A morphological and acoustic analysis of the vocal tract during the act of whistling

Kaburagi, Tokihiko
Faculty of Design, Kyushu University

Shimizu, Takuma
Graduate School of Design, Kyushu University

Uezu, Yasufumi
Graduate School of Design, Kyushu University

出版情報：Acoustical Science and Technology. 39 (3), pp.198-206, 2018. 日本音響学会
バージョン：

Acoust. Sci. & Tech. **39**, 3 (2018)

PAPER

A morphological and acoustic analysis of the vocal tract during the act of whistling

Tokihiko Kaburagi[1,*], Takuma Shimizu[2] and Yasufumi Uezu[2]

[1]*Faculty of Design, Kyushu University,*
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan
[2]*Graduate School of Design, Kyushu University,*
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan

(*Received 28 October 2016, Accepted for publication 10 November 2017*)

Abstract: The human whistle is a typical aeroacoustic sound. Downstream of a small orifice made by the lips, a jet is formed by airflow with a high Reynolds number. A sequence of vortex rings is then produced, and periodic air pressure changes result in a characteristic whistling sound. Although the vocal tract has been reported to act as an acoustic resonator determining the blowing pitch, the precise shape of the vocal tract and its resonance properties during whistling remain unclear. In the current study, the morphological and acoustic properties of the vocal tract were examined during the act of whistling in a single participant. The vocal tract was scanned in three dimensions using magnetic resonance imaging while four musical notes were produced. The data revealed that the tongue constricted the vocal tract in different ways depending on the note, and the location of the constriction moved forward when the blowing pitch increased. Acoustic analysis of the vocal tract showed that the second peak of the lip input impedance was largely in accord with the whistling pitch. In addition, specific regions in the vocal tract were highly acoustically sensitive to small deformations.

Keywords: Whistling, Blowing pitch, Vocal-tract shape, Lip impedance, Acoustic sensitivity

PACS number: 43.70.Aj, 43.75.Zz [doi:10.1250/ast.39.198]

1. INTRODUCTION

The human vocal tract plays a range of acoustic roles, in both linguistic and non-linguistic communication. In the production of speech, the vocal tract can act as a resonator, or as an anti-resonator, acoustically filtering source signals, determining the spectral characteristics of speech, and finally conveying linguistic content to a listener [1,2]. The vocal tract is also able to produce plosive and fricative sounds. In addition, when the fundamental frequency of voice approaches a resonance frequency of the vocal tract, the vocal tract influences vibrational movements of the vocal folds and the flow of air passing through the glottis [3–5]. This type of interaction between a sound source mechanism and an acoustic filter comprises the central blowing mechanism of wind instruments [6]. Importantly, the vocal tract of a wind instrument player can affect the sound, in addition to the bore of the instrument. For example, skilled players of the trumpet change the vocal-tract shape according to the blowing pitch [7].

*e-mail: kabu@design.kyushu-u.ac.jp

The vocal tract also plays a role in whistling, which is not generated by the vibration of the vocal folds. Whistling is different from fricative speech sounds, because the former typically entails a sensation of pitch. Wilson *et al.* studied the effect of the vocal-tract geometry on the generation of whistle [8]. They hypothesized that the vocal tract has two small orifices (one between the lips and the other between the tongue and the roof of the oral cavity), and constructed a cylinder model of the vocal tract to conduct aeroacoustic sound generation experiments. Based on their results, the authors concluded that there are three essential features in the production of whistling: the formation of a jet, the production of vortex rings from the resulting flow instability, and the interaction between these rings and a rigid boundary. The formation of a jet and flow instability commonly occur in the vicinity of the outlet of a narrow channel (e.g., the glottis in the case of voice, and lip aperture in the case of brass instruments). As a jet forms, the flow of air detaches from the wall and coincidentally generates a wake. For whistling, the flow separation phenomenon results in the generation of vortex rings. Rayleigh [9] and Wilson *et al.* [8] also propose that

the resonance characteristics of the vocal tract are essential for determining the frequency at which a sequence of vortex rings is generated. As a result, the air pressure at the outlet of a small orifice changes periodically at that frequency, enabling the sound of a whistle to exhibit an audible pitch.

Although Wilson *et al.* highlighted the importance of vocal-tract geometry in whistling, the actual geometry of the vocal tract has not been directly examined in a human subject. In addition, the hypothesis that the vocal tract has two orifices remains to be verified. We therefore carried out an experiment to observe the shape of the vocal tract in three dimensions using magnetic resonance imaging (MRI) while a participant whistled a range of musical notes. MRI is a safe, noninvasive imaging technique for use in humans, and is capable of obtaining a large amount of scan data [10,11]. A previous study of volumetric scanning of the vocal tract reported that the cross-sectional area function could be accurately estimated [12,13]. In addition, area function data has been reported to be useful for examining the acoustic characteristics of the vocal tract using an acoustic tube model including energy losses on the surface of the tract [14]. Based on these techniques developed in the research field of speech production, the characteristic features of the vocal tract were analyzed in the current study from both morphological and acoustic perspectives. The results revealed that, when the blowing pitch was low, the vocal tract appeared to have two orifices, as hypothesized by Wilson *et al.* [8]. However, when the blowing pitch was high, only a single constriction was made between the upper and lower lips and in the region behind them. Based on this finding, we propose that vortex rings are produced only at the orifice between the upper and lower lips during whistling. The acoustic characteristics of the vocal tract were computed under this assumption throughout the rest of the study.

As another acoustic analysis method of the vocal tract, we adopted a sensitivity function and examined the relationship between the vocal-tract shape and peaks of the acoustic resonance (i.e., the input impedance seen from the lips). The sensitivity function can theoretically be derived from a theorem reported by Ehrenfest [15]. This function represents the change in formant frequency related to small perturbations in the vocal-tract cross-sectional area [16] or the vocal-tract length [17]. Recent studies reported that this function can be used for recovering the vocal-tract geometry from specified formant frequencies [17–19]. However, these sensitivity functions did not fit the current experimental paradigm, because the vocal-tract resonance was determined for whistling as the peak of the input impedance. We therefore computed the perturbation relationship on the basis of a complex sensitivity function [20]. Using this sensitivity analysis, we demonstrated that the anterior cavity of the vocal tract was more sensitive than the posterior cavity to the frequency of an impedance peak. Thus, the blowing pitch was determined by fine control of the geometry of the vocal tract, particularly around the anterior portion.

