Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers

H. Hatano[1], T. Kitamura[1], H. Takemoto[2], P. Mokhtari[2], K. Honda[3], S. Masaki[4]

1) Faculty of Intelligence and Informatics, Konan University, Japan, 2) Universal Communication Research Institute, NICT, Japan, 3) University of Paris III, 4) ATR/ATR-Promotions, Japan

hatano.hiroaki@gmail.com, t-kitamu@konan-u.ac.jp

AIM OF OUR RESEARCH

1. To clarify the relationships of physical measures and acoustic parameters using a magnetic resonance imaging (MRI) database
   - Physical measures: vocal tract length (VTL) and body height
   - Acoustic parameters: pitch and formant frequencies
2. To verify whether the individual difference of formant frequencies is caused by the speakers' vocal tract length

MATERIALS AND METHODS

☐ATR vocal tract MRI data (released by ATR-Promotions)

Contents of the database

Speakers: 15 native Japanese adult male speakers

Speech materials: five Japanese vowels (/a/, /e/, /i/, /o/, /u/)

MRI data of the midsagittal plane obtained during speaking a vowel

Speech sounds recorded during the MRI scan

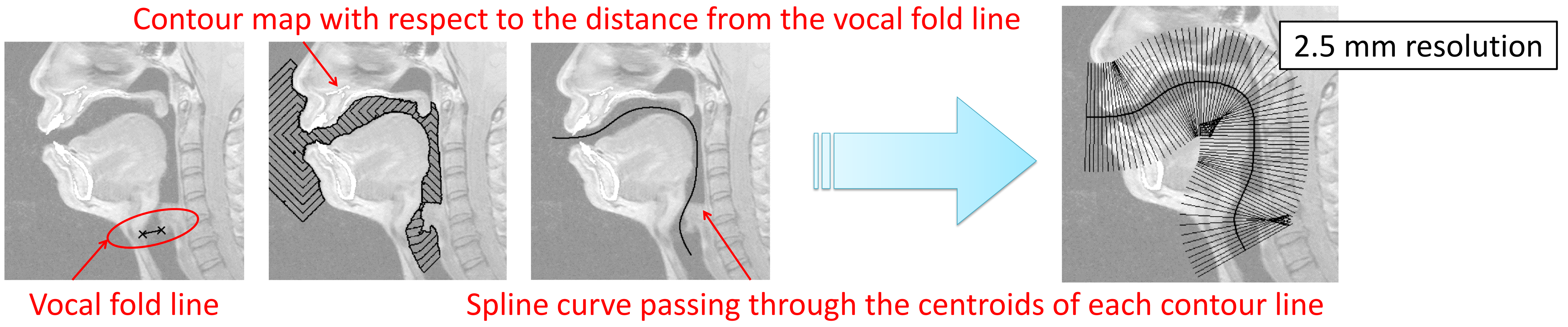
Vocal tract length extracted from the MRI data

Speakers' body heights (self-reported) and ages

☐Data acquisition methods

Vocal tract length

- The vocal tract length is defined by the glottis-to-lips length of the vocal tract midline with a 2.5 mm resolution obtained by the method proposed by Takemoto *et al.* (2006)



