19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

Judging sex and age: effect of glottal-pulse rate, vocal-tract length and original talker

PACS: 43.71.Bp

Smith, David R. R.[1,2]; Walters, Thomas C.[2]; Patterson, Roy D.[2]
[1]Dept. Psychology, University of Hull, Cottingham Rd., Hull, HU6 7RX, United Kingdom; d.r.smith@hull.ac.uk
[2]Centre for Neural Basis of Hearing, Dept. Physiology Development & Neuroscience, University of Cambridge, Downing St., Cambridge, CB2 3EG, United Kingdom; rdp1@cam.ac.uk, tcw24@cam.ac.uk

ABSTRACT
Glottal-pulse rate (GPR) and vocal-tract length (VTL) strongly influence the perceived sex and age of the speaker [Smith and Patterson, J. Acoust. Soc. Am. 118, 3177-3186 (2005)] [1]. Our previous research simulated the voices of variously-sized speakers of both sexes, by manipulating the recorded vowels of one adult male talker. The current study explored whether there are additional cues in the voices of men, women and children that influence judgements of speaker sex and age. We manipulated the recorded vowels of an adult man, adult woman, young boy and young girl, and determined the effect upon the distribution of sex and age responses ('man', 'woman', 'boy', 'girl'). Results show that the distribution of sex and age judgements across the GPR-VTL plane is heavily influenced by GPR and VTL, but it is also affected by the original talker's size (or age).

INTRODUCTION
The voices of men, women and children sound different from each other. Much of the difference between their voices is a consequence of speaker size-driven and gender-specific maturational processes. Key perceptual differences are caused by differences in the length and mass of the vocal folds [2], and the length of the vocal tract [3,4]. Recently, we have shown that glottal-pulse rate (GPR) and vocal-tract length (VTL) are important determinants of whether vowels are heard as being spoken by men, women, boys, or girls [1]. Our previous work simulated the voices of variously-sized speakers of both sexes (different GPR and VTL combinations), by manipulating the recorded speech of a single adult male speaker. Thus a range of voices was created from a single starting point in the GPR-VTL plane.

In this paper we explore how having four different original speakers (adult man and woman, young boy and girl), i.e., creating the same range of voices but starting from four different starting points in the GPR-VTL plane, affects the distribution of sex and age judgements ('man', 'woman', 'boy', 'girl') in the GPR-VTL plane. The question is 'Do additional cues in the voices of men, women and children alter judgements of sex and age significantly?'

METHOD
Listeners were presented isolated vowels recorded from four different speakers (adult man and woman, young boy and girl). The vowels were scaled over a range of GPR and VTL values. Listeners were required to judge whether a boy, girl, man or woman had spoken each scaled vowel.

A. Stimuli
Examples of the five English vowels (/a/, /e/, /i/, /o/, /u/) of an adult man and woman, and a young boy and girl, were recorded using a high-quality microphone (Shure SM58-LCE), with a sampling rate of 48 kHz and a 16-bit amplitude resolution. Speakers were required to utter a series of sustained vowels at a regular relaxed rate at a comfortable effort level. For each speaker, five examples of each of the five vowels were selected. The vowel's natural onset and

offset were retained. The age, weight, GPR, height and estimated VTL for the four different speakers are shown in Table I.

Table I. Physical variables for the four speakers.

SPEAKER	AGE [yr]	WEIGHT [kg]	GPR [Hz]#	HEIGHT [cm]	VTL [cm] #	HEIGHT [%]*	VTL [%]*
MAN	24	69.6	108	183	17.6	100	100
WOMAN	41	68	226	175	14.9	96	85
GIRL	9	36	239	143	13.2	78	75
BOY	6	22	256	121	12.5	66	71

#average across all vowels *expressed as a percentage normalized to the value for the adult male speaker.

The vowel sounds were normalized to the same rms level prior to scaling. Pilot listening indicated that the vowel sounds had similar loudness.

To scale the vowels of the four speakers to the same VTL, it is necessary to estimate the VTL of each speaker. This was done by analyzing the five recorded examples of each of the five vowels /a/ to /u/, for each speaker. The frequencies of formants *F*1 to *F*3 of each vowel were estimated using Praat [5], and largely agreed with those reported in Hillenbrand [6]. These formant frequencies were fed into a physical model of formant production [7]. The estimates were calibrated against MRI estimates of vocal tract length [4] and a vowel database [6]. This model performs a factor analysis with a single latent factor of speaker size [7]. Figure 1 shows estimates of VTL for each speaker, for each of the five vowels, using this model. The scaling factor for each speaker's estimated VTL was based on the average across all vowels for that speaker.