This paper is organized as follows. Section 2 provides a description of the image acquisition technique and the results of the morphological analysis. In Sect. 3, the method of calculation of input impedance and the sensitivity function are described. In addition, the results of the acoustic analysis are presented. Section 4 provides a discussion of our findings in the context of previous research, and the conclusions that can be drawn from the current results.

2. MORPHOLOGICAL ANALYSIS

2.1. Imaging of the Vocal Tract

Observation of vocal tract shape is important for understanding the mechanisms underlying the production of a whistling sound. Thus, in the current study, we used MRI to non-invasively image the vocal tract. The participant was an adult male with musical experiences as an amateur trombone player for about 10 years. Sectional images of the vocal tract were first scanned in three dimensions while the participant whistled each of four musical notes: F5 (698 Hz), B♭5 (933 Hz), F6 (1,396 Hz), and B♭6 (1,865 Hz). The shape of the upper and lower teeth was then measured independently, because they are as transparent as air in standard vocal-tract scanning. Following the method of Takemoto *et al.* [13], a contrast medium was inserted into the oral cavity so that the outline of the teeth could be distinguished from the oral cavity. The oral shape was then scanned in the resting state of the vocal tract without whistling, using the usual scanning procedures, as described below.

Each slice of volumetric data was measured in the sagittal plane. The slice thickness was 2 mm for the vocal tract and 1.5 mm for the teeth. There were 30 slices for the vocal tract and 51 slices for the teeth. Obtaining a set of volumetric data required a scanning time of approximately 30–40 s, and the participant was instructed to maintain as much stability as possible in the configuration of the vocal tract during the scanning. The resolution of each slice was 512×512 pixels covering a scanning area of 256×256 mm^2. The measurement was made by the ATR-Promotions Inc. (Kyoto, Japan).

Figure 1 shows the midsagittal images of the vocal tract during the production of each musical note. The main speech articulators, including the tongue, lips, and soft palate, can be clearly seen in the images. For the two low notes (F5 and B♭5), the tongue was located slightly towards the back of the tract, producing a vocal-tract constriction between the tongue and soft palate separately from the

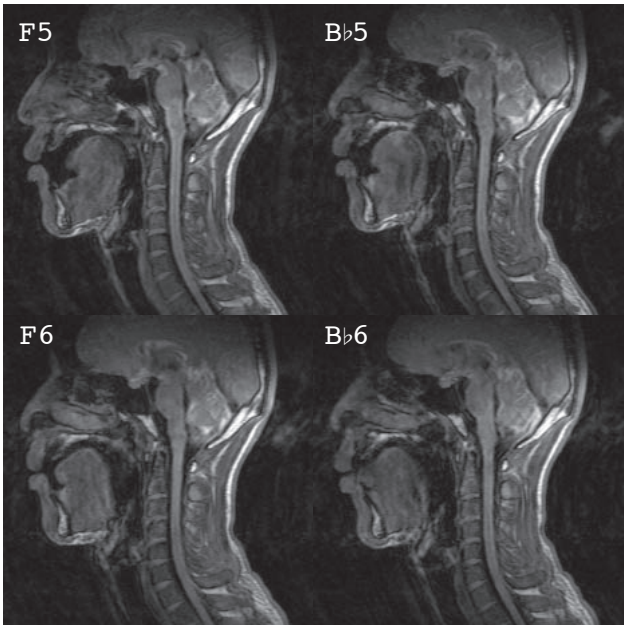

Fig. 1 Midsagittal images of the vocal tract during whistling, measured with MRI during the production of four musical notes.

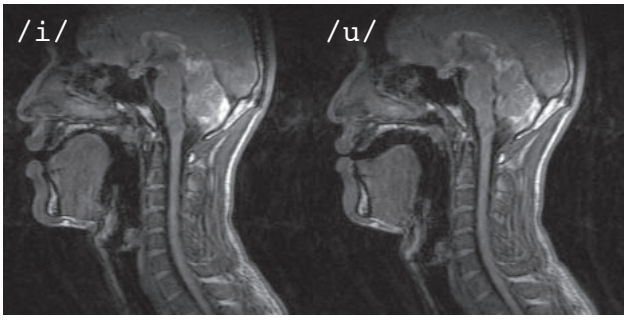

Fig. 2 Midsagittal images of the vocal tract measured with MRI during the production of /i/ and /u/ vowel sounds uttered by the same participant.

constriction made by the upper and lower lips. When the note increased in pitch to F6 and B♭6, the tongue was positioned more anteriorly. The vocal tract was then constricted between the tongue and the hard palate just behind the constriction of the lips.

The vocal tract is ordinarily used for producing speech. The lip opening is particularly small during the production of the vowels /i/ and /u/, and the tongue is located in the frontal or mid region of the vocal tract, creating a narrow constriction. Therefore, understanding the degree to which the posture of the vocal tract is similar between speech production and whistling may also be helpful. To compare the position of the vocal organs shown in Fig. 1 during whistling with the vocal organ position in the production of speech, Fig. 2 shows imaging results during the production of two Japanese vowels (/i/ and /u/) recorded in the same participant. The Japanese /u/ vowel is classified as a mid

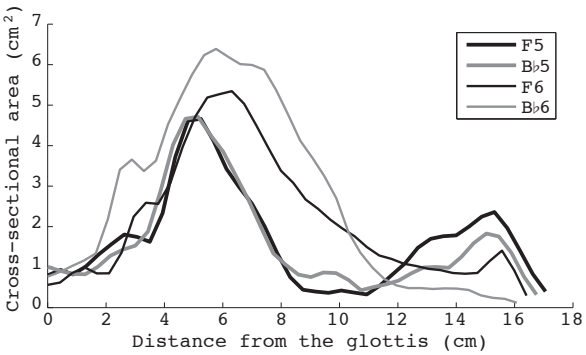

Fig. 3 Vocal-tract area function obtained from volumetric MRI. The ordinate is the cross-sectional area and the abscissa is the distance from the glottis.

vowel. The tongue was found to constrict near the border of the hard and soft palates, but the constriction was much narrower for the two low notes of the whistle. For the two high notes of whistle, the configuration of the narrow channel formed between the tongue and the hard palate was somewhat similar to that of the /i/ vowel, but the overall configuration of the tongue was markedly different, especially around the back and root portions of the tongue. The lips were rounded during whistling, and pulled sideways during the production of the /i/ vowel.

