The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age[a)]

David R. R. Smith and Roy D. Patterson
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 15 April 2005; revised 2 August 2005; accepted 3 August 2005)

Glottal-pulse rate (GPR) and vocal-tract length (VTL) are related to the size, sex, and age of the speaker but it is not clear how the two factors combine to influence our perception of speaker size, sex, and age. This paper describes experiments designed to measure the effect of the interaction of GPR and VTL upon judgements of speaker size, sex, and age. Vowels were scaled to represent people with a wide range of GPRs and VTLs, including many well beyond the normal range of the population, and listeners were asked to judge the size and sex/age of the speaker. The judgements of speaker size show that VTL has a strong influence upon perceived speaker size. The results for the sex and age categorization (man, woman, boy, or girl) show that, for vowels with GPR and VTL values in the normal range, judgements of speaker sex and age are influenced about equally by GPR and VTL. For vowels with abnormal combinations of low GPRs and short VTLs, the VTL information appears to decide the sex/age judgement. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2047107]

PACS number(s): 43.66.Lj, 43.71.Bp, 43.71.An [RLD] Pages: 3177–3186

I. INTRODUCTION

When the radio or the telephone presents us with a previously unknown speaker, we rapidly develop a distinct impression of whether the speaker is an adult or a child, and if an adult, whether it is a man or a woman. This paper is concerned with the acoustic cues that people use to make these judgements. One highly-salient cue is voice pitch; adult men have low pitches, young children have high pitches, and adult women lie in the middle. Pitch is determined by the rate of opening and closing of the vocal folds [glottal-pulse rate (GPR)]. Another potent cue is vocal-tract length (VTL); large adult men have the longest VTLs, children have the shortest VTLs, and women have intermediate VTLs (Fitch and Giedd, 1999). Differences in VTL lead to shifts in the frequency of the prominent spectral peaks (formants) of speech (Fant, 1970). We have shown that changes in simulated VTL of as little as 7% can be reliably discriminated (Smith, Patterson, Turner, Kawahara, and Irino, 2005). It is unclear how the different effects of GPR and VTL are combined to influence the perception of speaker size, sex, and age. The purpose of this paper was to measure the interaction of GPR and VTL in judgements of speaker size, and to the categorization of speakers according to sex and age (man, woman, boy, or girl).

Recently, we have shown that when listeners are given two sequences of four vowels, and the simulated VTL for one sequence is longer than for the other, listeners are capable of discriminating VTL differences of 6–10%, over a wide range of GPR and VTL values (Smith *et al.*, 2005). The experiments used a 2AFC discrimination task which only requires the listener to make a relative size judgement. A second motivation for the present paper was to determine the extent to which listeners can make consistent judgements about speaker size, and consistent judgements about the sex and age of the speaker (man, woman, boy, or girl).

A. Background

Much of the variability between the voices of men, women, and children is due to differences in the mass of the vocal folds and the length of the vocal tract. For a given vowel, these differences lead to significant differences in both the GPR (perceived as voice pitch) and the frequencies of the most prominent spectral peaks (formants). The length and shape[1] of the vocal tract (VT) causes certain frequencies to be reinforced and attenuated. The length of the supralaryngeal VT is highly-correlated with speaker height, increasing with both age and sex (Fitch and Giedd, 1999). The longer the VT, the more the formant frequencies are shifted towards lower frequencies (Fant, 1970). As a child grows between the ages of 4 and 12 (puberty), there is a steady increase in VTL with a concomitant decrease in the formant frequencies. The formant frequencies of adult males decrease by about 32% from their values at age 4, while the formant frequencies of adult females decrease by about 20% (Huber, Stathopoulos, Curione, Ash, and Johnson, 1999). Within groups of adult men and women, the correlation between speaker height and formant frequency weakens (González, 2004). Nevertheless, a quantitative analysis by Turner and Patterson (2003) of the variability in the classic vowel data of Peterson and Barney (1952) shows that, within a given vowel cluster, speaker size is the largest source of variation. There is also a strong correlation between body size and formant-related parameters in rhesus monkeys (Fitch, 1997),

[a)]Portions of this work were presented in "The perception of sex and size in vowel sounds," British Society of Audiology, UCL London, United Kingdom, 2004, and "Perception of speaker size and sex of vowel sounds," Acoustical Society of America, Vancouver, Canada, 2005.

and in the vowel-like grunts of baboons, the formants of adult males are about 25% lower in frequency than those of the females (Rendall, Owren, Weerts, and Hienz, 2004). Indeed, the presence of size information has been demonstrated in a diverse range of vertebrate species (e.g., frogs, Fairchild, 1981, Narins and Smith, 1986; birds, Fitch, 1999; lions, Hast, 1989; dogs, Riede and Fitch, 1999).

The relationship between GPR and speaker size is more complex. Certainly, there is a strong link between speaker sex and pitch (Darwin, 1871; Morton, 1977). Adult males have pitches about an octave lower than adult females primarily because the vibrating segments of the male vocal folds are about 60% longer than those of the female and, thus, they are much more massive (Titze, 1989). This sexually dimorphic difference in pitch is also present in the vowel-like grunts of adult baboons (Rendall *et al.*, 2004). In a statistical clustering analysis of human adult male and female speech sounds, both GPR and VTL were highly successful as single-factors for classifying speaker sex. However, GPR was much less effective than VTL in correctly classifying individual speakers (Bachorowski and Owren, 1999). The sexual dimorphism in GPR is attributable to increased testosterone at puberty in males which stimulates growth in the laryngeal cartilages (Beckford, Rood, and Schaid, 1985). However, there is no direct correlation between body size and GPR within adult men and women (e.g., Lass and Brown, 1978; Künzel, 1989; Hollien, Green, and Massey, 1994). This is to be expected because VTL is dictated by the size of the cranium whilst the vocal folds are not constrained by any bony structure (Negus, 1949). The correlation between GPR and speaker size is also weakened by our use of GPR variation to make prosodic distinctions, such as the rising pitch contour of the interrogative sentence. Thus, while GPR provides a strong cue to speaker sex in adults (cf. Bachorowski and Owren, 1999), it provides a more variable cue to speaker size.

B. The interaction of GPR and VTL in judgements of speaker size, sex, and age

We wished to determine how GPR and VTL interact in the perception of speaker size. Given the strong correlation of VTL with speaker size, we would expect that VTL has a substantial effect on the perception of speaker size. There is also a correlation between GPR and size, although it is not as strong, and pitch is a highly salient property of a person's voice. With regard to the perception of speaker sex and age, we wished to determine the combinations of GPR and VTL that are associated with the categories used naturally by people, that is, man, woman, boy, and girl. Specifically, we wished to demonstrate that listeners would reliably assign combinations of GPR and VTL found in the normal population to the expected category, and we wished to investigate how they would extend the use of the categories to combinations of GPR and VTL well beyond the range normally encountered. Finally, we wanted to compare the listener's speaker-size judgements with their use of the categories, man, woman, boy, girl, particularly in the extended region of GPR and VTL values.

II. METHOD

Listeners were presented isolated vowels scaled over a large range of GPR and VTL values, and requested to make two judgements about each vowel: the height of the speaker (seven point descriptive rating) and their natural category (man, woman, boy, or girl).

