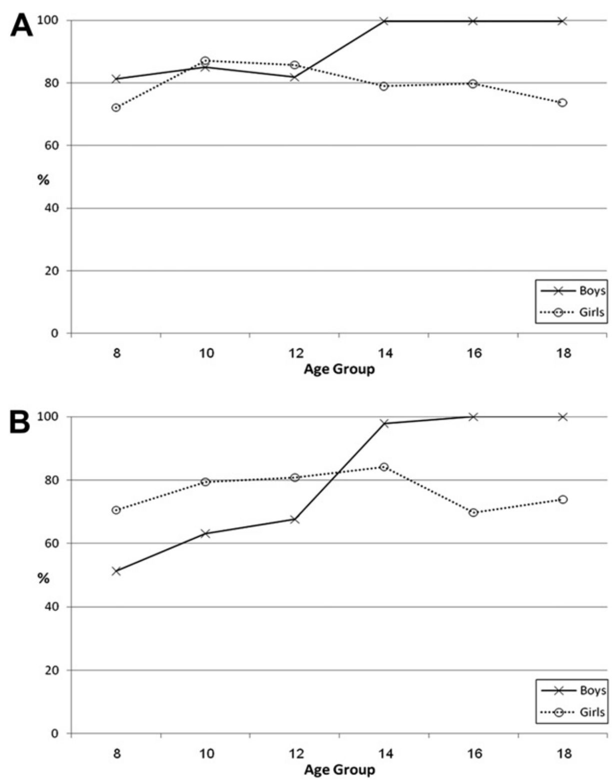

FIGURE 1. Correct gender identification rates for boys and girls in the six age groups for (**A**) sentences and (**B**) vowels.

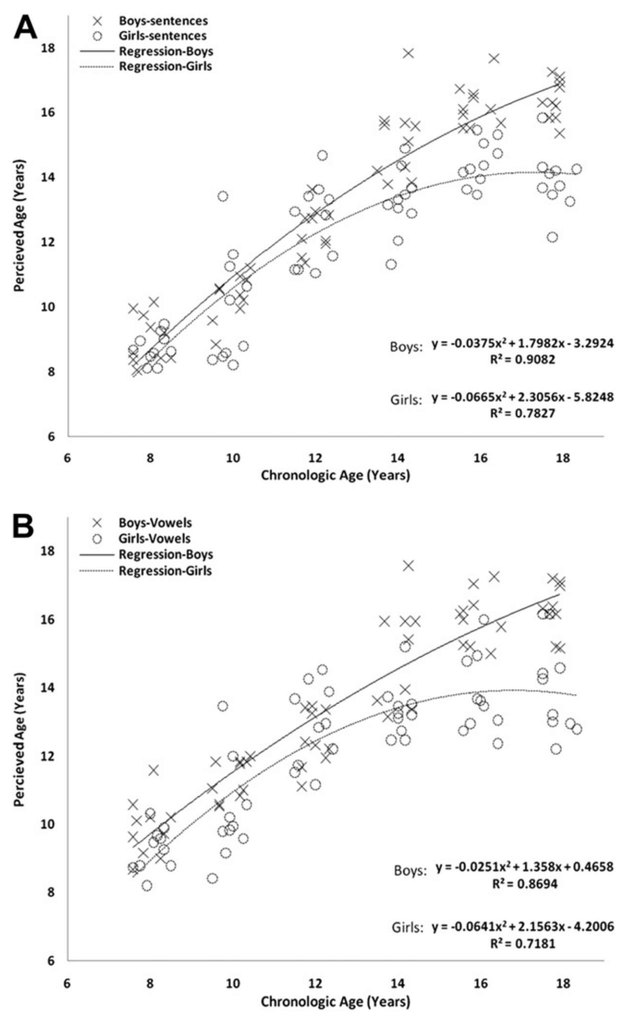

FIGURE 2. Correlation between the chronological and biological age in the boys and girls groups, for the (**A**) sentences stimuli and (**B**) vowel stimuli.

TABLE 3.
Summary of Correct Assignment to the Six Age Groups by Gender (Boys and Girls) and Stimuli (Sentences and Vowels)

		Gender			Stimuli		
Age Group (y)	Overall	Boys	Girls	Difference	Sentence	Vowel	Difference
8	52.70	45.53	59.87	−14.34*	61.58	43.82	17.79*
10	38.62	44.61	32.63	11.98*	41.32	35.92	5.40*
12	35.53	36.97	34.08	2.89^{NS}	35.92	35.13	0.79^{NS}
14	36.12	33.03	39.21	−6.18*	36.84	35.39	1.45^{NS}
16	38.03	50.53	25.53	25.00*	39.74	36.32	3.42^{NS}
18	21.90	35.66	8.16	27.50*	22.11	21.71	0.40^{NS}

Abbreviation: NS, not significant.
*$P < 0.05$.

