UNIVERSITY OF CAMBRIDGE

CNBH, Dept Physiology, Development & Neuroscience, Cambridge University
www.pdn.cam.ac.uk/cnbh

Role of input speaker upon judgements of sex and age P39

David R. R. Smith, Thomas C. Walters, Katie Walland and Roy D. Patterson drrs2@cam.ac.uk

1 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.

Recently, we have shown how glottal-pulse rate (GPR, pitch) and vocal-tract length (VTL) interact to determine whether we hear vowels as being spoken by a man, woman, boy or girl (Smith and Patterson, 2005). 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.

The purpose of this work was to explore how having *different* input speakers (adult man and woman, young boy and girl), affects the distribution of sex and age classifications (boy, girl, man, woman) in the GPR-VTL plane.

2 APPROACH

Vowels were scaled to simulate variously-sized people of both sexes, as defined by a range of GPR and VTL values (Fig.1 panel 4). The recorded vowels of four different speakers were used (Table 1 panel 7). Vowels were scaled using the vocoder STRAIGHT (Kawahara *et al.*, 1999). Panel 7 gives more details. Listeners were presented with isolated vowels in a single-interval, one-response paradigm.

Listeners were required to judge whether a boy, girl, man or woman had spoken each scaled vowel.

3 RESULTS & CONCLUSIONS

- The distribution of responses (boy, girl, man, woman) across the GPR-VTL plane for the four different speakers are shown as 2D surface plots (Fig.2 panel 5).
- Listeners' judgements (boy, girl, man, woman) are largely unaffected by the sex of the original speaker from whom the vowels have been scaled. Judgements are primarily driven by the GPR and VTL values of the vowel, with an important caveat.
- Perceptual distributions are similar between adult speakers (Fig.2 bottom rows all groups).
- Perceptual distributions are similar between young speakers (Fig.2 top rows all groups).
- Where differences do exist, as in the perception of boys (Fig.2 top-left) and women (Fig.2 bottom-right), the differences are likely to be attributable to child-adult oral-pharyngeal length differences.
- Speaker size has greater power than speaker sex in accounting for the variance in our data.

4

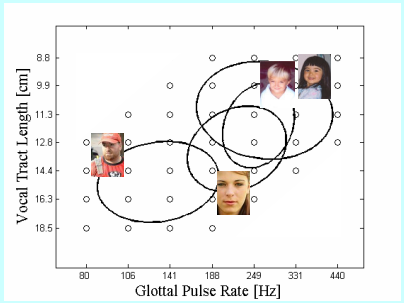


FIGURE 1. The open circles are GPR and VTL combinations to which vowels were scaled in the experiment. The four ellipses show estimates of the normal range of GPR and VTL values in speech for men, women, boys and girls (derived from Peterson and Barney, 1952). In each case, the ellipse encompasses 99% of individuals in the Peterson and Barney data for that category of speaker. The abscissa is GPR and the ordinate is VTL, plotted on logarithmic axes.

5

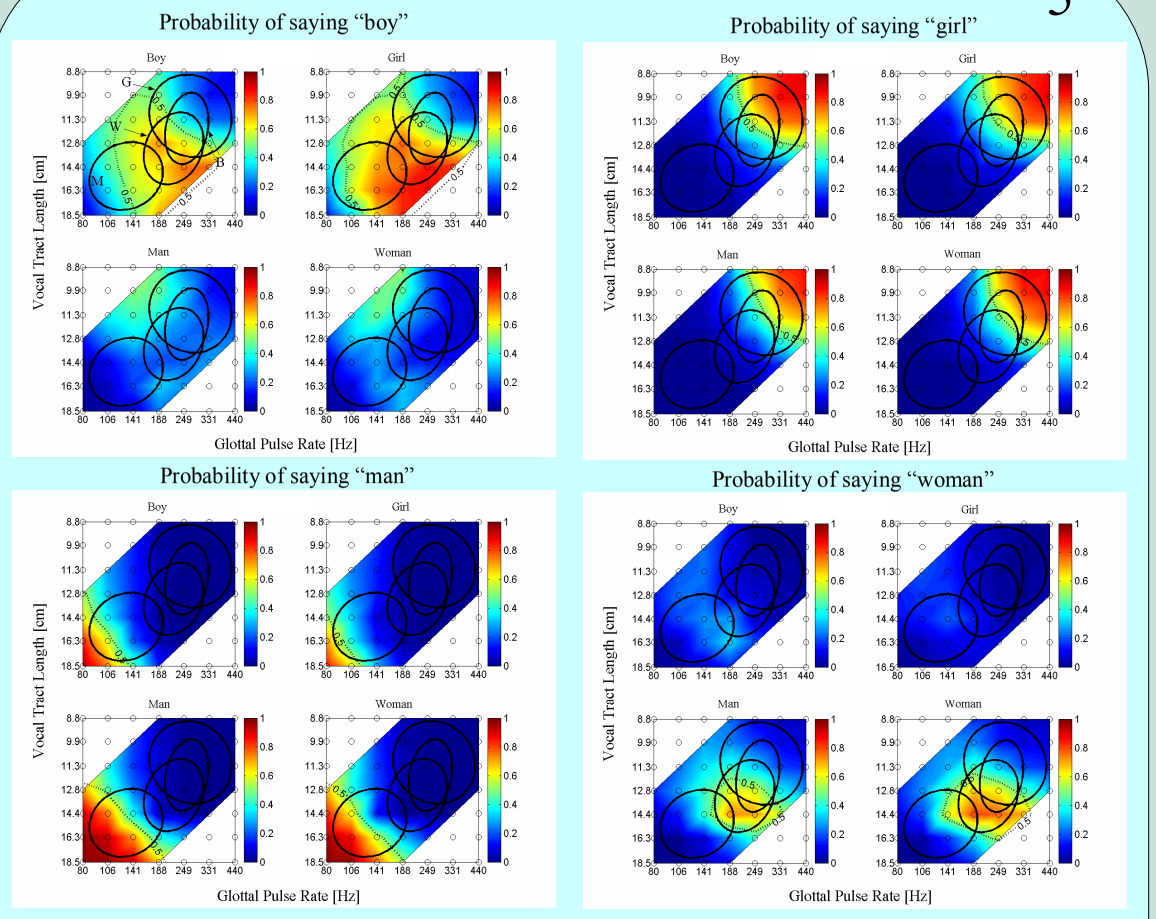
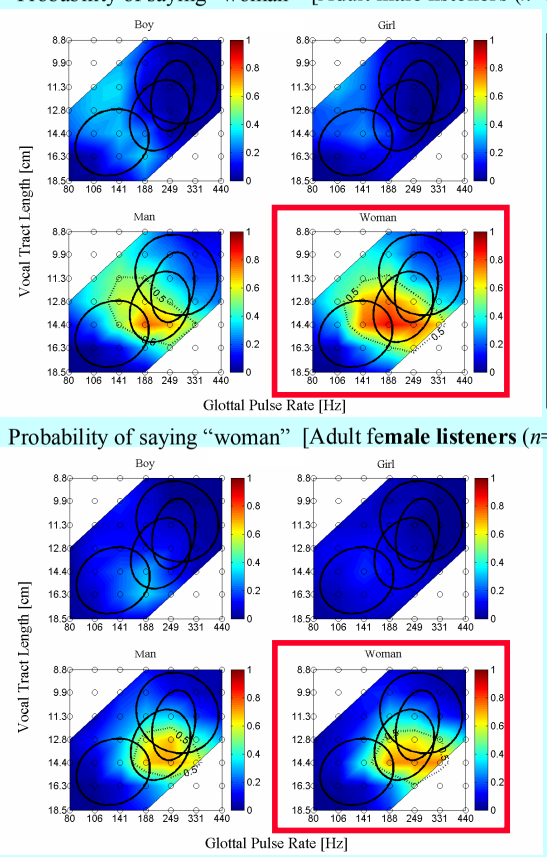