2.2. Cross-sectional Area Function

Next, we estimated the cross-sectional area function of the vocal tract using volumetric images. The volumetric images for the upper and lower teeth were first superimposed on the vocal-tract images, based on an affine congruent transformation [13]. Before applying the transformation, the position of the teeth was manually marked on images. When the transformation was performed and the marked teeth area was mapped onto the tract images, the correlation coefficient between the teeth and tract images was calculated for the marked area. The optimal affine transformation was then determined so that the value of the correlation coefficient was largest.

The area function was then obtained using the vocal-tract images with the teeth superimposed. First, the inferior and superior outlines of the vocal tract were traced by sight on the midsagittal plane. The center line of the tract, representing the axis of wave propagation, was determined so that the distances of each node point on the center line from the two outlines were equal. The vocal tract was then divided along the center line into 40 sections of equal length. Finally, the cross-sectional area in the direction perpendicular to the center line was obtained for each section.

The obtained cross-sectional area function data are plotted in Fig. 3, in which the ordinate is the cross-sectional area and the abscissa is the distance from the

glottis along the center line of the vocal tract. For the two low notes (F5 and B♭5), it can be seen that the vocal tract constricted around the point 10 cm to 11 cm away from the glottis, clearly separating the vocal tract into anterior and posterior cavities. The cross-sectional area of the constriction was almost the same as that at the lip end, supporting the two-constriction hypothesis of Wilson *et al.* [8]. On the other hand, the cross-sectional area, and hence the volume of the anterior cavity decreased as the pitch increased. The midsagittal images in Fig. 1 and the cross-sectional area data in Fig. 3 show that the vocal tract was no longer separated into anterior and posterior cavities when the note increased to F6 and B♭6. Although only a single subject participated in our experiment, our morphological data indicate that the two-constriction hypothesis [8] doesn't hold when the blowing note is high. The length of the vocal tract decreased as the note increased, and was estimated to be 17.1 cm for F5, 16.7 cm for B♭5, 16.4 cm for F6, and 16.1 cm for B♭6.

3. ACOUSTIC ANALYSIS

3.1. Input Impedance of the Vocal Tract at the Lips

The acoustic characteristics of the vocal tract were analyzed using cross-sectional area function data. A frequency-domain acoustic tube model [14], which properly accounts for energy losses along the wall of the vocal-tract tube, was used for the analysis. The morphological analysis revealed that the vocal tract constricted only at the lip-end portion when the blowing note was F6 or B♭6 (Fig. 3).

Because the sound source of the whistle is known to be a sequence of vortex rings generated near the outlet of a constriction, the results of our morphological analysis suggest that the sound source was likely to be located in the vicinity of the lip-end, at least for these notes. Based on this assumption, we calculated the input impedance of the vocal tract observed from the lips as an acoustic property that may be involved in adjustment of the blowing pitch. In addition, because the glottis is constantly open during whistling so that the flow of air from the lungs passes smoothly through the glottal orifice, we included a connection between the sub-glottal airway and the vocal tract in our analysis.

The area function data of the vocal tract was described in the previous section. For the sub-glottal airway, the morphological data were obtained from a study by Weibel [21]. The airway can be classified into conductive, transitory, and respiratory zones, where the cross-sectional area is relatively constant along a section approximately 20 cm below the glottis, but increases rapidly in the remainder of the conductive zone and in the transitory and respiratory zones. The branching of airways was not directly considered in our model, and the cross-sectional area for each stage was summed to form a single uniform tube [22]. The resulting area function data was then used to calculate the input impedance of the vocal tract, where the total length of the sub-glottal airway was shortened uniformly by a factor of 0.941 [22].

When the vocal tract sections from the lips to the glottis were defined as $1, 2, \ldots, N$ and the sub-glottal sections from the glottis to the lungs were defined as $1, 2, \ldots, M$, the propagation matrix for the vocal tract is given as

$$K_{\mathrm{v}} = \begin{pmatrix} A_{\mathrm{v}} & B_{\mathrm{v}} \\ C_{\mathrm{v}} & D_{\mathrm{v}} \end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix} A_{\mathrm{v}i} & B_{\mathrm{v}i} \\ C_{\mathrm{v}i} & D_{\mathrm{v}i} \end{pmatrix}, \tag{1}$$

the propagation matrix for the sub-glottal airway as

$$K_{\mathrm{s}} = \begin{pmatrix} A_{\mathrm{s}} & B_{\mathrm{s}} \\ C_{\mathrm{s}} & D_{\mathrm{s}} \end{pmatrix} = \prod_{i=1}^{M} \begin{pmatrix} A_{\mathrm{s}i} & B_{\mathrm{s}i} \\ C_{\mathrm{s}i} & D_{\mathrm{s}i} \end{pmatrix}, \tag{2}$$

and the propagation matrix for the total airway as

$$K = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A_{\mathrm{v}} & B_{\mathrm{v}} \\ C_{\mathrm{v}} & D_{\mathrm{v}} \end{pmatrix} \begin{pmatrix} A_{\mathrm{s}} & B_{\mathrm{s}} \\ C_{\mathrm{s}} & D_{\mathrm{s}} \end{pmatrix}. \tag{3}$$