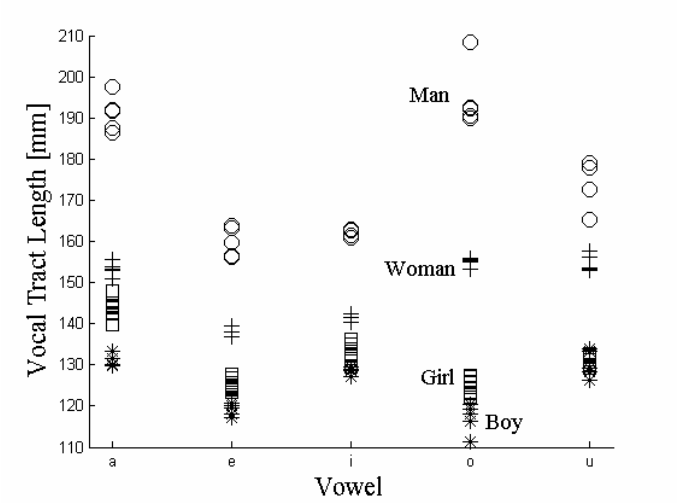

Figure 1. Estimates of vocal-tract length from formant frequency data using a physical model and a latent variable factor analysis [7]. Details for each of the speakers are provided in Table I.

The scaling of the vowels was performed by STRAIGHT [8]. See [9] for a description of how STRAIGHT is used to scale vowels to simulate different speaker sizes, and [10] for the underlying principles. The duration of all vowels was adjusted to 850 ms within STRAIGHT, by stretching/expanding the signal without altering the pitch or spectral content. The GPR and VTL values to which the vowels were scaled are shown in Fig. 2. The values were chosen to encompass the range of GPR and VTL values encountered in the normal population; the GPR varied from 80 to 440 in six, equal logarithmic steps, and the VTL ranged from 18.5 cm to 8.8 cm in six, equal logarithmic steps. 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 vowel database [11]. Each ellipse encompasses 99% of the individuals in the Peterson and Barney data for that category of speaker. Six points in the top-left corner (low GPR and short VTL combinations) and six points in the bottom-right corner (high GPR and long VTL

combinations) were not included because these combinations are particularly unusual and we wished to focus the listeners' attention on normal perception as far as possible.

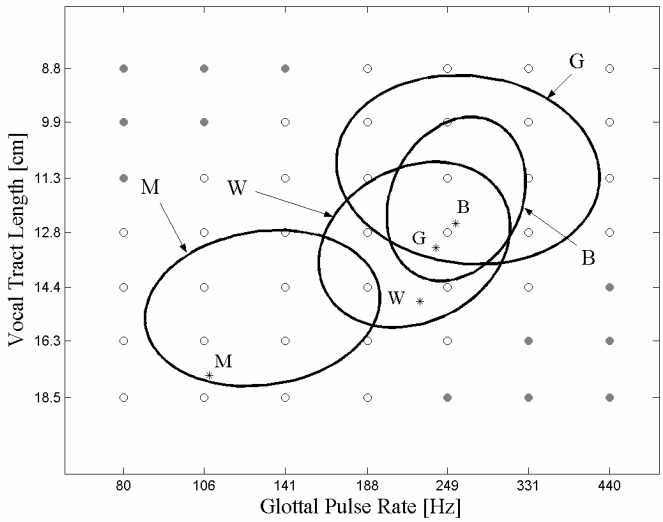

Figure 2. GPR and VTL values of scaled vowels (open circle symbols). Twelve GPR-VTL combinations were not presented (filled grey circles).

Listeners were seated in a double-walled, IAC, sound-attenuating booth. The stimuli were played by a 24-bit sound card through a TDT anti-aliasing filter with a sharp cutoff at 10 kHz, and presented diotically to the listener over AKG K240DF headphones. The sound level of the vowels at the headphones was ~60 dB SPL.

B. Procedures

The experiments were performed using a single-interval, single-response paradigm. The listener heard scaled versions of five stationary English vowels (/a/, /e/, /i/, /o/, /u/), and had to make a judgement about the sex/age of the speaker (man, woman, boy, girl). Sex/age judgements were made by selecting the appropriate button on a response box. Vowel level was roved in intensity over a 10 (±5) dB range. There was no feedback.

A run of judgements consisted of one presentation of each GPR-VTL combination for all five vowels and all four input speakers, presented in a pseudo-random order (a total of 37 GPR-VTL combinations X 5 vowels X 4 input speakers, or 740 trials). For each trial, there were five possible examples of the single vowel that could be played. Each run took about 50 min to complete. Each listener completed five runs in three sessions over a week. Ten listeners participated in the experiments, five male and five female. They ranged in age from 20 to 53 years, and were paid volunteers. All had normal absolute thresholds at 0.5, 1, 2, 4, and 8 kHz.