A. Stimuli

The five English vowels (/a/, /e/, /i/, /o/, /u/) of an adult male (author, R.P.) were recorded in natural /hVd/ sequences (i.e., *haad*, *hayed*, *heed*, *hoed*, *who'd*), using a high-quality microphone (Shure SM58-LCE) and a 44.1 kHz sampling rate. The vowels were sustained (e.g., *haaaad*) to allow isolation of a stationary vowel component of relatively long duration, which was free of co-articulation with the preceding /h/ and the following /d/.

The speaker's vocal-tract shape determines the vowel type. The speaker's VTL determines the scale of the resonance and, thus, the position of the vowel pattern along the frequency dimension. The scaling of the vowels was performed by STRAIGHT (Kawahara, Masuda-Kasuse, and de Cheveigne, 1999; Kawahara and Irino, 2004). This sophisticated speech processing software uses the classical source-filter theory of speech (Dudley, 1939) to segregate GPR information from the spectral-envelope information associated with the shape and length of the vocal tract. Liu and Kewley-Port (2004) have reviewed STRAIGHT and commented favorably on its ability to manipulate formant-related information. STRAIGHT produces a GPR-independent spectral envelope that accurately tracks the motion of the vocal tract throughout the utterance. Once STRAIGHT has segregated a vowel into a GPR contour and a sequence of spectral-envelope frames, the vowel can be resynthesized with the spectral-envelope dimension (frequency) expanded or contracted, and the GPR dimension (time) expanded or contracted. Moreover, the operations are largely independent. Utterances recorded from a man can be transformed to sound like a women or a child; examples are provided on our web page.[2] The resynthesized utterances are of high quality even when the speech is resynthesized with GPR and VTL values well beyond the normal range of human speech (provided the GPR is not much greater than the frequency of the first formant, cf. Smith *et al.*, 2005). STRAIGHT is reviewed in Kawahara and Irino (2004).

The scaling of GPR consists of expanding or contracting the time axis of the sequence of glottal events. The scaling of VTL is accomplished by compressing or expanding the spectral envelope of the speech linearly along a linear frequency axis. On a logarithmic frequency axis, the spectral envelope shifts along the axis as a unit. The change in VTL is described by the spectral envelope ratio (SER), that is, the ratio of the unit on the new frequency axis to that of the axis associated with the original recording. Values of SER less than unity indicate lengthening of the vocal tract to simulate larger men, and SERs greater than unity indicate shortening of the vocal tract to simulate smaller men, women, and children. The SER values of STRAIGHT can be converted to VTL values by noting that (a) the speaker of our original

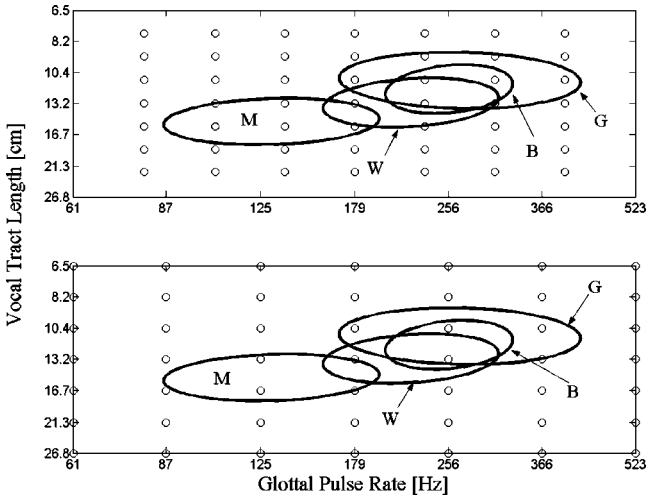

FIG. 1. The open circles show the GPR and VTL combinations of the stimuli used in the speaker size and sex/age categorization experiments. The circles in the top panel show the "narrower" range of (7×7) sample points (GPRs of 80, 105, 137, 179, 234, 306, and 400 Hz; VTLs of 7.8, 9.3, 11.0, 13.2, 15.7, 18.7, and 22.2 cm). The bottom panel shows the "wider" range (GPRs of 61, 87, 125, 179, 256, 366, and 523 Hz; VTLs of 6.5, 8.2, 10.4, 13.2, 16.7, 21.3, and 26.8 cm). The four ellipses show the normal range of GPR and VTL values in speech for men (M), women (W), boys (B), and girls (G), derived from the data of Peterson and Barney (1952). Each ellipse contains 99% of the individuals from the respective category.

vowels was of normal height, (b) that the VTL of the average-sized adult male is 15.5 cm (cf. Fitch and Giedd, 1999), and (c) assuming that formant frequencies scale linearly with VTL (Fant, 1970). The data in this study are reported in GPR and VTL units.

Following the scaling of GPR and VTL by STRAIGHT, a cosine-squared gating function (10-ms onset, 30-ms offset, 465-ms plateau) was used to select a stationary part of the vowel. The RMS level was set to 0.025 (relative to maximum ±1). The stimuli were played by a 24-bit sound card (Audigy 2, Sound Blaster), through a TDT anti-aliasing filter with a sharp cutoff at 10 kHz and a final attenuator, and presented binaurally to the listener over AKG K240DF headphones. Listeners were seated in a double-walled, IAC, sound-attenuating booth. The sound level of the vowels was 66 dB SPL.

B. Procedures

The experiments were performed using a single-interval, two-response paradigm. The listener heard a scaled version of one of five stationary English vowels (/a/, /e/, /i/, /o/, /u/), and had to make one judgement about the *size* of the speaker (very short, short, quite short, average, quite tall, tall, very tall)[3] and a second judgement about the *sex/age* of the speaker (man, woman, boy, girl). The order in which the judgements were made was left to the listener. Size and sex/age judgements were made by selecting the appropriate button on a response box displayed on a monitor in the booth. The level of the vowel was roved in intensity over a 10 dB range. Since the judgements are subjective there was no feedback.

The experiment was performed for two ranges of GPR and VTL values as shown in Fig. 1. The narrower range [Fig. 1(a)] was chosen to encompass the range of GPR and VTL values encountered in the normal population; GPR varied from 80 to 400 Hz in six logarithmic steps (seven sample points), and VTL ranged from 22.2 to 7.8 cm in six logarithmic steps (seven sample points). The four ellipses show estimates of the normal range of GPR and VTL values in speech for men, women, boys, and girls, derived from the Peterson and Barney (1952) vowel database. In each case, the ellipse encompasses 99% of the individuals in the Peterson and Barney data for that category of speaker.[4] The wider range [Fig. 1(b)] was chosen to extend the judgements well beyond the values encountered in everyday speech; GPR varied from 61 to 523 Hz in six logarithmic steps, and VTL ranged from 26.8 to 6.5 cm in six logarithmic steps. These VTLs simulate speakers ranging from a small child 0.6-m high (VTL=6.5 cm) to a giant 3.7-m high (VTL=26.8 cm).[5]

A run of judgements consisted of one presentation of each GPR-VTL combination for all five vowels, presented in a pseudo-random order (a total of 7 GPRs $\times$7 VTLs $\times$5 vowels, or 245 trials). Each run took approximately 30 min to complete. Each listener contributed a block of five runs to the database for the narrower range of judgements about speaker size and sex/age, and a block of five runs to the database for the wider range of judgements about speaker size and sex/age. The starting range [cf. Fig. 1(a) or Fig. 1(b)] was counterbalanced across listeners. The overlap in GPR and VTL values in the two ranges allows an across-condition test of the consistency of size and sex/age judgements. This helps us to see how different ranges of input sounds are stretched to the available seven point response, and how that mapping is influenced by the frames of reference provided by the two different ranges of GPR and VTL of the vowel sounds.

