CNBH, Physiology Department, Cambridge University
www.mrc-cbu.cam.ac.uk/cnbh

The perception of size and sex in vowel sounds P60

David R. R. Smith and Roy D. Patterson
david.smith@mrc-cbu.cam.ac.uk, roy.patterson@mrc-cbu.cam.ac.uk

1 INTRODUCTION

The voices of men, women and children sound different – men tend to have low frequency voices while children have high frequency voices, and women lie somewhere in the middle.

Our perception of these size differences (adult or child) and sex differences (male or female) is more complicated than it at first seems. There are two physical parameters that both vary with size and sex which are combined in our perception. One is the rate of opening and closing of the vocal folds (glottal pulse rate, GPR) determining voice pitch; the other is the length of the supra-laryngeal vocal tract (VTL) determining the centre of gravity of the frequencies.

Both GPR and VTL are linked to the physical size and sex of the speaker (Fig. 1 panel 4) – however, it is unclear how they interact to determine our perception of speaker size and sex.

The purpose of this work is to measure the relative contribution of GPR and VTL to judgements of:

- speaker size
- speaker sex and age

The work also extends our previous work on 2AFC size discrimination (Smith *et al*., 2003; Smith and Patterson, 2004) by using a rating procedure to measure speaker size.

2 APPROACH

Vowels were scaled to simulate people with a huge range of GPRs and VTLs, including many well beyond the normal range of the population (Fig. 1 panel 4). Vowels were scaled using the vocoder STRAIGHT (Kawahara *et al*., 1999). Panel 8 gives more details. Listeners were presented with vowels in a single-interval, two-response rating paradigm.

Listeners had to make one judgement about:

- speaker **size** (on a 7 point ordinal scale ranging from "very short" to "very tall")

and a second judgement about:

- speaker **sex** (man, woman, boy, girl)

3 RESULTS & CONCLUSIONS

SPEAKER SIZE

- Listeners are able to estimate the size of speaker of vowel sounds (Fig. 2 panel 5).
- The influence of VTL (SER) upon size estimates is up to six times greater than GPR (Fig. 3 panel 6).

SPEAKER SEX & AGE

- Listeners' judgement of the sex and age of the speaker is affected about equally by the GPR and VTL (SER) of the vowel (Fig. 4 panel 7).
- When listeners are presented with supra-normal combinations of GPR and VTL, the VTL information is more heavily weighted than the GPR information.

4

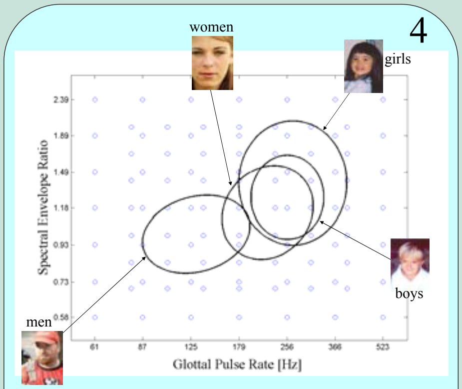

FIGURE 1. The four ellipses show estimates of the normal range of GPR and SER values in speech for men, women, boys and girls (derived from Peterson and Barney, 1952). In each case, the ellipse encompasses 90% of individuals in the Peterson and Barney data for that category of speaker. The open circles are combinations of GPR and SER values used in the experiment. The abscissa is GPR and the ordinate is SER, plotted on logarithmic axes.

5

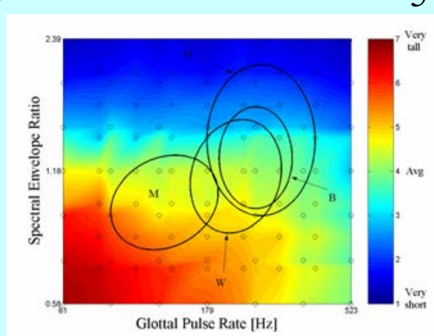

FIGURE 2. Speaker size rating judgements presented as a 2D surface plot with colour showing perceived speaker size. The 7-point ordinal speaker size rating scale goes from 1 (meaning "very short") to 7 (meaning "very tall"). Sample points are shown as circles with interpolation between the data points. Data is collapsed across the five vowels and four listeners giving 100 trials per point.

6

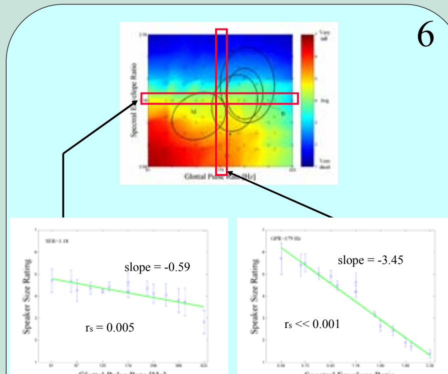

FIGURE 3. Speaker size rating as a function of GPR and as a function of SER (VTL). Error bars are one standard error of the mean (across four listeners). Best-fitting regression lines calculated for speaker size rating as a function of natural logarithm of parameter. Probabilities are one-tailed Spearman's rank order correlations for non-parametric variables.

7

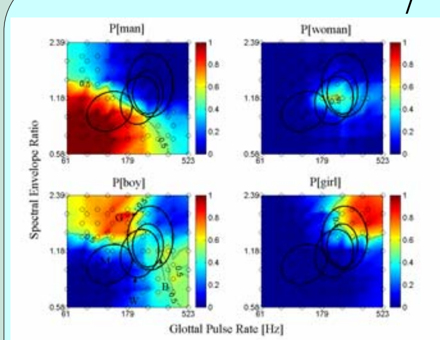

FIGURE 4. Sex categorisation performance. Data are presented as 2D surface plots with colour showing probability of assigning a given GPR-SER combination to one of four sex categories. The dotted black contour line marks our classification threshold, that is, a probability ≥0.50 of consistently choosing one category out of the four available. Data is collapsed across the four listeners giving 100 trials per point.

8 METHODS

CANONICAL VOWELS Vowels (/a/, /e/, /i/, /o/, /u/) were extracted from a natural /hVd/ speech sequence spoken by an adult male (RP) – *haad*, *hayed*, *heed*, *hoed*, *who'd*. Sounds were digitized with 16-bit quantification and a sampling rate of 44.1 kHz. All vowels were 500 ms.

SCALE MANIPULATION Vowels were manipulated to have a range of GPRs and simulated VTLs 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.

EXP We used a single-interval, two-response rating paradigm. Listeners (n=4) heard a scaled version of one of five English vowels (pseudo-randomly chosen vowel and GPR-SER value *cf*. Fig. 1), and had to make one judgement about the size of the speaker (very short, short, quite short, average, quite tall, tall, very tall) and a second judgement about the sex of the speaker (man, woman, boy, girl). The level of the vowel was roved in intensity over a 10 dB range. There was no feedback.

9 REFERENCES

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 Communication 27, 187-207.*

Smith, D. R. R., Patterson, R. D., and Jefferis, J. (2003). "The perception of scale in vowel sounds," *British Society of Audiology, Nottingham.* ***P35.***

Smith, D. R. R., and Patterson, R. D. (2004). "The existence region of scaled vowels in pitch-VTL space," *18th International Conference on Acoustics, Kyoto Japan, vol I., 453-456.*

ACKNOWLEDGEMENTS

Research supported by the UK MRC (G9901257; G9900369) and the German Volkswagen Foundation (VWF 1/79 783).