RESULTS

The results averaged across all listeners are presented in Fig. 3, where each group of four panels shows the probability of each of the four different speakers being heard as a 'boy', 'girl', 'woman', or 'man'. Each of the sixteen panels shows the probability of classifying a vowel sound as the given speaker type, as a function of GPR and VTL. The probability of classification is shown by colour, ranging from 0 (dark-blue) meaning "never classified" to 1 (brown-red) meaning "always classified". Within each panel 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 combinations in the top-left and bottom-right corners of the GPR-VTL plane were not presented, and therefore, are omitted from the interpolated surface. The dotted black lines outline regions of the GPR-VTL plane where listeners consistently chose 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 [11].

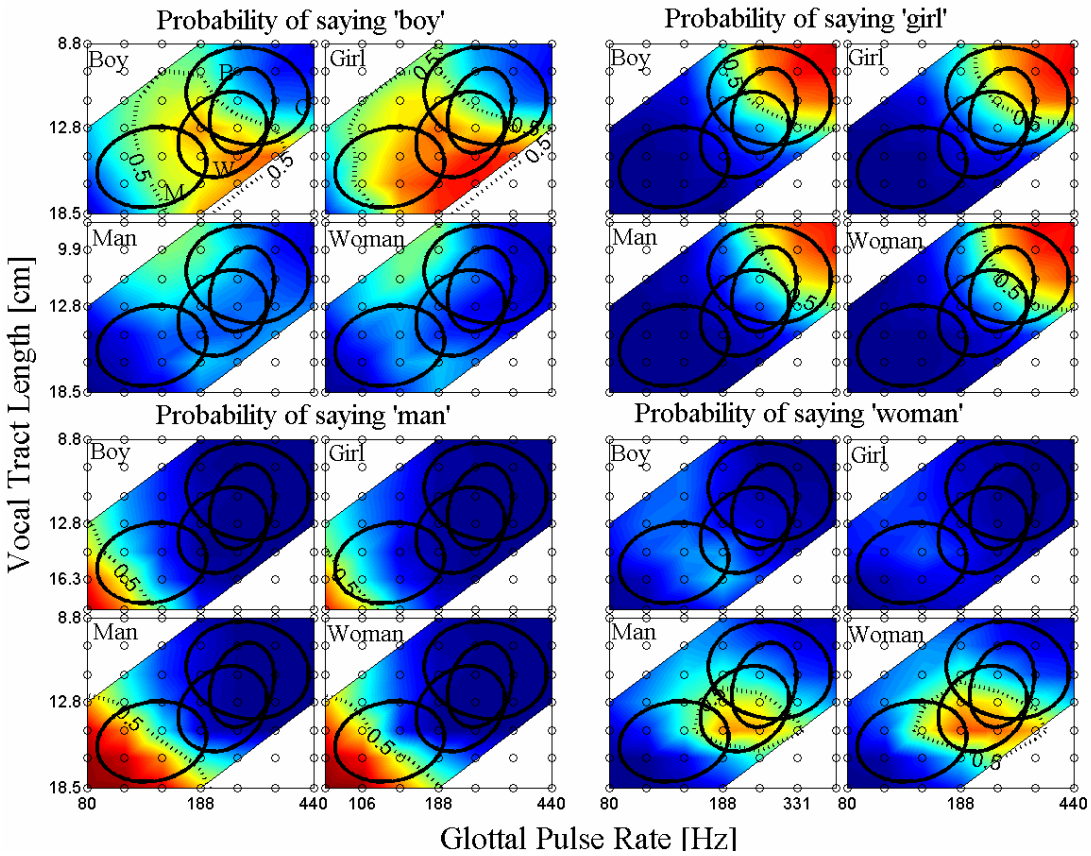

Figure 3. Sex and age judgements for the four different speakers (young boy, young girl, adult man, and adult woman) averaged over the five vowels and over all listeners (*n*=10). The probability at each sample point (open circle symbols) is based on 250 trials.