Eight listeners participated in the experiments, three male and five female. They ranged in age from 21 to 39 years. All had normal absolute thresholds at 0.5, 1, 2, 4, and 8 kHz.

III. RESULTS

Broadly speaking, the results from the two stimulus ranges show the same effects (Fig. 2), and they show that judgements of speaker size (Fig. 3) and sex/age (Fig. 4) are affected both by GPR and VTL. Listeners reliably reported that vowels spoken with a very low GPR and a very long VTL came from a very tall person (Fig. 3), and increasing the GPR or shortening the VTL reliably reduced the reported size of the speaker (Fig. 3). The influence of VTL upon these size judgements was very strong, as shown by the marked fall-off in reported speaker size as VTL shortened. Examination of the speaker size judgements over the course of the experiment showed little evidence of learning; listeners can do the task at near asymptotic levels almost straightaway. In the perception of sex and age (man, woman, boy, or girl), GPR and VTL had about the same influence in the narrower range about the normal ellipses (Fig. 4), but in the wider range, for the more unusual combinations of GPR and VTL, it is VTL information which appears to decide the sex/age judgement (Fig. 4).

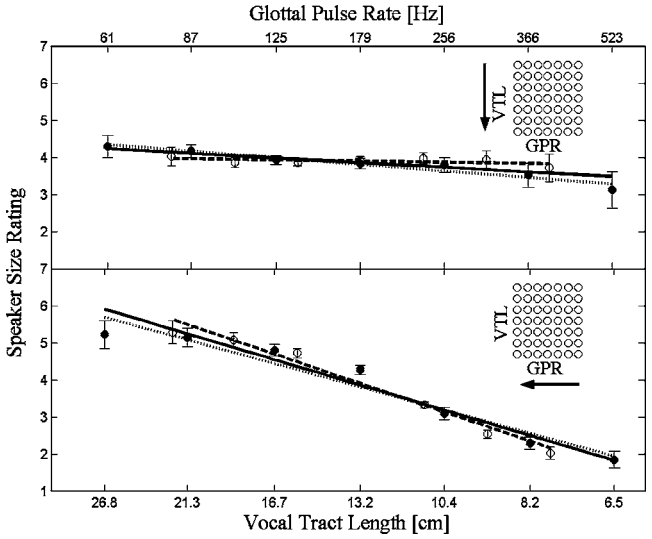

FIG. 2. Speaker size judgements collapsed across VTL (upper panel) and GPR (lower panel), separately for the narrower and the wider ranges (cf. Fig. 1). The arrows on the inset show the dimension over which the data were collapsed. The open circles show the data from the narrower range and the solid circles from the wider range. The dotted line is the best fitting line for the wider range; the dashed line is the best fitting line for the narrower range, and the solid thick line is the best fitting line for the combined data. The error bars are ± one standard error of the mean (calculated from the average of the eight listeners, where each listener's average is based on the seven values per point over which the data were collapsed). Each datum point is based on 1400 trials.

A. The effect of stimulus range on speaker size judgements

We will begin by comparing the size judgements obtained from the two ranges of GPR and VTL values [cf. Fig. 1(a) and 1(b)] because the results show that they are essentially sampling the same size surface, and so the data from the two ranges can be combined for subsequent analyzes. Figure 2 shows the column and row averages for both the narrower and the wider ranges; specifically, the upper panel shows the data for the two ranges collapsed across VTL (column averages), and the lower panel shows the data collapsed across GPR (row averages), as indicated by the inset schematic. In both panels, the data from the two ranges are seen to fall along similar lines (dashed and dotted for the narrower and wider ranges, respectively). For the GPR column averages in the upper panel, the slope of the line fitted to the data from the narrower range is slightly shallower than the slope of the line fitted to the data from the wider range. For the VTL row averages in the lower panel, the reverse is true; the slope for the wider range is slightly shallower than that for the narrower range. In both cases, when a single line (solid) was fitted to the combined data from the two ranges, it was found to provide an excellent fit to the full data set. Accordingly, the data from the two ranges were combined for subsequent analyzes.

B. The interaction of GPR and VTL in judgements of speaker size

The size judgements for both the wider and narrower ranges are presented in Fig. 3 as a 2D surface plot, averaged over the five vowels and eight listeners. The abscissa is GPR and the ordinate is VTL, both on logarithmic axes; color shows perceived speaker size. The GPR-VTL points where speaker size ratings were measured are shown by the open circles; between the data points, the surface was derived by interpolation.[6] The consistency of the size ratings across the two ranges (cf. Fig. 1) is shown by the similarity of the ratings for adjacent stimuli from the two data sets. The seven categories of the size rating scale, from "very short" to "very tall," were assigned ordinal values from 1 to 7, and they are represented by the spectrum of colors from dark-blue (1) to brown-red (7). The surface shows, as expected, that the combination of a long vocal tract with a low pitch is consistently heard as a large or very large person, and the combination of a short vocal tract with a high pitch is consistently heard as a small or very small person. The four ellipses show the normal range of GPR and VTL in speech for men, women, boys, and girls (Peterson and Barney, 1952). In each case, the ellipse encompasses 99% of the individuals in the Peterson and Barney data for that category of speaker (man, woman, boy, or girl). The figure shows that, although the perception of speaker size is affected both by VTL and GPR, the effect of VTL is stronger than that of GPR, at least in this coordinate system. For instance, for a constant GPR of 61 Hz, as we move vertically from a long VTL of 26.8 cm to a short VTL of 6.5 cm, the size rating goes from 6.2 ("tall") to 1.7 ("short"). The greatest change in perceived size as a function of change in GPR is for a VTL of 26.8 cm, where the size rating goes from 6.2 ("tall") at 61 Hz to 4.0 ("average") at 523 Hz.

The change in the perception of speaker size as a function of GPR and VTL was quantified in terms of the slopes of lines across the size surface in Fig. 3 parallel to the GPR and VTL axes. Perceived speaker size is shown as a function of GPR for three values of VTL in Fig. 5, namely, the two extreme VTLs (6.5 and 26.8 cm) associated with very short and very tall people, and a central value (13.2 cm) associated with an average-sized woman. Regression lines were fitted to the speaker size ratings as a function of the natural logarithm of GPR (solid lines in Fig. 5). They show that changes in GPR have the most effect when VTL is at its longest (26.8 cm; slope of −1.04). As VTL decreases to 13.2 cm, the slope decreases by about 60% (slope of −0.40), and as it decreases further to 6.5 cm, the slope becomes flat (0.01), indicating no change in speaker size whatsoever. The negative correlation between GPR and perceived speaker size is highly significant at the longer VTLs of 13.2 and 26.8 cm ($\rho < 0.001$ and $\rho \ll 0.001$, respectively, based on a one-tailed Spearman's rank order correlation test for nonparametric variables); the correlation is obviously not significant when VTL is 6.5 cm.