Matrix components for the vocal tract ($A_{\mathrm{v}i}$, $B_{\mathrm{v}i}$, $C_{\mathrm{v}i}$, and $D_{\mathrm{v}i}$) are determined by the length (L_i) and cross-sectional area (S_i) for the ith section using the acoustic tube model [14] such that $A_{\mathrm{v}i} = \cosh(\sigma L_i/c)$, $B_{\mathrm{v}i} = -(\rho c/S_i)\gamma \sinh(\sigma L_i/c)$, $C_{\mathrm{v}i} = -(S_i/\rho c)(\sinh(\sigma L_i/c))/\gamma$, and $D_{\mathrm{v}i} = \cosh(\sigma L_i/c)$. α, β, γ, and σ are frequency-dependent parameters from the literature, such that $\alpha = \sqrt{j\omega c_1}$, $\beta = j\omega\omega_0^2/\{(j\omega + a)j\omega + b\} + \alpha$, $\gamma = \sqrt{(\alpha + j\omega)/(\beta + j\omega)}$, and $\sigma = \gamma(\beta + j\omega)$, where $a = 130\pi$ rad/s, $b = (30\pi)^2\,(\mathrm{rad/s})^2$, $c_1 = 4\,\mathrm{rad/s}$, $\omega_0^2 = (406\pi)^2\,(\mathrm{rad/s})^2$, and $j = \sqrt{-1}$. ρ is the air density and c is the speed of sound. In this study, ρ was set to $1.184 \times 10^{-3}\,\mathrm{g/cm^3}$ and c to 34630 cm/s, respectively.

Matrix components for the sub-glottal airway ($A_{\mathrm{s}i}$, $B_{\mathrm{s}i}$, $C_{\mathrm{s}i}$, and $D_{\mathrm{s}i}$) are also calculated using the same acoustic tube mode [14], but the wall loss is greater for the sub-glottal airway than that for the vocal tract because of an effective increase in the wall area. This effect was approximated by changing the parameter value of the tube model as $a = 780\pi\,\mathrm{rad/s}$ and $c_1 = 16\,\mathrm{rad/s}$ so that the input impedance of the sub-glottal airway seen from the glottis was in accord with the numerical data reported in the literature [22]. Figure 4 shows the calculated input impedance of the sub-glottal airway.

The input impedance of the total airway including the vocal tract and sub-glottal airway, Z, is then calculated as

$$Z = -\frac{AZ_{\mathrm{p}} + B}{CZ_{\mathrm{p}} + D}, \tag{4}$$

where $Z_{\mathrm{p}} = \rho c^2/(j\omega V)$ is the terminal impedance modeled as an air volume (V) of 500 ml [23]. In addition, the transfer function (H_{v}) and the input impedance (Z_{v}) of the vocal tract are calculated as follows under the assumption of complete glottal closure:

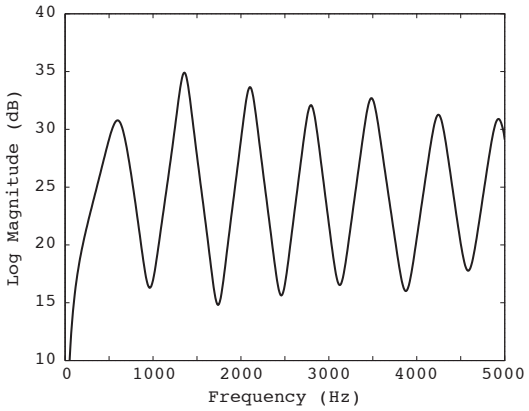

Fig. 4 Calculated input impedance of the sub-glottal airway seen from the glottis.

$$H_{\mathrm{v}} = \frac{1}{A_{\mathrm{v}} - C_{\mathrm{v}} Z_{\mathrm{r}}} \tag{5}$$

and

$$Z_{\mathrm{v}} = -\frac{A_{\mathrm{v}}}{C_{\mathrm{v}}}, \tag{6}$$

where Z_{r} is the radiation impedance of the lip aperture [24]. Note that the sign for the traveling wave of the volume velocity from the lungs to the lips is positive, and the input impedance is defined as the ratio between the acoustic pressure and the volume flow at the lip end when the vocal tract is seen from the lips. The transfer function is the volume velocity at the lip end over the volume velocity at the glottis.

Figure 5 shows the numerical results. When the input impedance Z_{v} (thin lines) for a specific musical note is compared with the transfer function H_{v} (broken lines) for the same note, it is clear that each peak of the transfer function corresponded to an impedance dip of the same frequency. Here, note that Z_{v} and H_{v} were obtained under the assumption of glottal closure. The frequency for the 1st impedance peak was 200, 200, 200, and 193 Hz for F5, B♭5, F6, and B♭6, respectively, then almost the same among four notes. The corresponding frequencies for the 2nd impedance peak were 708, 892, 1,358, and 1,755 Hz, and we found that the 2nd peak was approximate agreement with the pitch of each note. The absolute error proportions between the pitch and the peak frequency were 1.4%, 4.4%, 2.7%, and 5.9% for F5, B♭5, F6, and B♭6, respectively. These results suggest that the 2nd peak of the input impedance is a crucial acoustic factor. Thus, it is possible that the mouth shape is changed in accord with the