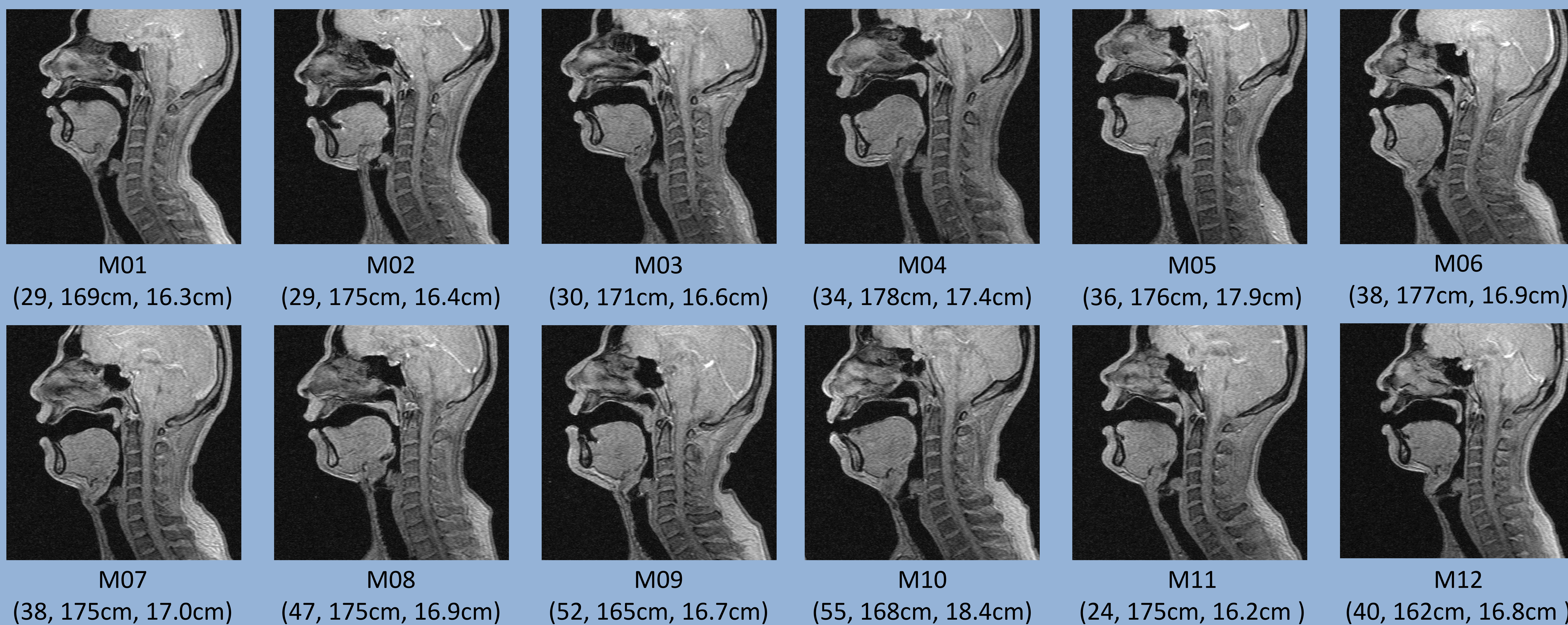
Formant frequencies

- The F1 - F4 were measured from the log spectral envelopes of the vowel segments calculated by the unbiased log spectral estimation (Imai & Furuichi 1987) and averaged with respect to the frames (frame length: 64ms, frame period: 16ms, order of the cepstrum: 60, number of iterations: 3)

Pitch frequencies

- The pitch frequency was extracted using the Pitch Contour function of WaveSurfer with its default parameters (frame length: 7.5ms, frame period: 10ms)

RESULTS



	Range	Mean	S.D.
Age	24 - 55	37	9
Body height	162 - 184	173	6
VTL	15.3 – 19.3	17.0	0.8

Figure 1: *MR images of each speaker during uttering the vowel /a/. The speakers' age, body height, and vocal tract length which averaged five vowels are showed under their images. The table on the bottom-right corner indicates the range, mean, and standard deviation of speakers' age, body height, and vocal tract length.*

Table 1: *Correlation coefficient (r) between vocal tract length and formant frequencies.*

	/a/	/e/	/i/	/o/	/u/
F1	-.16	-.65	-.27	-.48	-.46
F2	-.15	-.56	-.07	-.50	.05
F3	-.21	-.43	-.20	-.41	.22
F4	-.24	-.58	.38	-.28	.23