Similarly, perceived speaker size is shown as a function of VTL, for three GPR values in Fig. 6; namely, 61, 179, and 523 Hz. Again, they are the extreme values from the wider range (61 and 523 Hz), and the central value associated with an average-sized woman (179 Hz). Regression lines were fitted to the speaker size ratings as a function of the natural logarithm of VTL (solid lines in Fig. 6). The slopes of these VTL lines are all steeper than those of the GPR lines in Fig. 5. The slopes of these VTL lines become steeper as GPR

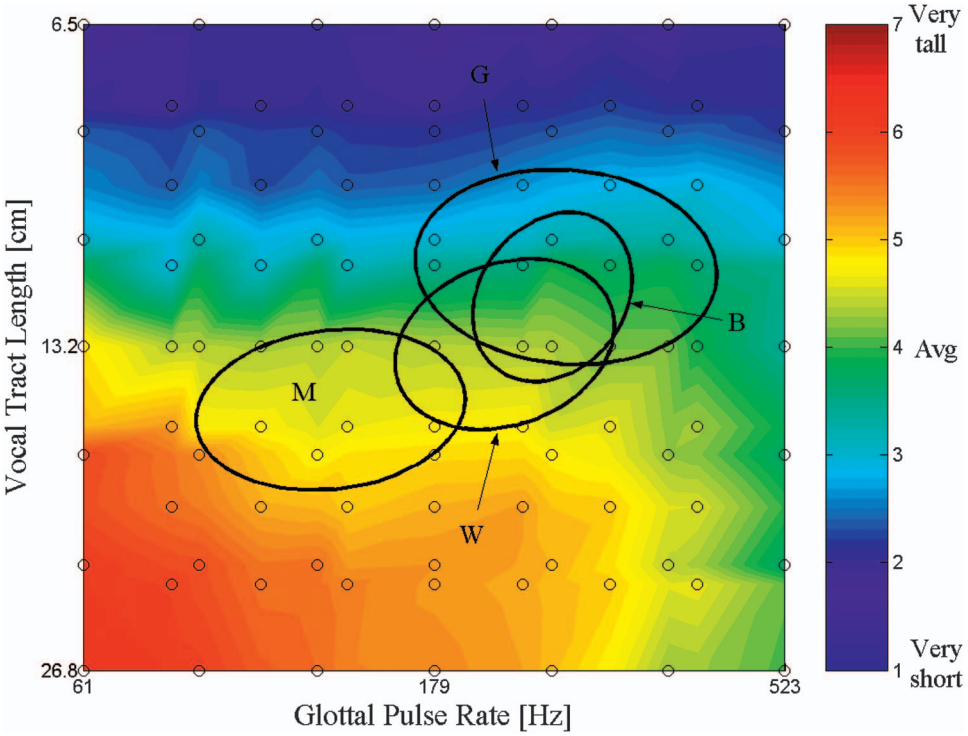

FIG. 3. Perceived size (in color) as a function of GPR and VTL on logarithmic axes. The size scale from "very short" to "very tall," is represented by the spectrum of colors from dark-blue (1) to brown-red (7). The points where speaker size ratings were measured are shown by the open circles; between the data points, the surface was derived by interpolation. The data were averaged across all five vowels and eight listeners, so each point is based on 200 trials. The four ellipses show the range of GPR and VTL in speech for men (M), women (W), boys (B), and girls (G), as derived from the data set of Peterson and Barney (1952).

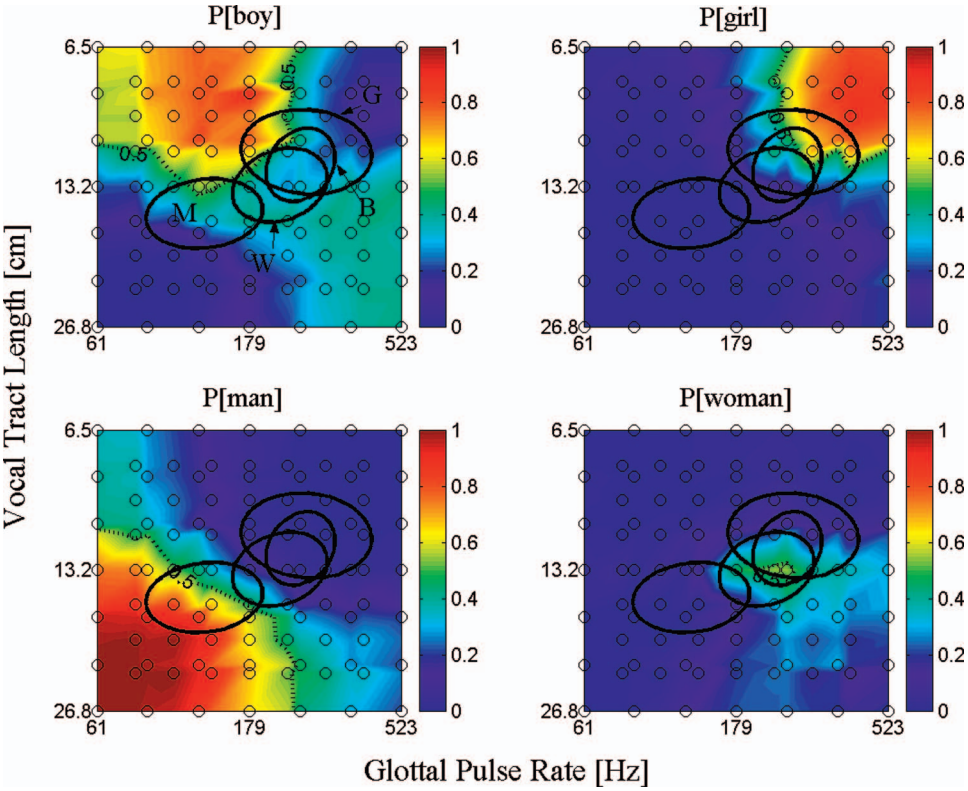

FIG. 4. Sex and age categorizations. The data are presented as 2D surface plots with color showing probability of assigning a given GPR-VTL combination to one of four categories (man, woman, boy, or girl). The points where sex/age judgements were collected are shown by the open circles; between the data points the surface was derived by interpolation. At each GPR-VTL point, the probabilities from the four panels sum to 1 (imagine the four separate 2D maps stacked vertically and aligned over each other). The data is averaged across all five vowels and eight listeners (each sample point probability based on 200 trials). The dotted black contour line marks the classification threshold, that is, a probability $\geqslant 0.50$ of consistently choosing one category out of the four available. The region of GPR-VTL values enclosed by this line defines a region categorized as one particular sex or age. The four ellipses show the range of GPR and VTL in speech for men (M), women (W), boys (B), and girls (G), as derived from the data set of Peterson and Barney (1952).