Data show that boys were assigned to their correct age group more frequently than girls (41.05% vs 33.23%, respectively). In general, assignment to the correct age group was performed better, based on recordings of sentences rather than vowels (39.58% vs 34.71%, respectively).

Statistical analysis, using analysis of variance with a mixed model revealed a main effect for gender [$F(1,37) = 64.18$, $P < 0.001$], age group [$F(5,185) = 67.49$, $P < 0.001$], and stimuli [$F(1,37) = 24.96$, $P < 0.001$]. A significant gender × age interaction was found [$F(5,185) = 49.54$, $P < 0.001$]. To explore this interaction, a test of effect slices was conducted and revealed significant gender differences in all age groups, except for the 12-year-old group (*corrected* $P < 0.01$). As demonstrated in Table 3, girls in the youngest age group (8 years) were assigned correctly to their age group at the highest rates, compared with all other groups. In contrast, girls in the two older groups (16 and 18 years) were typically assigned incorrectly to a younger category than their actual age group.

A significant stimuli × age interaction was found [$F(5,185) = 7.62$, $P < 0.001$]. To explore this interaction, a test of effect slices revealed that although a consistent advantage for sentences was observed, it was statistically significant only in the two younger age groups ($P < 0.05$), whereas it failed to reach statistical significance in all other age groups.

A significant gender × stimuli interaction was found [$F(1,37) = 5.69$, $P = 0.022$]. A test of effect slices revealed a significant gender difference in both stimuli ($P < 0.01$). Although the advantage for sentences over vowels was observed in both genders, this difference was statistically significant only for boys ($P < 0.001$) but failed to reach statistical significance for girls ($P = 0.073$).

Figure 3 illustrates the correct assignment rates of the children to their age groups, based on recordings of sentences (Figure 3A) and vowels (Figure 3B), for boys and girls in the six age groups.

To further explore listeners' ability to identify children's age based on their voice, the distribution of the listeners' responses to the age question was examined. Table 4 presents the distribution of the listeners' responses to the age question, for both genders within each age group, and mean correct response rates for each age group, when an error of ±1 age group was allowed.

With a ±1 age group error, mean correct age identification rate was 87.04% for boys and 74.45% for girls, with an overall mean value of 80.75%. Inspection of the data shows that when listeners judged boys' age inaccurately, they tended to identify them as older, while tending to judge older girls (age 14–18) as younger.

Physical characteristics effect

To evaluate the possible contribution of the children's physical characteristics to the listeners' ability to identify speakers' gender and age, a linear regression analysis was performed. In this analysis, the dependent variable was defined as percentage of

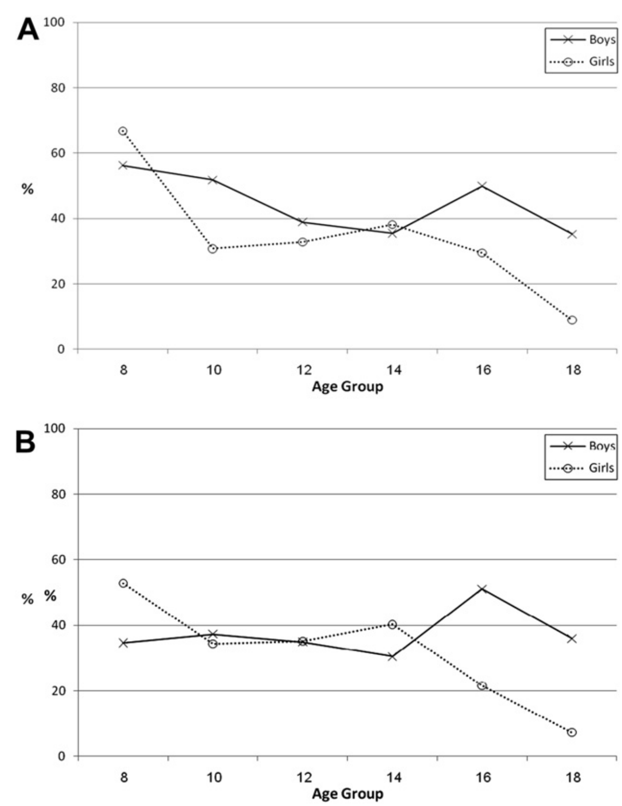

FIGURE 3. Correct age group identification rates for recordings of (**A**) sentences and (**B**) vowels in the six age groups.