Role of GPR, VTL and original speaker in sex and age judgements

The distribution of sex/age judgements across the GPR-VTL plane is relatively unaffected by the sex and age of the original speaker of the vowels for certain categories of response. For instance, the distribution of 'man' responses in the bottom-left quadrant of Fig. 3 is very similar for the four speakers, in that vowels with a combination of low GPR and long VTL tend to be categorized as being spoken by a man. The ellipse for adult men shows that this is the natural category to adopt for vowels scaled to these combinations of GPR and VTL. This result replicates Smith and Patterson [1], who also found that vowels in this region of the GPR-VTL plane are reported as being spoken by men. In Smith and Patterson [1], the vowels were scaled from a single adult male speaker. In the present paper, we can see that vowels with combinations of very low GPRs and very long VTLs are categorized as being spoken by adult men, regardless of whether they are scaled from vowels of an adult man or woman, or from vowels of a young boy or girl. Generally, vowels with combinations of low GPRs and long VTLs are categorized as 'men' though those scaled from the boy or girl original speaker have to have more extreme values compared to vowels scaled from the adult original speakers.

The perceptual maps associated with different speakers are very similar for the 'girl' response (top-right quadrant group of Fig. 3). Vowels with combinations of high GPRs and short VTLs, which appear in the upper-right corner of each of the GPR-VTL planes, are overwhelmingly categorized as being spoken by girls. This corner 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 natural category to adopt. This replicates Smith and Patterson [1] who also found that vowels in this region are categorized as being spoken by a girl. There is little effect of original speaker upon the distribution of girl responses.

The perceptual maps are very similar for the adult male and adult female original speakers (bottom row of each perceptual response group of Fig. 3); and the perceptual maps are also very similar for the young male and female original speakers (top row of each perceptual response group of Fig. 3).

However, there is a strong effect of original speaker when response is 'boy' or 'woman' (top-left and bottom-right quadrant groups in Fig. 3); the distributions in the upper row are different from those in the lower row for both of these response groups.

DISCUSSION

Broadly speaking, for the adult speakers, the distribution of sex and age judgements across the GPR-VTL plane is largely unaffected by the sex of the speaker; compare the plots for the man and the woman in the bottom row of each judgement group (Fig. 3). Similarly, for the children, the distribution of sex and age judgements is largely unaffected by the sex of the speaker; compare the plots for the young boy and girl in the top row of each judgement group (Fig. 3). The distribution of sex and age judgements is largely unaffected by the sex or age of the speaker for the 'man' and the 'girl' response groups (bottom-left and top-right judgement groups in Fig. 3). However, the pattern of judgements for the juvenile speakers is quite different from that for the adult speakers for the 'boy' and 'woman' response groups; compare the upper row with the lower row in the top-left and bottom-right groups of four panels (Fig. 3).

Sex and age judgements

Motivated by recent work on auditory size perception [9,12], we measured the interaction of GPR and VTL in sex and age judgements [1]. We found that both GPR and VTL contribute to listeners' perception of the sex and age of the speaker. However, this work simulated the voices of variously-sized speakers of different sexes (different GPR and VTL combinations), by manipulating the recorded speech of a single adult male speaker. Scaling all the vowels from one speaker meant that the pattern of formant ratios for a given vowel was fixed regardless of VTL. This is not typical of the human population [13]. Children have disproportionately large heads relative to their bodies, when compared with adults. As a consequence, there is a large difference in the oral/pharyngeal length ratio between children and adults [14]; this is clearly shown in Fig. 4. The details of the relationship between formant ratio and the oral/pharyngeal length is complicated, and beyond the scope of this paper. For the purposes of this paper, it is sufficient to note that cavity length determines the resonant frequency, or formant, produced by that cavity [3]. As a result, it is likely that there are different patterns of formant ratios for the vowels of children and adults, and it is likely that this plays a role in whether a speaker is heard as an adult or a child.

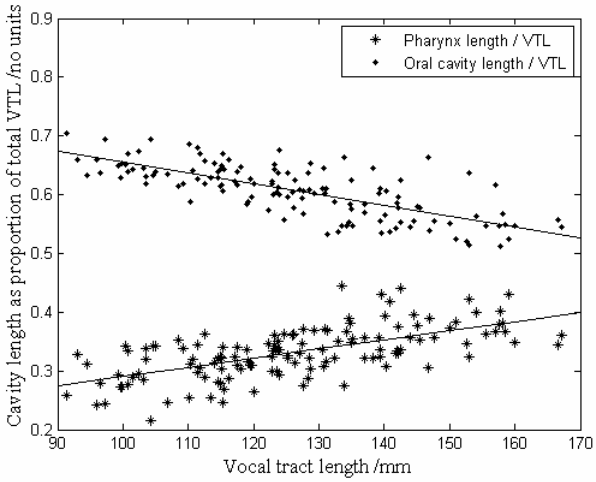

Figure 2. Oral and pharyngeal cavity lengths, expressed as a proportion of total VTL, and plotted as a function of speaker VTL (re-drawn from [7]). Cavity lengths derived from MRI measurements [4]. Lines show the best-fitting linear regression.