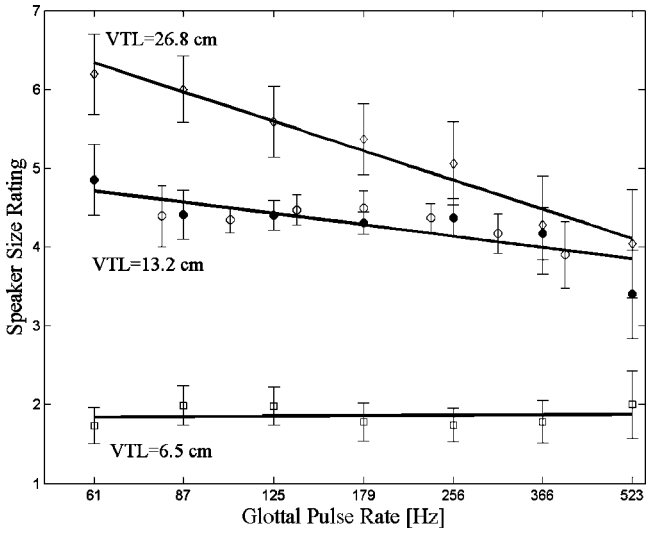

FIG. 5. Perceived speaker size as a function of GPR, for VTLs of 6.5, 13.2, and 26.8 cm. The open and solid circles show data from the narrower and wider stimulus ranges, respectively. The solid lines show the best-fitting regression lines for perceived speaker size rating as a function of the natural logarithm of GPR. The error bars are ± one standard error of the mean (calculated from the average of the eight listeners). Each datum point is based on 200 trials.

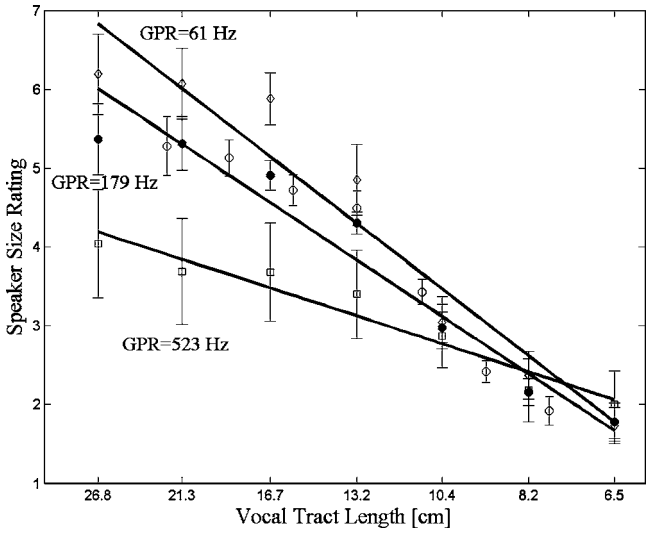

FIG. 6. Perceived speaker size as a function of VTL, for GPRs of 61, 179, and 523 Hz. The solid lines show the best-fitting regression lines for speaker size rating as a function of the natural logarithm of VTL. For all other details see Fig. 5.

decreases; the gradient is 1.50, 3.06, and 3.57 for GPRs of 523, 179, and 61 Hz, respectively. The correlation between VTL and perceived speaker size is highly significant for all three lines ($\rho \ll 0.001$ based on a one-tailed Spearman's rank order correlation test for nonparametric variables).

Figures 5 and 6 show an interaction between GPR and VTL in the perception of speaker size, especially at extreme GPR or VTL values. Simulated speakers that would only stand two feet tall, with very short VTLs (Fig. 5, VTL =6.5 cm), are always judged as short regardless of their GPR. Simulated giants of 12 feet (Fig. 5, VTL=26.8 cm) are always heard as above average height, but their estimated height declines as GPR increases. Figure 6 shows that the perception of speaker size is strongly affected by VTL, but that the effect weakens as GPR increases (cf. the decrease in slope for GPRs of 61, 179, and 523 Hz).

C. The interaction of GPR and VTL in judgements of sex and age

The speaker sex and age judgements from both the narrower and wider ranges of GPR-VTL values (cf. Fig. 1) are presented as 2D surface plots in Fig. 4; the results have been averaged over the five vowels and eight listeners. The results from the two ranges are entirely compatible, just as they were in the size rating experiments. The four panels show the probability of classifying a vowel with a specific GPR-VTL combination as a man, woman, boy, or girl. The probability of classification is shown by color, ranging from 0 (dark-blue) to 1 (brown-red). For each combination of GPR and VTL, the probabilities from the four panels sum to 1.0. The abscissa is GPR and the ordinate is VTL, both on logarithmic axes. The open circles show the combinations of GPR and VTL presented to the listeners; between these data points, the surfaces have been generated by interpolation. The dotted black lines bound regions of GPR-VTL where listeners consistently choose one category out of the four available to them. Within these regions, the probability of choosing the given combination of sex and age is greater than 0.5. The four ellipses show estimates of the normal range of GPR and VTL in speech for men, women, boys, and girls (Peterson and Barney, 1952). The ellipse for men does not intersect with the ellipses for girls and boys, whereas the ellipse for women intersects with the ellipses of all of the other groups. The ellipse for boys lies almost entirely within that for girls, and the overlap of the ellipses for boys and women is about 50%. Figure 4 shows that both GPR and VTL affect the perception of a speaker's sex and age, as expected, and that they interact, producing consistent responses in different regions across the GPR-VTL plane.

In the two quadrants of the GPR-VTL plane that represent the majority of normal human voices (lower left and upper right), the relationship between the sex/age category that the listener perceives and the combination of GPR and VTL in the vowel is straightforward. Vowels with low GPRs and long VTLs, in the lower left-hand quadrant of the GPR-VTL plane, are overwhelmingly categorized as men (lower left-hand panel of Fig. 4), and this quadrant contains the ellipse for men. This quadrant also contains vowels with lower than normal GPRs and longer than normal VTLs, and listeners consistently adopt the nearest category which is "man," as would be expected. Outside the $p \geqslant 0.5$ contour (dashed line), the probability of responding "man" drops rapidly, and only a small proportion of the "man" responses occur in the region above the negative diagonal.

Vowels with high GPRs and short VTLs, in the upper right-hand quadrant of the GPR-VTL plane, are predominately categorized as girls (upper right-hand panel of Fig. 4); this quadrant contains the ellipse for girls *and* the ellipse for boys, but the ellipse for girls extends to higher GPRs and shorter VTLs, so it is arguably the more natural category to adopt. This quadrant also contains vowels with higher than

normal GPRs and shorter than normal VTLs, and once again, listeners consistently adopt the nearest category which in this case is "girl." Outside the $p \geq 0.5$ contour, the probability of responding "girl" drops rapidly and the response "girl" is almost never used in the region below the negative diagonal. So, the categories "man" and "girl" are used consistently, and the combinations of GPR and VTL associated with these sex/age categories are mutually exclusive.

In the two remaining quadrants of the GPR-VTL plane (upper left and lower right), the majority of the vowels have combinations of GPR and VTL that are not typical of the normal population of voices; this is a consequence of the experimental design in which two related responses (size and sex/age) are obtained on each trial. It means that, as the GPR and VTL values become more extreme, the listener has to make a projection from normal experience, and decide which of the four categories of speaker would be most likely to produce the sound presented. As a result, the relationship between the sex/age category perceived and the combination of GPR and VTL in the vowel is not as straightforward as for the other two quadrants. Nevertheless, the use of the category names is consistent and seems entirely reasonable. Vowels with low GPRs but short VTLs, in the upper left-hand quadrant of the plane, are predominantly categorized as boys (upper left-hand panel of Fig. 4); the probability of "boy" is greater than 0.5 throughout most of this quadrant (although the ellipse for boys in the normal population does not even fall in this quadrant). In retrospect, the reason is not difficult to deduce; these voices with their short VTLs and low GPRs sound like males who, for some reason, are unusually short. This condition exists for male dwarves who are quite uncommon, but not unknown. Their pitch drops significantly at puberty but their vocal tract does not increase proportionately in length because their bodies do not grow in the usual way. There is no corresponding vocal category for female dwarves; they continue to sound like girls because they do not grow to the normal height and, although their pitch may decrease in the normal way, this decrease is relatively small, and the drop in pitch would rarely be so great as to shift the voice into the left-hand section of the GPR-VTL plane. The listeners rarely use the categories "woman" or "girl" for vowels in this quadrant; on the border where the pitch is 179 Hz, the response "boy" is far more likely than "woman" or "girl."