TABLE 4.
Distribution of Listeners' Responses to the Age Question for Boys and Girls Within Each Age Group and Overall Identification Rates with ±1 Age Group

		Perceived Age, %						
Gender	Age Group (y)	8	10	12	14	16	18	±1 Age Group, %
Boys	8	45.53	39.21	12.24	2.89	0.13	0.00	84.74
	10	14.61	44.61	27.11	12.63	1.05	0.00	86.32
	12	1.71	23.42	36.97	28.03	8.03	1.84	88.42
	14	0.00	1.71	11.71	33.03	35.39	18.16	80.13
	16	0.00	0.13	1.97	18.03	50.53	29.34	97.89
	18	0.00	0.26	0.66	14.34	49.08	35.66	84.74
Girls	8	59.87	31.84	6.84	1.32	0.13	0.00	97.71
	10	37.89	32.63	18.42	8.03	2.24	0.79	88.95
	12	1.71	20.92	34.08	28.29	12.50	2.50	83.29
	14	0.92	11.58	31.32	39.21	13.16	3.82	83.68
	16	0.79	8.68	21.71	32.76	25.52	10.53	68.82
	18	0.39	7.37	30.53	31.45	22.11	8.16	30.26

correct identification of age/gender and the candidate variables for entering the model were age, gender, height, weight, neck circumference, and the age × gender interaction.

Analyses were performed separately for the two stimuli (vowels and sentences). Three model-building approaches were used: forced entry (all variables in the model), forward selection, and backward elimination. Whenever the interaction between age and gender was significant, both parameters were forced into the model to ensure correct estimation. Table 5 presents *P* values and parameter estimates for the measures included in the three analyses.

Results show that correct age identification in vowels was associated significantly with gender ($P < 0.0001$) and age × gender interaction ($P < 0.0001$) but not with age. For the sentences stimuli, age identification was associated significantly with chronological age ($P = 0.0139$), and age × gender interaction ($P = 0.0122$) but not with gender. None of the children's physical characteristics (height, weight, and neck circumference) were associated with correct age identification in both stimuli.

Different results were obtained when the same analysis was performed for gender identification. These are presented in Table 6. For the vowels stimuli, gender, age × gender interaction, and height were significantly associated with gender identification ($P = 0.0003$, 0.0004, 0.0088, respectively). Age, weight, and neck circumference were not associated with gender identification. For sentences stimuli, the linear regression yielded inconsistent results: the forward selection approach resulted in a model including height ($P = 0.0017$), age × gender interaction ($P = 0.0411$), and gender ($P = 0.2224$). However, when age was forced into the model, none of the parameters showed significant association with gender identification. Weight and neck circumference were not associated with gender identification in either of the stimuli.

Finally, to examine the possibility that listeners who have children might respond differently from those who do not have children, a comparison between the mean correct identification score (calculated average for gender and age identification obtained for each listener) in the two subgroups was performed. Of the 38 listeners, 19 had children and 19 did not. The mean correct identification score was 59.68% for the "mothers" group and 59.26% for the "nonmothers" group. This minor difference was fond statistically insignificant, using a two-sample *t* test ($t_{36} = 0.49$, $P = 0.625$). Furthermore, a Pearson correlation between the number of children each listener had and her correct response rates yielded a nonsignificant result ($r = 0.0279$, $P = 0.868$). These results show that listeners' responses were not different for mothers and nonmothers.

TABLE 5.
Summary of Results of the Linear Regression Model for Age Identification

		Vowels		Sentences	
Variable	DF	Parameter Estimate	*P* Value	Parameter Estimate	*P* Value
Intercept	1	0.30458	0.0003	0.66453	<0.0001
Age	1	0.00045	0.3744	−0.00141	0.0139
Gender	1	0.50426	<0.0001	0.21519	0.0955
Age × gender	1	−0.00359	<0.0001	−0.00203	0.0122

Abbreviation: DF, degrees of freedom.