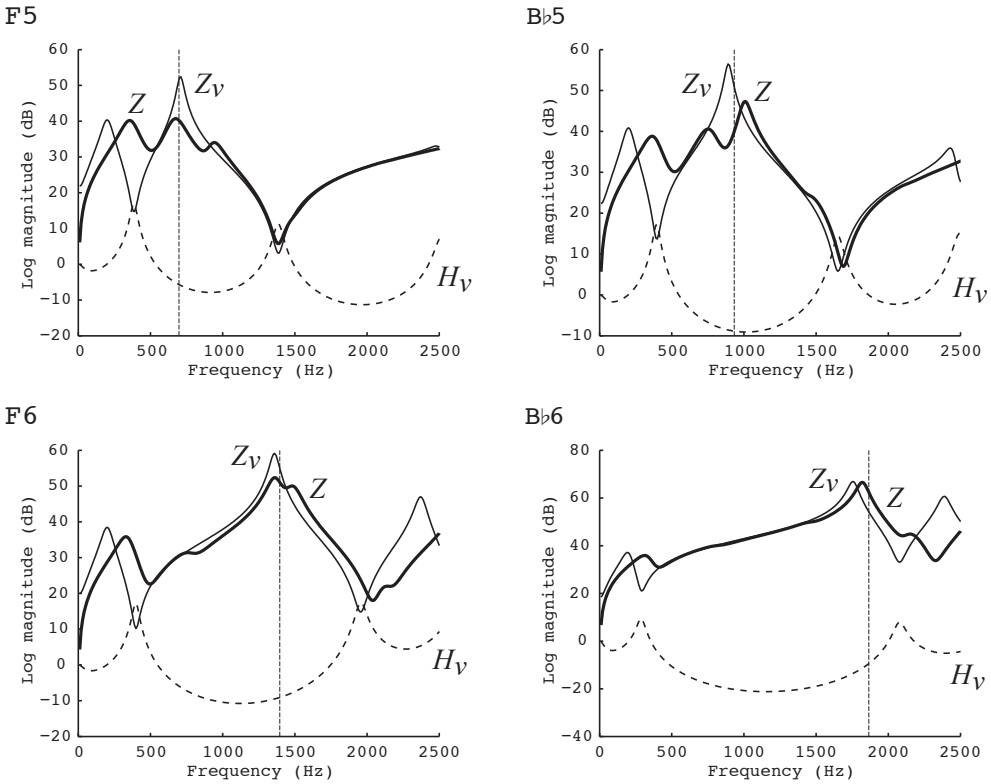

Fig. 5 Calculated acoustic characteristics of the vocal tract for the musical notes F5, B♭5, F6, and B♭6. The thick line shows the input impedance observed from the lip aperture (Z) when the vocal tract is acoustically connected to the sub-glottal airway. The thin and broken lines are the input impedance (Z_{v}) and the transfer function (H_{v}) of the vocal tract, respectively, when the complete closure of the glottis is assumed. The vertical line represents the pitch of each musical note.

blowing pitch during whistling, and the 2nd impedance peak may determine the frequency at which a sequence of vortex rings is periodically produced from the lip aperture, constituting the sound production mechanism of a whistle.

The thick lines in Fig. 5 show the results of the analysis of input impedance, Z, when the glottis was not closed and the vocal tract was acoustically connected to the sub-glottal airway. The acoustic characteristics of the sub-glottal airway then affected the impedance value and we observed the following differences when the impedance Z was compared with Z_v. First, the frequency for the 1st peak of Z was greater than the corresponding peak frequency of Z_v. The magnitude of each peak decreased and the band width increased in Z. Additional peaks and dips were also observed, which were not seen when glottal closure was assumed.

The frequency for the 2nd peak of Z did not substantially change from that of Z_v, except for B♭5, for which the peak frequency increased from 892 Hz to 1,008 Hz as a result of coupling to the sub-glottal airway. This finding suggests that the change in peak frequency may have been due to a decrease in input impedance for the sub-glottal airway just below 1,000 Hz, as seen in Fig. 4. The frequencies of the 2nd peak were 675, 1,361, and 1,820 Hz for F5, F6, and B♭6, respectively. For B♭5, the 3rd peak corresponded to the blowing pitch, and its frequency was 1,008 Hz. The absolute error proportions were 3.3%, 8.0%, 2.5%, and 2.4%, for F5, B♭5, F6, and B♭6, respectively.

3.2. Sensitivity Analysis between the Cross-sectional Area and Impedance Peaks

As reported previously, acoustic sensitivity function is useful for investigating the relationship between the geometry of the vocal tract and the resulting resonance properties [16,17]. If the sensitivity is high for a specific portion of the tract, a change in the cross-section area or the length of that portion may cause a change in vocal-tract resonance to a large degree. In the current study, we compared the acoustic sensitivity of the vocal tract during whistling, in which the sound source is located at the lip aperture, with that during the production of speech, in which the sound source is at the glottis, especially for vowels.

To perform sensitivity analysis, a sensitivity function must be newly derived with regard to the peaks of input impedance, because previously reported sensitivity functions [16,17] are related to peaks of the vocal-tract transfer function. To derive a sensitivity function, we used the basic concept of a complex sensitivity function [20]. Unlike the existing sensitivity functions derived from a physical principle [15], complex sensitivity functions are derived formally. Consequently, these functions are flexible, and capable of treating *any* type of resonance, or even anti-resonance. Kreuzer and Kasess [20] also reported that a complex sensitivity function for formants was largely in accord with traditional, physically-derived sensitivity functions.

The acoustic characteristics for vowel-like sound production in a vocal tract can be determined from the cross-sectional areas under the assumption of plane wave propagation and the terminal condition at both sides of the tract. In addition, only the area perturbation was considered in this study, and the length of the vocal tract was not changed. Each complex pole of the tract, z_i, can thus be represented as a function of the area parameters, S_j such as:

$$z_i = z_i(S_1, S_2, \ldots, S_j, \ldots, S_N), \tag{7}$$

where N is the number of the vocal-tract sections. When the cross-sectional area changes slightly, such as $S_j + \delta S_j$, suppose that the logarithm of a complex pole changes as $\log z_i + \delta \log z_i$. By taking the total differential, we obtain

$$\delta \log z_i = \sum_{j=1}^{N} \frac{\partial \log z_i}{\partial S_j} \delta S_j. \tag{8}$$

Thus, it holds that

$$\frac{\delta \log z_i}{\arg z_i} = \sum_{j=1}^{N} \frac{\partial \log z_i}{\partial S_j} \frac{S_j}{\arg z_i} \frac{\delta S_j}{S_j} = \sum_{j=1}^{N} R_{ij} \frac{\delta S_j}{S_j}, \tag{9}$$

where

$$R_{ij} = \frac{\partial \log z_i}{\partial S_j} \frac{S_j}{\arg z_i}. \tag{10}$$