In the final quadrant of the GPR-VTL plane (lower right-hand), where the vowels have high GPRs in combination with *long* VTLs, the most common responses are "man," when the GRP is relatively low and the VTL is long (lower left-hand panel of Fig. 4), and "boy" when the GRP is relatively high and the VTL is long (upper left-hand panel of Fig. 4). The "man" responses are just the natural extension of the large "man" region in the lower, left-hand quadrant of the GPR-VTL plane. "Boy" responses seem reasonable for voices that have a high pitch and are perceived to come from tall people.

The response "woman" is predominant only in one small region near the center of the GPR-VTL plane as shown in the lower right-hand panel of Fig. 4. In this region, the probability of the response exceeds the criterion value of 0.5. Moreover, the peak of the region is close to the center of the ellipse for women (although the listeners had no knowledge of these distributions other than their personal experience). The response "woman" is also used for a proportion of the vowels produced with GPRs that are greater than those for normal women, *provided* the VTL is the same, or longer, than that for normal women; these responses appear in the lower, right-hand quadrant of the lower, right-hand panel. If the VTL becomes shorter, listeners consistently use "girl" instead of "woman," and if it becomes longer, they consistently use "boy" instead of "woman." Nevertheless, the relationship between the response category, "woman," and the combination of GPR and VTL in the vowel seems reasonable.

The four panels of Fig. 4 also make it clear that the distribution of responses across the four sex/age categories is not uniform; the overall probabilities for man, woman, boy, and girl, are 0.36, 0.11, 0.36, and 0.17, respectively. The relatively low probability of responding "woman" in the women's ellipse is consistent with the large degree of overlap of the woman's ellipse with the boy and girl ellipses. The nonuniform distribution of response is largely attributable to the fact that the GPR and VTL values span a rectangular plane of combinations, whereas the normal population of voices is concentrated on combinations that cluster the central section of the positive diagonal in the GPR-VTL plane. The listener has to extend the use of the normal categories to the novel stimuli and, in general, they do this reasonably and consistently, but it does lead to a nonuniform distribution of responses across the four categories.

All of the vowels in these experiments were synthesized from the vowels of one adult male speaker, and so the formant ratios for a given vowel are the same for all combinations of GPR and VTL in the experiment. This has the advantage of minimizing one source of variability in the experiment; however, it is not typical of the human population (Diehl, Lindbolm, Hoemeke, and Fahey, 1996). For a given vowel, the formant ratio $F2/F1$ increases slowly as height increases because throat length grows relatively more than mouth length as we grow up. The effect is illustrated in Turner, Walters, and Patterson (2004) using MRI measurements of mouth and throat length from Fitch and Giedd (1999). It is not entirely clear what effect this might have on judgements of size and sex/age like those in this paper. However, we might expect that, if the experiment were rerun using vowels from an adult female speaker, listeners would hear a difference and the proportion of "woman" responses would be higher over much of the range of our measurements. In the end, however, although the effects of the sex and size of the original speaker on these judgements are interesting, they are beyond the scope of this paper.

IV. DISCUSSION

The size rating experiments show that listeners make consistent judgements about speaker size given a sequence of vowel sounds (Fig. 3). Both GPR and VTL affect judgements of speaker size (Figs. 5 and 6), and the effect of VTL is strong enough to change speaker size estimates from tall to

short. The sex and age judgements are also affected both by the GPR and the VTL of the vowels (Fig. 4). The data show that sex and age are not dictated solely by GPR or VTL; rather, there is an interaction between these variables that means that specific combinations of GPR and VTL act as robust indicators of sex and age.

A. Speaker size - interaction of GPR and VTL

Previous studies on the perception of speaker size were limited by the restricted range of heights of the speakers. For instance, listeners were asked to judge the height of speaker of recordings made from adult men only (Lass and Davis, 1976; van Dommelen and Moxness, 1995; Collins, 2000). Although listeners made consistent judgements about the size of these adult speakers, these estimates were not very accurate (van Dommelen and Moxness, 1995; Collins, 2000), though Lass and Davis (1976) did report better than chance correct categorization. These studies only used normal-range adult voices which have recently been reported to show a significant but weak correlation between speaker size and formant frequency within same sex adults (González, 2004). Given the weak correlation between speaker size and formant frequency in same sex adults, the task of accurately judging the physical height of same sex adults might prove difficult (Rendall, Vokey, Nemeth, and Ney, 2005). Nevertheless, listeners make consistent perceptual decisions as if they were receiving strong valid acoustic cues to speaker size (e.g., Collins, 2000). One way to reconcile this apparent conflict is to hypothesize that the correlation between speaker height, VTL and formant frequency, observed in close hominoid species such as rhesus monkeys (Fitch, 1997), has become disassociated in adult humans, possibly because of human-specific vocal-tract changes such as the descent of the larynx in adult men (Fitch, 1997; Fitch, 2000).

A better mapping between perceptual judgements of speaker size and physical speaker size might arise if a wider range of speaker heights were used, say from very small children to very large men. However, for all natural recorded voices there will always be the problem that GPR and VTL cues are confounded. To tease out the separate effects of each of the cues, and to simultaneously provide listeners with a suitably wide range of potential heights, it is necessary to use synthetic speech. Fitch (1994, Ph.D. thesis) used a rating scale to gather listeners' judgements of speaker size using computerized vowels. Even though the vowels were restricted to the middle to upper normal range for men only, he found main effects of both GPR and VTL on listeners' size ratings. Our study uses a much greater range of GPR and VTL values, simulating tiny children, giants, castrati, and dwarves, as well as everyday speech combinations. It was made possible by the recent development of the high-quality vocoder, STRAIGHT (Kawahara *et al.*, 1999; Kawahara and Irino, 2004).

GPR can be used to distinguish between male and female speakers (Bachorowski and Owren, 1999) but not to draw reliable *intra*-sex inferences about speaker size. Unlike the vocal tract, which is related to the size of the cranium and hence body size, the vocal folds are not constrained by any hard bony structure (Negus, 1949; discussed in Fitch, 1997). Prosody is also routinely used to make sentence distinctions, e.g., "The baby is happy" with constant pitch is a statement, but the same sentence with rising pitch is a question. The advantage of vocal-tract information is clear. Measurements made with magnetic resonance imaging show that VTL is highly correlated with speaker height (Fitch and Giedd, 1999). There is a highly significant correlation between age and formant frequency in humans (Huber *et al.*, 1999), and a strong relationship between body size and formant-related parameters in rhesus monkeys (Fitch, 1997). The reliability of VTL for speaker size (as signalled by perceptually salient shifts in formant frequency) may have weakened within human adults of the same sex (González, 2004), but it is still strong between groups of children, women, and men. Within group, the correlation between GPR and speaker size is surprisingly weak, both in humans, and close hominoid species (Fitch, 1997; Rendall *et al.*, 2004).