FIGURE 2. Sex/age judgements for the four different speakers (boy, girl, man, and woman) for all listeners (n=10). Data presented as 2D surfaces with colour showing probability of assigning a given GPR-VTL combination to one of four perceptual categories (boy, girl, man, or woman). Data for each perceptual response shown separately as a group of four panels, where each panel is a different speaker. Sample points (based on 250 trials) are shown as circles with interpolation between data points. Dotted black contour lines marks a probability $p \geq 0.5$ of consistently choosing one category out of the four available.

6



GENDER OF LISTENERS DIFFERENCES

Adult male listeners are more likely to rate a vowel as being spoken by a woman than are adult woman listeners (p=0.05, Mann-Whitney test, Fig.3). This is the case only if the vowels were scaled from a woman speaker and had the appropriate GPR and VTL combinations for a woman. This condition is ringed in red.

FIGURE 3. Probability of reporting "woman" for the four different speakers (boy, girl, man, and woman), across the GPR-VTL plane, for adult male and adult female listeners separately. All other details the same as described in Fig. 2.

7

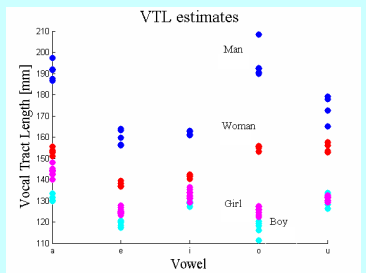
CANONICAL VOWELS We recorded four different speakers (adult man and woman, young boy and girl) saying the vowels /a/, /e/, /i/, /o/ and /u/. For details of the speakers cf. Table I. Sounds were digitized with 16-bit quantification and a sampling rate of 48 kHz. All vowels were balanced for loudness.

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 adult male speaker

TABLE 1. Physical variables for the four speakers

SCALE MANIPULATION Vowels were manipulated to have a range of GPR and simulated VTL values using STRAIGHT (Kawahara et al., 1999). STRAIGHT produces a pitch-independent spectral envelope that accurately tracks the motion of the vocal tract through an utterance. Once STRAIGHT has segregated a vowel into its 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, and the operations are largely independent.



EQUIVALENT VTL ACROSS SPEAKERS For all vowels to be scaled to the same (equivalent) VTL value for different-sized speakers, it is necessary to take into consideration the real VTL of each speaker. The VTL of each speaker was estimated by analyzing several recorded examples of the vowels, as spoken by each speaker. Formant tracks were extracted and fed into a model of formant production (Turner *et al*., 2004), which provided estimates of VTL for each speaker.

ACKNOWLEDGEMENTS

Research supported by the MRC (G9900369 and G0500221) and the German Volkswagen Foundation (VWF 1/79 783).

REFERENCES

Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (1999). *Speech Communication* **27**, 187-207.

Peterson, G.E., and Barney, H.L. (1952). *J. Acoust. Soc. Am.* **24**, 175-184.

Smith, D.R.R. and Patterson, R.D. (2005). *J. Acoust. Soc. Am.* **118**, 3177-3186.

Turner, R.E., Walters, T.C., and Patterson, R.D. (2004). *British Society Audiology,* UCL London **P61**.