TABLE 6.
Summary of Results of the Linear Regression Model for Gender Identification

		Sentences		Vowels	
Variable	DF	Parameter Estimate	*P* Value	Parameter Estimate	*P* Value
Intercept	1	0.25168	0.3096	−0.52357	0.0416
Age	1	−0.00016	0.9038	0.00143	0.2991
Gender	1	0.15062	0.3352	0.59757	0.0003
Height	1	0.00440	0.0900	0.00707	0.0088
Age × gender	1	−0.00152	0.1324	−0.00376	0.0004

Abbreviation: DF, degrees of freedom.

DISCUSSION

The aim of this study was to evaluate listeners' ability to identify children's gender and age, based on their voice. In general, listeners demonstrated the ability of identifying gender and age well above chance. Gender was identified correctly in close to 82% of cases, and age group was assigned correctly in over 37%. Moreover, when a single age group error was allowed, success in the age group assignment task exceeded 80%.

Listeners' ability to identify speakers' gender in *adults*, based on recordings of their voice, was demonstrated in many previous studies.[5,7,14,25] On the other hand, the ability to identify *children's* gender, based on their voice, was examined only in a relatively limited number of studies. This could be probably attributed mainly to the classic view that until puberty there are no pronounced physical differences between the male and female larynges,[17] thus assuming that there would be no vocal differences between genders. Nonetheless, studies that examined this view have questioned this assumption. In one study[18] in which voices of children in the age range of 6–8 years were evaluated, it was reported that gender identification rates ranged between 61% and 81%. In a later study[26] that included children in four age groups (4, 8, 12, and 16 years), it was reported that correct gender identification rates varied between 56% and 99.7%. In contrast to these two studies, Traunmüller and Bezooijen[19] reported that listeners had difficulties in identifying young children's gender, based on their voice. In that study, listeners were asked to identify the gender of children aged 5, 7, 9, and 11 years, and correct identification rates ranged between 8% and 40%.

The differences in the results among these past studies can be explained primarily by methodological differences. Listeners' success in the gender identification task is, apparently, affected by children's age, their gender, and the type of the recorded signal. Whereas previous studies examined limited age ranges, the present study examined voice recordings of children between the ages of 8–18 years (ie, pre- and postpuberty). Data show that, indeed, these three parameters (children's age, gender, and stimuli) have affected gender identification rates. Whereas girls were identified correctly at a consistent rate of approximately 80%, across all age groups (pre- and postpuberty), boys' identification rates varied significantly. Before the voice transition at puberty (ie, age groups 8–12 years), boys were identified correctly at approximately 60%, based on productions of isolated vowels and at approximately 80% based on sentences. In contrast, after puberty (ie, age groups 14–18 years), boys were identified correctly almost unanimously.

These results lead to several conclusions. First, the increase in correct perceptual rates of gender and age of younger versus older boys was observed around the age in which voice change (ie, vocal mutation) typically occurs.[17] Our data show that the voices of boys after puberty carry acoustic information that enables listeners to identify their gender with absolute certainty, whereas girls in the same age groups are identified correctly at lower rates. Because the present study is perceptual, we currently cannot ascertain the acoustic features on which listeners relied in making the gender judgment. Yet, it is clear that whereas girls are identified similarly at all age groups, boys' identification rates improve after puberty. Previous studies have concluded that the rapid laryngeal development observed in boys during puberty leads to a prominent acoustic change of a lowering of the fundamental frequency, which is perceived by listeners as a pronounced drop in pitch.[27–29] This vocal change was reported to occur between the ages of 12.5 and 14.5 years.[30] The parallel change in girls is lesser in magnitude, and it is spread over a longer period than in boys; thus it produces a smaller and slower perceptual change in pitch. This fact is evident in the prominent elevation in correct identification rates of male voices around puberty, in comparison with the relatively stable rates observed for female voices.

Differences in vocal tract dimensions between boys and girls were also explored in previous studies. Aronson,[17] for example, reported that until puberty, there are no physical differences in larynx size between boys and girls. Fitch and Giedd[29] added to that and reported a lack of morphological differences between genders in articulatory mechanism. On the other hand, other studies have reported on specific gender-related differences between genders, such as pharyngeal length[31] or oral tract length.[32] Regardless of this controversy, most researchers agree that even prepuberty voice carries gender-related acoustic information.[26,33,34] These studies suggested that the acoustic cues for gender in prepubescent children are related primarily to the supraglottis (ie, vocal track), while fundamental frequency is perceived as providing only secondary cues at this age range. However, after puberty, as fundamental frequency differences grow to be more prominent, they become a dominating cue for gender recognition.