The strong effect of VTL on the perception of speaker size may reflect the extremely wide range of VTL values used in our study. In normal speech, pitch is more salient than vocal tract length, perhaps because the just noticeable difference for voice pitch is about 2%, whereas the just noticeable difference for a change in VTL is 6–10% (Smith *et al.*, 2005). STRAIGHT enabled us to simulate the vowels of very small children and giants. This could have the effect of encouraging listeners to lend additional weight to VTL, especially if VTL has more natural relevance to speaker size than GPR.

Our earlier size discrimination experiments showed that listeners were capable of discriminating changes in speaker size of 6–10% when the sounds were presented in two temporal intervals of a forced choice experiment (Smith, Patterson, and Jefferis, 2003; Smith and Patterson, 2004a; Smith *et al.*, 2005). The size perception experiments reported in this paper show that listeners can also make consistent and sensible size judgements about vowels which are presented in a single temporal interval. The listener in this rating task cannot discriminate speaker size *relative* to another vowel sound presented immediately after the first vowel sound; rather, they have to make a judgement about speaker size relative to the frame of reference provided by all the other vowel sounds in the set (and presumably all the vowel sounds they have experienced over their lives). That our listeners can do this task as well as they do, supports our belief that size information can be extracted from individual voiced sounds to inform perceptual decisions.

B. Speaker sex and age: the interaction of GPR and VTL

Previous research attempting to identify those acoustic properties of male and female voices responsible for our perception of sex type, have used either statistical clustering methods (e.g., Childers and Wu, 1991; Wu and Childers, 1991; Bachorowski and Owren, 1999) or perceptual categorization experiments (e.g., Schwartz, 1968; Schwartz and Rine, 1968; Ingemann, 1968; Lass *et al.*., 1976). The statistical clustering studies have consistently highlighted GPR

and vocal tract related variables as explaining most of the variance between the speech sounds of adult males and females (Childers and Wu, 1991; Bachorowski and Owren, 1999). Some studies have shown that vocal tract information alone can be used to identify speaker sex (Schwartz, 1968; Ingemann, 1968; Schwartz and Rine, 1968). Other studies have reported that GPR is a much stronger cue to speaker sex than VTL (Lass *et al.*, 1976). Statistical clustering studies suggest that GPR and VTL are highly correlated (Childers and Wu, 1991; Wu and Childers, 1991). Other studies suggest that formant information can be important in discriminating speaker sex (Coleman, 1976; Whiteside, 1998) but generally pitch is dominant (Whiteside, 1998). Recently, Bachorowski and Owren (1999) have shown that sex classification accuracy is excellent using only GPR or only VTL, but best using both.

Our reasons for wishing to measure the interaction of GPR and VTL in sex/age judgements were based on two main factors. First, we believe that the auditory system employs a scale invariant neural transform to normalize natural sounds for size prior to more central processes like speaker identification (e.g., Irino and Patterson, 2002; Turner, Al-Hames, Smith, Kawahara, Irino, and Patterson, 2005). We have recently reported evidence that human listeners are able to discriminate and use size information in speech sounds (vowels), suggesting that size information is actively used in auditory perception (Smith, Patterson, and Jefferis, 2003; Smith and Patterson, 2004a; Smith *et al.*, 2005). We were thus interested in how speaker size information, as mediated by VTL and GPR cues, influenced decisions in natural sex/age categorization (man, woman, boy, or girl). Second, both statistical and perceptual classification studies are limited to databases of sounds that are from *normal* groups, i.e., recorded from largely homogeneous (usually adult) males and females. Thus the range over which the independent variables could be manipulated was necessarily limited. The vocoder STRAIGHT (Kawahara *et al.*, 1999; Kawahara and Irino, 2004) enabled us to manipulate the GPR and VTL of vowels independently of each other over a huge range. These speech sounds are of high quality even when pushed well beyond the normal range of speech. This allows unprecedented control over our main experimental variables, across a much wider range of GPRs and VTLs than has been used previously.

We found that *both* GPR and VTL contribute to listeners' perception of the sex and age of a speaker (Fig. 4). If GPR was the *sole* perceptual determinant of the sex and age of the speaker (man, woman, boy, or girl), then listeners would only be able to reliably classify most men (GPR $\lesssim$155 Hz) and the higher-pitched girls (GPR$\gtrsim$330 Hz). If VTL was the *only* perceptual marker to sex and age then listeners would only be able to reliably classify taller men (with VTL$\gtrsim$16 cm) and shorter girls (with VTL$\lesssim$10 cm). The sex classification performance of our listeners is much better than this.

V. SUMMARY AND CONCLUSIONS

Listeners were presented with vowels in a single-interval, two-response paradigm. The listener heard a vowel scaled in GPR and VTL, and had to make one judgement about the size of the speaker (on a seven-point ordinal scale ranging from "very short" to "very tall") and a second judgement about the sex/age of the speaker (man, woman, boy, or girl). The results from the speaker size judgement experiment show that VTL has a strong influence upon perceived speaker size (Figs. 3, 5, 6). The strength of this effect presumably reflects the high correlation of VTL with speaker size. The results of the sex/age categorization experiments show that judgements of speaker sex/age are influenced by the interaction of GPR and VTL (Fig. 4). In the normal range of GPR and VTL values, judgements of sex/age are consistent with listeners combining both GPR and VTL information about equally to give a robust indicator of sex and age. When listeners are presented with unusual GPR and VTL combinations, where low GPRs are combined with short VTLs, the VTL information appears to decide the sex/age judgement.

ACKNOWLEDGMENTS

This research was supported by the UK MRC (G9901257; G9900369) and the German Volkswagen Foundation (VWF 1/79 783). Some of the data were reported in abstract form (Smith and Patterson, 2004b; Smith and Patterson, 2005). We thank Richard Turner for providing the ellipses showing the GPR-VTL values for men, women, boys, and girls as derived from the data of Peterson and Barney (1952).

[1]The shape of the vocal tract is largely determined by the placement of the tongue within the oral cavity. The shape affects the positioning of the formants relative to each other—different vowels having different vector angles in a multi-dimensional vowel space. For the purposes of our argument, we assume the same fixed vocal tract shape across all speakers, i.e., the speakers are uttering the same vowel.

[2]http://www.mrc-cbu.cam.ac.uk/cnbh/web2002/framesets/Soundsframeset.htm. Click on "Scaled vowels."

[3]In British English, "quite" means "to some extent" in this context. Since the size scale was presented graphically and continuously, it is very likely that the responses involving the word "quite" were interpreted as meaning half way between the sizes above and below them in each case.

[4]The set of formant values for each of the 76 speakers in the classic study of Peterson and Barney (1952) were converted to VTL values using the VTL data that Fitch and Giedd (1999) extracted from magnetic resonance images of a large population of subjects. Each ellipse represents the mean ± three standard deviations for each category of speaker. The calibration details are presented in Turner, Walters, and Patterson (2004); this poster can be found at http://www.mrc-cbu.cam.ac.uk/cnbh/web2002/framesets/posters_talksframeset.htm.

