Introduction
Vocal tract plays an important role in speech articulation, and the shape of which formed by the tongue, lips, and jaw determines the acoustic features of speech sounds. General theory of the relationships between anterior and posterior formant frequency suggests that the first formant frequency (F1) is inversely relevant with the tongue height, and the second formant frequency (F2) relates to the anterior-posterior movements of tongue. The principles tend to consider the vocal tract a bent uniform tube with a mechanical valve of vocal cords, and may be used to manifest the physiological and individualized features of tongue during speech. In this study, we presented a new method for objectively measuring the anterior cavity and the posterior tube that produce vowel F1 and F2 using ultrasound imaging and processing in healthy young adults.

Materials & Methods

• Participant and Voice Recording:
  Eighteen healthy adults (8 males, 10 females) aged between 20 and 40 years were enrolled, and a female who underwent midline resection of hyoid bone post Sistrunk surgery was also included for providing referential images of epiglottis and tongue base without masking of the acoustic shadow from hyoid bone. The participants were requested to sustain vowels /a/, /i/, and /u/ for at least 6 s at an intensity of ≥90 dB to gain a good SNR in the environmental noises of ultrasound examination room.

• Ultrasonography of Tongue and Mandible:
  Clinical ultrasound was used to obtain the mandible-to-incisor thickness (m) which indicated the anterior end for oral-cavity resonance (Fig. 1A). T_{inc} was measured from the most inferior edge of mandible to the tip of lower incisor. The participants revealed a T_{inc} of 4.93 ± 0.55 cm (mean ± standard deviation) for the males and 4.37 ± 0.39 cm for the females. We respectively assumed T_{inc} = 5 cm for males and T_{inc} = 4.55 cm for females which indicated the anterior end for oral-cavity resonance.

• Acoustic Analysis of Voice Signals:
  Vocal signals of the first 0.5 s after voice onset were bypassed for avoiding instability at the beginning of phonation. For estimating the F2, though the difference in shapes of tongue base to the superior border of hyoid bone (Fig. 3 FTS). However, L_{AOC} in case of post Sistrunk surgery (Fig. 3 D to F), the tongue base border appeared in close proximity to the epiglottis for vowel /a/ so that the contour of this segment may not affect the length of resonance tube producing F1. The length of anterior oral cavity (L_{AOC}) was the linear distance from the peak to the lower incisor tip (Fig. 3, AOC), and was to show the length and shape of the anterior oral cavity. Because the incisor area was mostly masked by the acoustic shadow of mandible and thus could not be observed on an ultrasound image, TH was used to provide a reference point for identifying the lower incisor tip which was thought in close proximity to the superior end of T_{inc} anatomically. For vowel /a/, as the frontal chamber was elongated by rounding lips during phonation, 1 cm was added to L_{AOC} for complementing the essential but ultrasonographically invisible feature before further analysis.

• Statistics:
  The SPSS 17.0 for windows was used for statistics. A significant difference or correlation was assumed if p < .05. Pearson’s correlation analysis and multivariate linear regression analysis were employed to examine the relationship of the formant frequencies with TH, TA, L_{AOC}, and L_{PTS}.

Results
  As shown in Fig. 4, there was a significant, negative but poor correlation between the F1 and TH (Fig. 4A; r = 0.31, p = .002). The F1 and TA showed a moderately significant correlation (r = 0.5 0.37, p = .002). The correlations of the F2 with TA (Fig. 4B) and mandible-to-incisor thickness (m) (Fig. 4C) were both significant (p = .001), and TH showed to have a slightly better estimate of the F2 (r = 0.85) than TA (r = 0.78). Fig. 5 showed that L_{AOC} and L_{PTS} were better parameters to present the differences of tongue position and shape of the production of vowels than were TH and TA. In multivariate linear regression analysis, the F1 was significantly correlated only with L_{AOC} and the F2 was significantly correlated with TA and L_{PTS} (r = -0.90, p < .001).

Discussion and Conclusion
  The study results suggest that L_{AOC} and L_{PTS} are better estimators of the first two vowel formants than the tongue height and tongue advancement. The results also show that L_{PTS} allows us to estimate the F1 with better accuracy, but also reveal that the production of the F1 for various vowels mainly depends on both the placement and shape of tongue base. The results also deduce that sound waves might travel along the surface of tongue like traveling through a bent tube rather than a chamber. For estimating the F2, though the difference in r value between the L_{AOC} and TA may not have a significant effect on the clinical outcome of the F2 estimation, it will be beneficial for clinical examination to efficiently estimate the F2 from tongue position by ultrasound measurement of L_{PTS} due to its simplicity and high viability of oral cavity.

Financial Disclosure: the research was supported by the grant MOST 103-2314-B-010-055-MY3 of Ministry of Science and Technology, Taipei, Taiwan.

Fig. 1: Measurement of (A) mandible (B) tongue peak to a perpendicular line running through the hyoid bone (Fig. 2, TA). TH was then used for comparing the correlation with the F1 and TA was used for analyzing the correlation with F2.

Fig. 2: Tongue height (TH) and tongue advancement (TA) in vowels

Fig. 3: Lingual ultrasound images in a healthy adult (A,C) and a case post Sistrunk surgery (D,F)

Fig. 4: Pearson’s correlation analysis

Fig. 5: TH and L_{PTS} for vowels /a/, /i/, and /u/ (A) and TA and L_{AOC} for vowels /a/, /i/, and /u/ (B). * p < 0.05; † p < 0.01; ‡ p < 0.001.