Equation (9) is the perturbation relationship between the area parameters and the impedance peaks. Equation (10) is the complex sensitivity function. From the imaginary part of both sides of Eq. (9), we obtain

$$\frac{\delta \arg z_i}{\arg z_i} = \sum_{j=1}^{N} \Im\{R_{ij}\} \frac{\delta S_j}{S_j}, \tag{11}$$

because $\Im\{\log z_i\} = \arg z_i$. $\Im\{\ \}$ represents the imaginary part of a complex variable. If the sampling frequency is f_s, the frequency of the impedance peak is $f_i = \arg z_i \cdot f_s/(2\pi)$. The left side of Eq. (11) then represents the relative change in the peak frequency. The perturbation relationship in Eq. (11) suggests that the relative change in peak frequency may be related to a relative change in the area parameters through the sensitivity function.

The partial differential $(\partial \log z_i/\partial S_j)$ should be obtained numerically in Eq. (10). Suppose that the value of $\arg z_i$ changed to $\arg z_i'$ when the area parameter for the jth section changed to $S_j + \Delta S$. The imaginary part of the

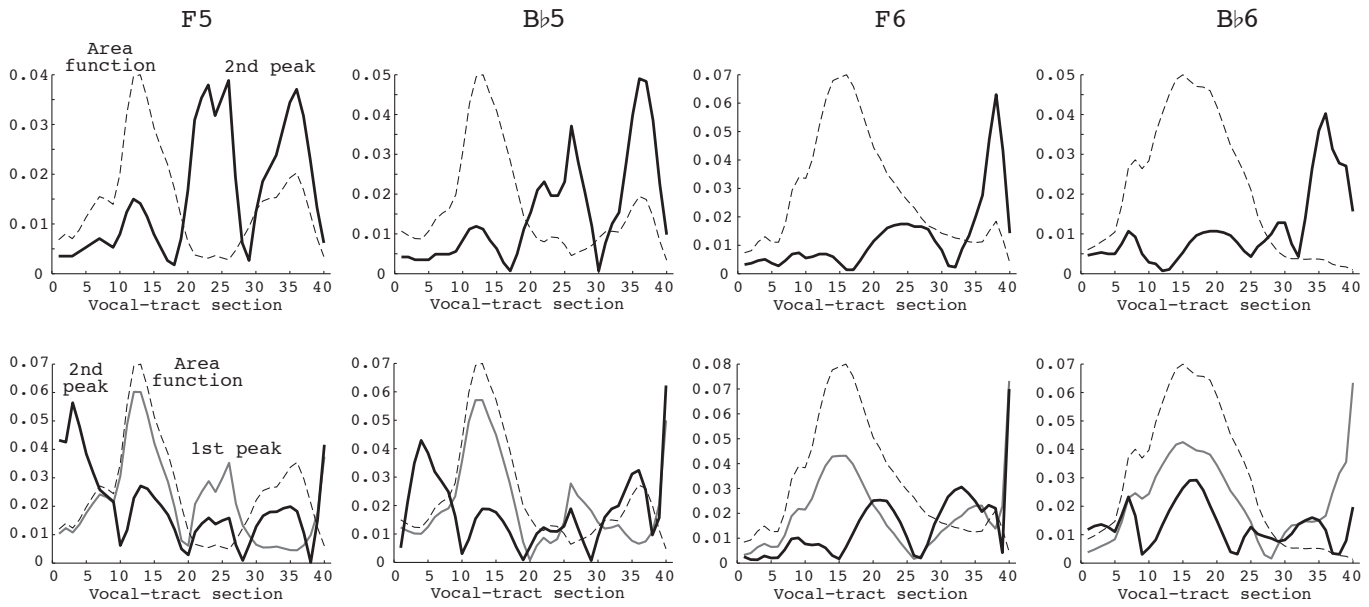

Fig. 6 Calculated acoustic sensitivity function in terms of the cross-sectional area of the vocal tract during whistling. In the upper part of the figure, the solid line shows the absolute value of the sensitivity function for the 2nd peak of the input impedance (Z_v). In the lower part, the absolute value of the sensitivity function is shown with regard to the 1st (thin) and 2nd (thick) peaks of the vocal-tract transfer function (H_v). The cross-sectional area function is also shown with broken lines for reference. The 1st section of the vocal tract corresponds to the glottis, and the 40th corresponds to the lips.

complex sensitivity function in Eq. (10) can then be obtained as:

$$\Im\{R_{ij}\} = \frac{\arg z_i' - \arg z_i}{\Delta S} \frac{S_j}{\arg z_i}. \tag{12}$$

As a result, the sensitivity function with regard to the peak frequency is a real number, and it is not necessary to calculate the logarithm of the complex pole ($\log z_i$) explicitly. In the sensitivity analysis, the input impedance was first calculated from the cross-sectional area data (S_j) and the argument of a complex pole ($\arg z_i$) was determined by finding the ith peak of the input impedance.

Numerical results were obtained for four musical notes and plotted in Fig. 6. First, the sensitivity function was calculated with regard to the peaks of the transfer function (i.e., formants) following previously reported methods [16–18]. The absolute value of the sensitivity function was plotted in the lower part of the figure for the 1st (thin lines) and 2nd (thick lines) peaks. Sensitivity is a function of vocal-tract section, and the cross-sectional area data were also plotted by the broken line to indicate the vocal-tract shape of each musical note. The sensitivity implies the capability, or ease, associated with changing the peak frequency with a small perturbation of the cross-sectional area. For note F5, for example, the figure shows that sensitivity was strongly correlated with vocal-tract shape. In addition to the vocal-tract constriction observed around the 20th to 26th sections, high sensitivity for the 2nd peak was observed in the anterior cavity region and in the posterior cavity region, in which sensitivity for the 1st peak was also noticeable. Sensitivity for the 2nd peak was quite significant in the vicinity of the lip aperture and in the region directly above the glottis.