[5]An estimate of the size of speaker for a given SER was derived by extrapolating from the VTL *versus* height data in Fitch and Giedd [1999 cf. Fig. 2(a)]. In Fitch and Giedd, the average VTL for seven men aged 19 to 25 was 15.54 cm. An SER of 0.58 means that the spectral envelope of the initial vowel has been compressed by a factor of 1.72 (=1/0.58), while an SER of 2.39 means that the spectral envelope has been dilated by 0.42. Assuming linear scaling between VTL and formant frequency, these SER values are equivalent to VTLs of 26.8 and 6.5 cm, respectively. The former would be a giant and the latter a tiny child.

[6]The two 7×7 ranges (cf. Fig. 1) were merged to form one 13×13 matrix (the middle row and column of both ranges is the same). Any empty cell in the matrix was filled by the average of all adjoining cells where a speaker size rating had been collected. The data surface was derived by interpolation between the sample points and their averaged neighbors.

Bachorowski, J., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker sex identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.
Beckford, N. S., Rood, S. R., and Schaid, D. (**1985**). "Androgen stimulation and laryngeal development," Ann. Otol. Rhinol. Laryngol. **94**, 634–640.
Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.
Coleman, R. O. (**1976**). "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice," J. Speech Hear. Res. **19**, 168–180.
Collins, S. A. (**2000**). "Men's voices and women's choices," Anim. Behav. **60**, 773–780.
Darwin, C. (**1871**). "The Descent of Man and Selection in Relation to Sex" (Murray, London).
Diehl, R. L., Lindbolm, B., Hoemeke, K. A., and Fahey, R. P. (**1996**). "On explaining certain male-female differences in the phonetic realization of vowel categories," J. Phonetics **24**, 187–208.
Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Fant, G. (**1970**). "Acoustic Theory of Speech Production", 2nd ed. (Mouton, Paris).
Fairchild, L. (**1981**). "Mate selection and behavioural thermoregulation in Fowler's toads," Science **212**, 950–951.
Fitch, W. T. (**1994**). "Vocal tract length perception and the evolution of language," Ph.D. dissertation, Brown University.
Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus monkeys," J. Acoust. Soc. Am. **102**, 1213–1222.
Fitch, W. T. (**1999**). "Acoustic exaggeration of size in birds by tracheal elongation: Comparative and theoretical analyses," J. Zool. (London) **248**, 31–49.
Fitch, W. T. (**2000**). "The evolution of speech: a comparative review," Trends in Cognitive Science **4**, 258–267.
Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.
González, J. (**2004**). "Formant frequencies and body size of speaker: a weak relationship in adult humans," J. Phonetics **32**, 277–287.
Hast, M. (**1989**). "The larynx of roaring and non-roaring cats," J. Anat. **163**, 117–121.
Hollien, H., Green, R., and Massey, K. (**1994**). "Longitudinal research on adolescent voice change in males," J. Acoust. Soc. Am. **96**, 3099–3111.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T., and Johnson, K. (**1999**). "Formants of children, women and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.
Ingemann, F. (**1968**). "Identification of the speaker's sex from voiceless fricatives," J. Acoust. Soc. Am. **44**, 1142–1144.
Irino, T., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform," Speech Commun. **36**, 181–203.
Kawahara, H., and Irino, T. (**2004**). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. Divenyi, Kluwer Academic, Massachusetts, 167–180.
Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds," Speech Commun. **27**, 187–207.
Künzel, H. J. (**1989**). "How well does average fundamental frequency correlate with speaker height and weight?" Phonetica **46**, 117–125.
Lass, N. J., and Brown, W. S. (**1978**). "Correlational study of speakers' heights, weights, body surface areas and speaking fundamental frequencies," J. Acoust. Soc. Am. **63**, 1218–1220.
Lass, N. J., and Davis, M. (**1976**). "An investigation of speaker height and weight identification," J. Acoust. Soc. Am. **60**, 700–703.
Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (**1976**). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," J. Acoust. Soc. Am. **59**, 675–678.
Liu, C., and Kewley-Port, D. (**2004**). "STRAIGHT: a new speech synthesizer for vowel formant discrimination," ARLO **5**, 31–36.
Morton, E. S. (**1977**). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," Am. Nat. **111**, 855–869.
Narins, P. M., and Smith, S. L. (**1986**). "Clinal variation in anuran advertisement calls—basis for acoustic isolation," Behav. Ecol. Sociobiol. **19**, 135–141.
Negus, V. E. (**1949**). *The Comparative Anatomy and Physiology of the Larynx* (Hafner, New York).
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. D. (**2004**). "Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners," J. Acoust. Soc. Am. **115**, 411–421.
Rendall, D., Vokey, J. R., Nemeth, C., and Ney, C. (**2005**). "Reliable but weak voice-formant cues to body size in men but not women," J. Acoust. Soc. Am. **117**, 2372.
Riede, T., and Fitch, W. T. (**1999**). "Vocal tract length and acoustics of vocalization in the domestic dog Canis familiaris," J. Exp. Biol. **202**, 2859–2867.
Schwartz, M. F. (**1968**). "Identification of speaker sex from isolated, voiceless fricatives," J. Acoust. Soc. Am. **43**, 1178–1179.
Schwartz, M. F., and Rine, H. E. (**1968**). "Identification of speaker sex from isolated, whispered vowels," J. Acoust. Soc. Am. **44**, 1736–1737.
Smith, D. R. R., and Patterson, R. D. (**2004a**). "The existence region of scaled vowels in pitch-VTL space," *18th International Conference on Acoustics*, Kyoto, Japan, Vol. I, 453–456.
Smith, D. R. R., and Patterson, R. D. (**2004b**). "The perception of sex and size in vowel sounds," paper presented at British Society of Audiology, UCL London P49..
Smith, D. R. R., and Patterson, R. D. (**2005**). "Perception of speaker size and sex of vowel sounds," J. Acoust. Soc. Am. **117**, 2374.
Smith, D. R. R., Patterson, R. D., and Jefferis, J. (**2003**). "The perception of scale in vowel sounds," paper presented at British Society of Audiology, Nottingham P35.
Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.
Turner, R. E., Al-Hames, M. A., Smith, D. R. R., Kawahara, H., Irino, T., and Patterson, R. D. (**2005**). "Vowel normalisation: Time-domain processing of the internal dynamics of speech," in *Dynamics of Speech Production and Perception*, edited by P. Divenyi (IOS Press) (in press).
Turner, R. E., and Patterson, R. D. (**2003**). "An analysis of the size information in classical formant data: Peterson and Barney (1952) revisited," J. Acoust. Soc. Am. **33**, 585–589.
Turner, R. E., Walters, T. C., and Patterson, R. D. (**2004**). "Estimating vocal tract length from formant frequency data using a physical model and a latent variable factor analysis," paper presented at British Society of Audiology, UCL London P61.
van Dommelen, W. A., and Moxness, B. H. (**1995**). "Acoustic parameters in speaker height and weight identification: sex-specific behaviour," Lang Speech **38**, 267–287.
Whiteside, S. P. (**1998**). "Identification of a speaker's sex from synthesized vowels," Percept. Mot. Skills **86**, 595–600.
Wu, K., and Childers, D. G. (**1991**). "Gender recognition from speech. Part I: Coarse analysis," J. Acoust. Soc. Am. **90**, 1828–1840.