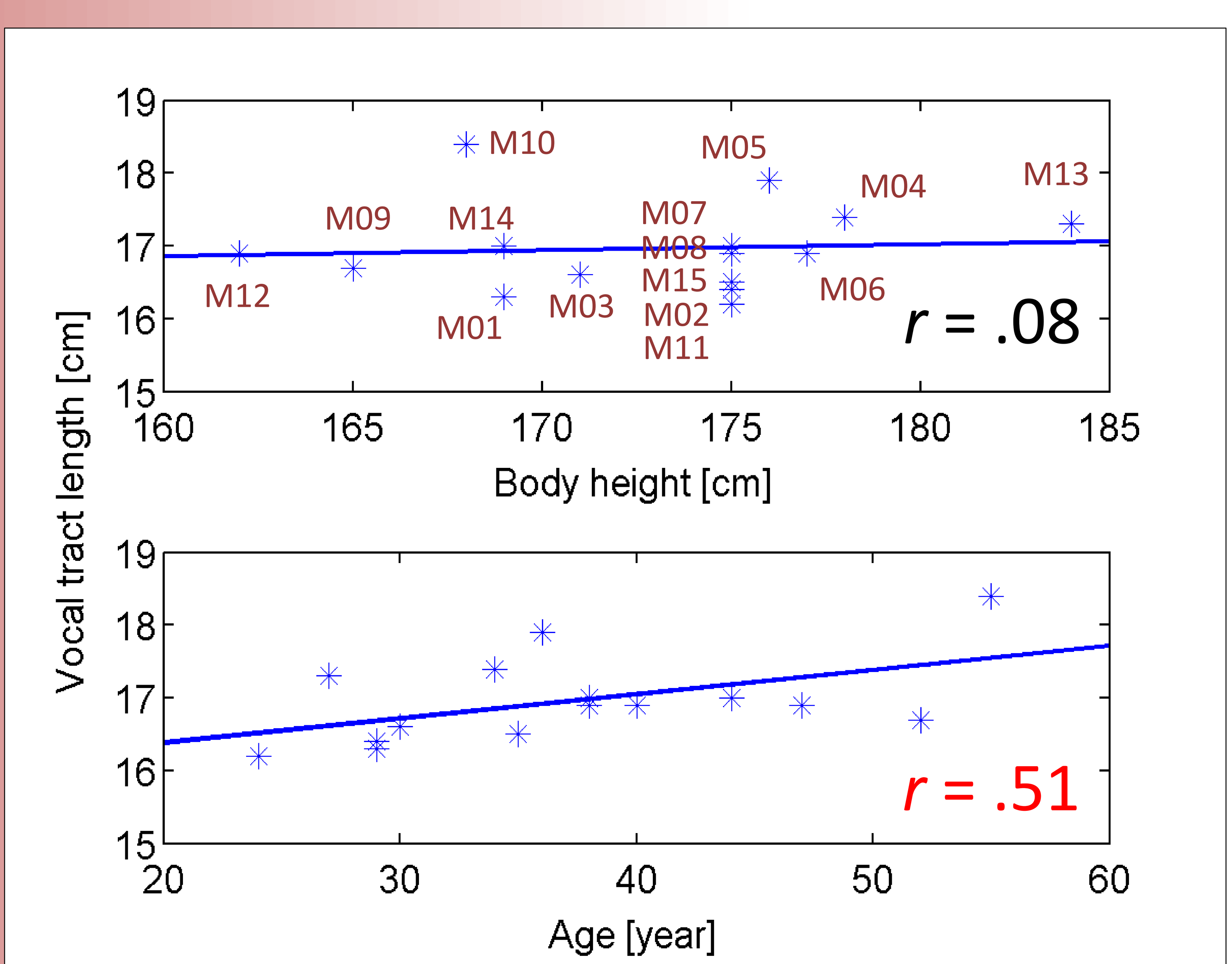


Figure 2: *Upper panel: correlation between mean vocal tract length and body height. Lower panel: correlation between vocal tract length and age.*