Next, the sensitivity function was calculated with regard to peaks of input impedance (Z_v) using the method described above. The absolute value of the sensitivity function, $|\Im\{R_{ij}\}|$, was then plotted in the upper part of the figure for the 2nd impedance peak as a function of the vocal-tract section, because the 2nd peak of the input impedance was closely related to the blowing pitch, and appears to be an important mechanism in whistling production. For F5 and B♭5, the figure shows that the sensitivity of the impedance peak exhibited two noticeable peaks: one around the vocal-tract constriction and the other around the anterior cavity. On the other hand, less sensitivity was observed around the posterior cavity and the region directly above the glottis. For high notes (F6 and B♭6), sensitivity was noticeable particularly around the region directly behind the lips. Our sensitivity analysis revealed that the sensitivity for the impedance peak differed markedly from the sensitivity calculated for formants.

4. SUMMARY AND CONCLUSION

To investigate the mechanisms underlying the production of whistling, the morphological and acoustic characteristics of the human vocal tract were examined. First, the vocal tract of one human participant was scanned in three dimensions using MRI, and the cross-sectional area

function was determined from images during the production of four musical notes. The images and area function data revealed that the tongue constricted the vocal tract in different ways depending on the note produced. The area function data were also used to estimate the acoustic characteristics of the vocal tract. The main findings obtained in this study can be summarized as follows.

(1) Scanned images showed that the tongue was located towards the back of the tract during the production of F5 and B♭5, and was located more anteriorly for F6 and B♭6. Vocal-tract constriction was observed between the tongue and soft palate when the subject produced notes F5 and B♭5. While producing notes F6 and B♭6, a narrow channel was formed between the tongue and the hard palate.

(2) Quantitative cross-sectional area function data revealed the vocal-tract geometry, demonstrating that vocal-tract constriction clearly separated the tract into anterior and posterior cavities during the production of F5 and B♭5. However, a clear two-cavity separation was not produced for F6 and B♭6. Thus, the results of the morphological analysis did not support the hypothesis of Wilson *et al.* [8].

(3) In the acoustic analysis of the vocal tract, the input impedance observed from the lips was calculated from the cross-sectional area function data. The results revealed that the 2nd peak of input impedance was largely in accord with the blowing pitch, and the frequency error was less than 10% for each blowing pitch when the coupling between the vocal tract and the subglottal cavity was taken into consideration. For B♭5 only, an additional peak was created by the coupling, and the 3rd impedance peak then corresponded to the blowing pitch.

(4) A sensitivity analysis was performed to examine changes in the frequency of the 2nd impedance peak when the cross-sectional area of the vocal tract was changed at each section independently. The results revealed that the vocal-tract sections corresponding to the vocal-tract constriction and the anterior cavity were highly sensitive during the production of F5 and B♭5. In contrast, for F6 and B♭6, the sensitivity value was prominent around the region directly behind the lips.

Whistling is considered an aeroacoustic sound, produced by shedding vortex rings [8]. The vocal tract is also believed to play an essential role in determining the frequency at which vortex rings are produced, and hence blowing pitch [8,9]. Importantly, the results of our acoustic analysis suggest that the input impedance of the vocal tract is a crucial acoustic factor. While the transfer function of the vocal tract is obtained as the ratio between the volume velocity at the lip end and the volume velocity at the glottis, the input impedance characterizes the vocal tract as the ratio between the acoustic pressure and the volume velocity, both located at the lip-end of the tract, so is closely related to the aeroacoustic mechanisms underlying whistling. In addition, our sensitivity analysis suggested that the peak frequency of the input impedance can be accurately adjusted by the geometry of the anterior portion of the vocal tract. The current results suggest that the tongue is the main articulator responsible for this geometric control, as observed in the imaging data.

A previous study in our lab investigated the shape of the vocal tract of a male trumpet player [7], revealing that the tongue moved forward as blowing pitch increased. For low and mid pitches (311 Hz and 466 Hz), the tongue was found to be located slightly back in the tract, to produce a constriction between the tongue and soft palate. When the pitch was high (698 Hz), the tongue was observed to be more anterior, and the vocal tract was constricted near the border of the hard and soft palates. The high note in our previous study was F5, corresponding to the lowest note in the current study. Although we tested a different subject in each of the two studies, we found that the configuration of the vocal tract was relatively similar for this note around the constriction made by the tongue (see Fig. 1 in the present study, and Fig. 1 in our previous study [7]). For the brass and woodwind instruments, the input impedance of the vocal tract seen from the lips or reeds is connected with the input impedance of the instrument bore in series [25,26]. These observations also agree with our conclusion that the input impedance of the vocal tract is a crucial factor for determining the blowing pitch of a whistle.

In the current study, we calculated the input impedance with and without the subglottal airway. The error between the 2nd peak frequency and the blowing frequency ranged from 1.4 to 5.9% for four notes when the glottis was assumed to be closed and coupling with the subglottal airway was not considered. Although the glottis is partially open during whistling, meaning that the vocal tract is acoustically connected with the subglottal airway, when we modeled this connection the frequency error slightly increased, ranging from 2.4 to 8.0%. We suspect that insufficient spatial resolution of the MRI scanner caused the error we observed in the estimated impedance peak frequency. In addition, the area function data for the subglottal airway was taken from the literature [21], which may have degraded our quantitative data.

The present study is rather preliminary, since it examined a single subject. However, the obtained results proved that the methodology used was effective for investigating the relationship between the geometry and acoustic properties of the vocal tract, thus shedding light on the true nature of the acoustic mechanism of whistling. Future studies examining a larger number of participants and a wider range of vocal-tract shapes for various musical notes may be helpful for confirming and generalizing the

present findings. In addition to the implications of the current results with regard to the vocal tract during whistling, this study demonstrated the usefulness of a complex sensitivity function for analyzing the acoustic characteristics of the vocal tract. In future work, we plan to use this technique to further examine the role of the vocal tract in the use of brass and woodwind instruments.