We explored whether these additional cues in the voices of men, women and children, significantly changed the distribution of sex and age classifications across the GPR-VTL plane.

We found that the perceptual distributions are very similar for adult speakers (Fig. 3, bottom row in all groups), and for juvenile speakers (Fig. 3, top row in all groups) considered separately. Thus sex of speaker has little effect. The greatest differences between distributions are for juvenile *versus* adult speakers for the 'boy' and 'woman' responses (top-left and bottom-right quadrant groups Fig. 3 respectively). Thus, size of speaker has a strong effect.

Finally, it is hard to make children sound like a woman (due presumably to formant ratio differences), yet we can make children sound like a man (bottom-left quadrant group Fig. 3). This is probably because the extreme GPR-VTL combinations in the bottom-left corner of the GPR-VTL plane override the more subtle formant ratio cues. Note, however, that the vowels of the young boy and girl have to be driven to more extreme values than those of the adult speakers before listeners report hearing a man speaking. A similar argument could be advanced for why high GPR and short VTL combinations are heard as girls, regardless of the original speaker (top-right quadrant group of Fig. 3).

SUMMARY AND CONCLUSIONS

Listeners were presented with vowels recorded from four different speakers (young boy and girl, adult man and woman) scaled to the same GPR and VTL values. Listeners were required to judge whether a boy, girl, man, or woman had spoken the scaled vowel. The results show that the distribution of responses across the GPR-VTL plane is largely independent of the sex and age of the original speaker for adult speakers (bottom row in each group Fig. 3) and, separately, for the juvenile speakers (top row in each group Fig. 3). Where differences do exist, as in the perception of boys (top-left quadrant group Fig. 3), and women (bottom-right quadrant group Fig. 3), the differences could be attributable to differences in the oral/pharyngeal length ratio between juveniles and adults. While GPR and VTL are major determinants of sex and age judgements, there is at least one more distinguishing characteristic in the voices, and it is associated with the overall size of the speaker (adult or juvenile) rather than their sex (male or female).

Acknowledgements

Research supported by UK MRC (G9900369, G0500221) and German Volkswagen Foundation (1/79 783)

References

[1] D.R.R. Smith, R.D. Patterson: The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. Journal of the Acoustical Society of America **118** (2005) 3177-3186

[2] I. R. Titze: Physiologic and acoustic differences between male and female voices. Journal of the Acoustical Society of America **85** (1989) 1699-1707

[3] G. Fant: *Acoustic Theory of Speech Production* 2nd ed. (1970) Mouton, Paris

[4] W. T. Fitch, J. Giedd: Morphology and development of the human vocal tract: A study using magnetic resonance imaging. Journal of the Acoustical Society of America **106** (1999) 1511-1522

[5] P. Boersma: Praat, a system for doing phonetics by computer. Glot International **5:9/10** (2001) 341-345

[6] J. M. Hillenbrand, L. A. Getty, M. J. Clark, K. Wheeler: Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America **97** (1995) 3099-3111

[7] R. E. Turner, T. C. Walters, R. D. Patterson: Estimating vocal tract length from formant frequency data using a physical model and a latent variable factor analysis. British Society of Audiology, UCL London **P61** (2004). http://www.pdn.cam.ac.uk/groups/cnbh/research/posters_talks/BSA2004/TWPBSA04.pdf.

[8] H. Kawahara, I. Masuda-Kasuse, A. de Cheveigne: Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds. Speech Communication **27**(3-4) (1999) 187-207

[9] D. R. R. Smith, R. D. Patterson, R. Turner, H. Kawahara, T. Irino: The processing and perception of size information in speech sounds. Journal of the Acoustical Society of America **117** (2005) 305-318

[10] H. Kawahara, T. Irino: Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech separation by humans and machines*, P. Divenyi (Ed.), Kluer Academic, Massachusetts, (2004) 167-180

[11] G. E. Peterson, H. L. Barney: Control methods used in a study of the vowels. Journal of the Acoustical Society of America **24** (1952) 175-184

[12] D. T. Ives, D. R. R. Smith, R. D. Patterson: Discrimination of speaker size from syllable phrases. Journal of the Acoustical Society of America **118** (2005) 3816-3822

[13] R. L. Diehl, B. Lindbolm, K. A. Hoemeke, R. P. Fahey: On explaining certain male-female differences in the phonetic realization of vowel categories. Journal of Phonetics **24** (1996) 187-208.

[14] G. Fant: A note on vocal tract size factors and non-uniform F-pattern scalings. STL-QPSR **4** (1966) 22-30.