The assumption that listeners rely on supraglottal features of speech for identification of gender in young children is supported by our results. When young boys produced isolated vowels, they were identified correctly as males at a rate of approximately 60%. On the other hand, when they produced speech, listeners were able to identify their gender correctly at over 80%. This improvement in gender identification was observed only for boys, but not for girls. Apparently, the young boys' speech signal carries acoustic information that listeners can use for identification of their gender. This acoustic information is not presented (or less pronounced) in isolated vowels. The speech signal provides coarticulation and prosody dynamic cues, beyond the fundamental frequency and the static supraglottal cues, which are provided in isolated vowels. Therefore, it is assumed that boys are making vocal tract and articulatory adjustments during speech, which enables listeners to identify them correctly. The explanation for the fact that young boys modify their speech differently than girls and that this modification is perceived by listeners as a perceptual cue for gender is beyond the scope of the present study. Yet, these findings seem to have implications to cultural and social dimensions of gender boundaries and gender identity, which should be explored in future research.

Identification of chronological age of adult speakers was shown to be a feasible task for listeners.[21–23,35,36] We were unable, however, to identify previous studies that evaluated listeners' ability to perform this task based on voices of children. Our results show that listeners were able to identify children's age well above chance. Moreover, when an error of ±1 age group was allowed, listeners were able to identify the children's age correctly between 68% and 98%.

Interestingly, although similar age identification rates were observed for both genders in most age groups, the one consistent finding was that older girls (ages 16–18) were identified as younger than their chronological age (Figure 3). Our findings do not provide a clear explanation for this result. In should be noted that although the increase in correct identification rate in the younger group (8 years) might be partially explained by a "floor effect," the lowering in the identification rate in the 18-year-old group cannot be explained by any equivalent statistic effect. It is possible, thus, that sociolinguistic factors have contributed to this finding, as it cannot be readily explained by physiological or developmental factors. Lee et al,[34] for example, suggested that there are gender differences in phonation and articulation that can be observed in childhood and adolescence and that these differences are attributed to sociocultural effects. It was suggested that boys attempt to produce a more masculine voice and speech pattern by adopting a lower pitch than their optimal pitch, lowering their jaw, and modifying the extent of lip rounding.[33,34,37] It is possible, thus, that such factors have contributed to the fact that the older girls were judged as younger, whereas boys were identified at a relatively stable rate across all age groups.

CONCLUSIONS

Although listeners' perception of speakers' gender and age is considered an intuitive task, relatively little is known about the manner in which it is done. The present study demonstrated that listeners are capable of identifying gender and age of children and adolescents, based on brief recordings of their voice and speech. It provides support to the assumption that there are audible differences between genders, that can be perceived by listeners even before the voice change that occurs in puberty. Future studies should correlate these findings with acoustic and more extensive physiological data and sociocultural factors.