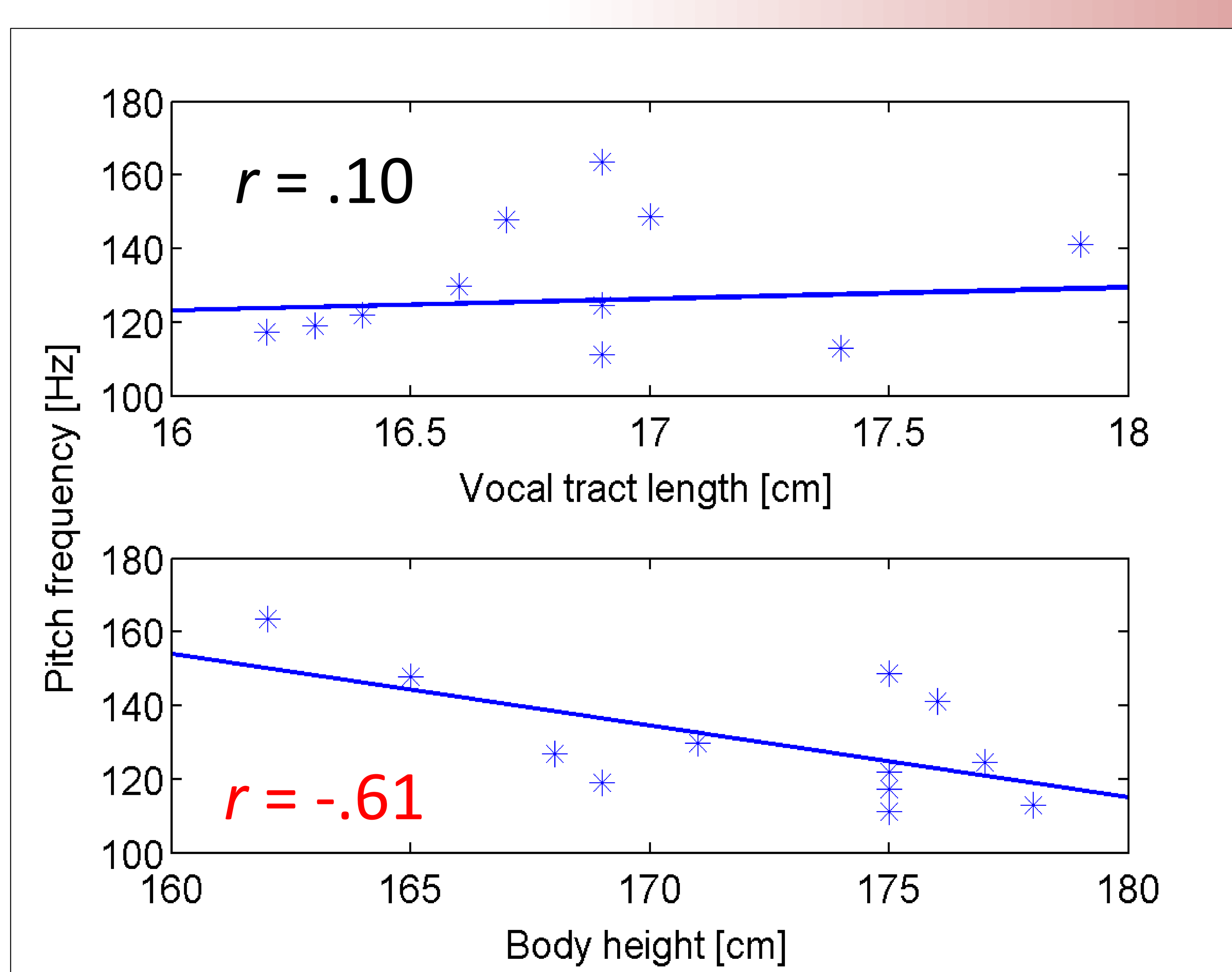


Figure 3: *Upper panel: correlation between mean vocal tract length and pitch frequency. Lower panel: correlation between body height and pitch frequency.*

DISCUSSION

- ☐ There is no correlation between VTL and body height (r = .08)
  - ➢ Fitch and Giedd (1999) indicated that correlation coefficient between them is .92.
  - ➢ Participants' ages is 2 to 24 in their study, but is 24 to 55 in our study.
  - ➢ In the former range, the vocal tract shape is strongly changed with growth. In our range, speech apparatus is already fully developed.

- ☐ There are negative correlations between VTL during uttering the vowel /e/ and its formant frequencies (except for F3)
  - ➢ In the production of /e/, there is no strong constriction in the vocal tract. Thus the vocal tract shape of /e/ is closest to a uniform tube among the Japanese vowels.
  - ➢ For the other vowels, the position and the cross-sectional area at the constriction much affect formant frequencies.

- ☐ There is a positive correlation between VTL and age (r = .51)
  - ➢ The larynx is lowered with increasing age, especially for elderly people.

- ☐ There is a negative correlation between the pitch frequency and body height (r = -.61)
  - ➢ This suggests that the length of the vocal fold may track the body height, even for a fully developed, adult population.

CONCLUSION & FUTURE WORKS

- ☐ Conclution
  - ✓ Individual differences in the formant frequencies, which contribute to voice characteristics, might not to be mainly caused by differences in the VTL within the adult male group.
  - ✓ One of the dominant factors causing speaker-to-speaker differences in the formant frequencies is the length of the vocal tract for the vowel /e/.
  - ✓ The age could be one of the factors for variability of the VTL even for adults.

- ☐ Future works
  - ➢ To reinforce the results obtained in this study by increasing the number of speakers.
  - ➢ To investigate the correlation between the physical and acoustic factors for adult female speakers.

Acknowledgements

This study was supported by JSPS KAKENHI (Nos. 21300071 and 21500184).
The MRI data and recorded sound data used in this study are parts of the "ATR vocal tract MRI data for Japanese vowels" database that was acquired at Human Information Science Laboratories in Advanced Telecommunications Research Institute International (ATR) and released by ATR-Promotions Co. Ltd. The use of the data is under licensed agreement with ATR-Promotions Co. Ltd.

References

Takemoto, H., Honda, K., Masaki, S., Shimada, Y. and Fujimoto, I., "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," J. Acoust. Soc. Am., 119, 2, 1037-1049 (2006).

Imai, S. and Furuichi, C., "Unbiased estimator of log spectrum and its application to speech signal processing," Trans. IEICE, J70-A, 3, 471–480 (1987).

Fitch, W. T. and Giedd, J., "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am., 106, 3, 1511–1522 (1999).