REFERENCES

1. Scherer KR. Vocal communication of emotions: a review of research paradigms. *Speech Commun*. 2003;40:227–256.
2. Krauss RM, Freyberg R, Morsella E. Inferring speaker's physical attributes from their voices. *J Exp Soc Psychol*. 2002;38:618–625.
3. Collins SA, Missing C. Vocal and visual attractiveness are related in women. *Anim Behav*. 2002;65:997–1004.
4. Sataloff RT. *Voice Perspective*. London, San Diego: Singular Publishing Group, Inc.; 1998.
5. Childers DG, Wu K. Gender recognition from speech. Part II: Fine analysis. *J Acoust Soc Am*. 1991;90:1842–1856.
6. Evans S, Neave N, Wakelin D. Relationship between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biol Psychol*. 2006;72:160–163.
7. Gelfer MP, Mikos VA. The relative contribution of speaking fundamental frequency and format frequencies to gender identification based on isolated vowels. *J Voice*. 2005;19:544–554.
8. Cerrato L, Falcone M, Paoloni A. Subjective age estimation of telephonic voices. *Speech Commun*. 2000;31:107–112.
9. Fitch WT. The evolution of speech: a comparative review. *Trends Cogn Sci*. 2000;4:258–267.
10. Hardus ME, Lameira AR, Van Schaik CP, Wich SA. Tool use in wild orangutans modifies sound production: a functionally deceptive innovation? *Proc Biol Sci*. 2009;276:3689–3694.
11. Bruckert L, Lie'nart J, Lacroix A, Kreutzer M, Leboucher G. Women use voice parameters to assess men's characteristics. *Proc Biol Sci*. 2006;273: 83–89.
12. Gonzalez J. Estimation of speakers' weight and height from speech: a reanalysis of data from multiple studies by Lass and colleagues. *Percept Mot Skills*. 2003;96:297–304.
13. Smith DRR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sound. *J Acoust Soc Am*. 2005; 117:305–318.
14. Linville SE, Fisher HB. Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *J Acoust Soc Am*. 1985; 78:40–48.
15. Whiteside SP. Identification of a speaker's sex: a fricative study. *Percept Mot Skills*. 1998;86:587–591.
16. Lass NJ, Hendricks C, Iturriaga MA. The consistency of listener judgments in speaker height and weight identification. *J Phon*. 1980;8:439–448.
17. Aronson AE. *Clinical Voice Disorders*. 2nd ed. New York, NY: Theime Inc.; 1985.
18. Bennett S, Weinberg B. Sexual characteristics of preadolescent children's voices. *J Acoust Soc Am*. 1979;65:179–189.
19. Traunmüller H, van Bezooijen R. The auditory perception of children' age and sex. *Proceedings of the Third International Conference on Spoken Language Processing (ICSLP)*; Yokohama, Japan: The Acoustical Society of Japan; 3, 1171–1174. Available at: http://www.ling.su.se/staff/hartmut. Accessed July 16, 2007.
20. Ptacek PH, Sander EK. Age recognition from voice. *J Speech Hear Res*. 1966;9:273–277.
21. Shipp T, Hollien H. Perception of the aging male voice. *J Speech Hear Res*. 1969;12:703–710.
22. Ryan WJ, Burk KW. Perceptual and acoustic correlates of aging in the speech of male. *J Commun Disord*. 1974;7:181–192.
23. Neiman GS, Applegate JA. Accuracy of listener judgment of perceived age relative to chronological age in adults. *Folia Phoniatr*. 1990;42:327–330.

24. Minematsu M, Yamauchi K, Hirose K. Automatic estimation of perceptual ae using speaker modeling techniques. *Paper presented at the 8th European Conference on Speech Communication and Technology*; Geneva, Switzerland; 3005–3008. Available at: http://isca-speech.org/archive/eurospeech_2003/e03_3005.html. Accessed October 19, 2008.
25. Whiteside SP. Identification of a speaker's sex: a study of vowels. *Percept Mot Skills*. 1998;86:579–584.
26. Perry TL, Ohde RN, Ashmead DH. The acoustic bases for gender identification from children's voices. *J Acoust Soc Am*. 2001;109: 2988–2998.
27. Huber JE, Stathopoulos ET, Curione GM, Ash TA, Johnson K. Formants of children, women, and men: the effect of vocal intensity variation. *J Acoust Soc Am*. 1999;106:1532–1542.
28. Fuchs M, Froehlich M, Hentschel B, Stuermer I, Kruse E, Knauft D. Predicting mutual changes in the speaking voice of boys. *J Voice*. 2007;21: 169–178.
29. Fitch WT, Giedd J. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J Acoust Soc Am*. 1999;106: 1511–1522.
30. Hollien H, Green R, Massey K. Longitudinal research on adolescent voice change in males. *J Acoust Soc Am*. 1994;96:2646–2654.
31. Bennett S, Weinberg B. Acoustic correlates of perceived sexual identify in preadolescent children's voices. *J Acoust Soc Am*. 1979;66:989–1000.
32. Xue SA, Cheng RWC, Ng LM. Vocal tract dimensional development of adolescents: an acoustic reflection study. *Int J Pediatr Otorhinolaryngol*. 2010;74:907–912.
33. Whiteside SP, Hodgson C. Acoustic characteristics in 6-10-year-old children's voices: some preliminary findings. *Logoped Phoniatr Vocol*. 1999; 24:6–13.
34. Lee S, Potamianos A, Naratanan S. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *J Acoust Soc Am*. 1999;105:1455–1468.
35. Hartman DE. The perceptual identity and characteristics of aging in normal male adult speakers. *J Commun Disord*. 1979;12:51–61.
36. Huntley R, Hollien H, Shipp T. Influence of listener characteristics on perceived age estimations. *J Voice*. 1987;1:49–52.
37. Fox RA, Nissen SL. Sex-related acoustic changes in voiceless English fricatives. *J Speech Lang Hear Res*. 2005;48:753–765.