ACKNOWLEDGMENT

This research was party supported by JSPS KAKENHI Grant Number 16K00242.

REFERENCES

[1] G. Fant, *Acoustic Theory of Speech Production with Calculations Based on X-ray Studies of Russian Articulations*, 2nd printing (Mouton, The Hague, 1970).
[2] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, 2nd Ed. (Springer Verlag, New York, 1972).
[3] I. T. Tokuda, J. Horáček, J. G. Švec and H. Herzel, "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," *J. Acoust. Soc. Am.*, **122**, 519–531 (2007).
[4] I. R. Titze, "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.*, **123**, 2733–2749 (2008).
[5] T. Kaburagi, "Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling," *J. Acoust. Soc. Am.*, **129**, 1554–1567 (2011).
[6] M. E. McIntyre, R. T. Schumacher and J. Woodhouse, "On the oscillations of musical instruments," *J. Acoust. Soc. Am.*, **74**, 1325–1345 (1983).
[7] T. Kaburagi, N. Yamada, S. Fukui and E. Minamiya, "A methodological and preliminary study on the acoustic effect of a trumpet player's vocal tract," *J. Acoust. Soc. Am.*, **130**, 536–545 (2011).
[8] T. A. Wilson, G. S. Beavers, M. A. DeCoster, D. K. Holger and M. D. Regenfuss, "Experiments on the fluid mechanism of whistling," *J. Acoust. Soc. Am.*, **50**, 366–372 (1971).
[9] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 2, p. 224.
[10] M. Rokkaku, K. Hashimoto, S. Imaizumi, S. Niimi and S. Kiritani, "Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique," *Ann. Bull. RILP*, **20**, 47–54 (1986).
[11] T. Baer, J. C. Gore, L. C. Gracco and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J. Acoust. Soc. Am.*, **90**, 799–828 (1991).
[12] B. H. Story, I. R. Titze and E. A. Hoffman, "Vocal tract area function from magnetic resonance imaging," *J. Acoust. Soc. Am.*, **100**, 537–554 (1996).
[13] H. Takemoto, T. Kitamura, H. Nishimoto and K. Honda, "A method of tooth superposition on MRI data for accurate measurement of vocal tract shape and dimensions," *Acoust. Sci. & Tech.*, **25**, 468–474 (2004).
[14] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust. Speech Signal Process.*, **35**, 955–967 (1987).
[15] M. Greenspan, "Simple derivation of the Boltzmann-Ehrenfest adiabatic principle," *J. Acoust. Soc. Am.*, **27**, 34–35 (1955).
[16] G. Fant and S. Pauli, "Spatial characteristics of vocal tract resonance modes," *Speech Comm. Sem. Stockholm*, pp. 1–3 (1974).
[17] S. Adachi, H. Takemoto, T. Kitamura, P. Mokhtari and K. Honda, "Vocal tract length perturbation and its application to male-female vocal tract shape conversion," *J. Acoust. Soc. Am.*, **121**, 3874–3885 (2007).
[18] B. H. Story, "Technique for tuning vocal tract area functions based on acoustic sensitivity functions," *J. Acoust. Soc. Am.*, **119**, 715–718 (2006).
[19] T. Kaburagi, "A method for estimating vocal-tract shape from a target speech spectrum," *Acoust. Sci. & Tech.*, **36**, 428–437 (2015).
[20] W. Kreuzer and C. H. Kasess, "Tuning of vocal tract model parameters for nasals using sensitivity functions," *J. Acoust. Soc. Am.*, **137**, 1021–1031 (2015).
[21] E. R. Weibel, *Morphometry of the Human Lung* (Springer-Verlag, Berlin, 1963), pp. 136–140.
[22] K. Ishizaka, M. Matsudaira and T. Kaneko, "Input acoustic-impedance measurement of the subglottal system," *J. Acoust. Soc. Am.*, **60**, 190–197 (1976).
[23] van den Berg, "An electrical analogue of the trachea, lungs and tissues," *Acta Physiol. Pharmacol. Neerl*, **9**, 361–385 (1960).
[24] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, 2nd Ed. (Springer Verlag, New York, 1972), pp. 36–38.
[25] J. Backus, "The effect of the player's vocal tract on woodwind instrument tone," *J. Acoust. Soc. Am.*, **78**, 17–20 (1985).
[26] J. Wolfe, M. Garnier and J. Smith, "Vocal tract resonances in speech, singing, and playing musical instruments," *HFSP J.*, **3**, 6–23 (2009).

APPENDIX

Cross-sectional area of the vocal tract obtained from our MRI measurement is shown below in cm^2. The glottal end is at section 1 and the lip end at section 40. *L* indicates the length of each section in cm.

Section	F5	B♭5	F6	B♭6	Section	F5	B♭5	F6	B♭6
1	0.80	1.01	0.56	0.77	21	0.44	0.81	3.08	4.76
2	0.93	0.91	0.62	0.88	22	0.40	0.76	2.67	4.12
3	0.82	0.83	0.86	1.02	23	0.37	0.87	2.44	3.70
4	1.02	0.82	1.00	1.16	24	0.42	0.85	2.18	3.25
5	1.32	1.01	0.85	1.34	25	0.37	0.68	1.95	2.70
6	1.57	1.28	0.85	2.18	26	0.33	0.43	1.76	1.87
7	1.81	1.43	1.35	3.41	27	0.54	0.51	1.50	1.35
8	1.75	1.53	2.25	3.66	28	0.82	0.58	1.29	0.99
9	1.62	1.88	2.59	3.37	29	1.09	0.67	1.20	0.71
10	2.33	2.88	2.56	3.63	30	1.48	0.85	1.09	0.55
11	3.72	4.02	3.09	4.52	31	1.70	0.99	1.03	0.48
12	4.62	4.66	3.98	5.16	32	1.76	1.00	0.96	0.48
13	4.66	4.71	4.71	5.75	33	1.78	0.98	0.93	0.46
14	4.11	4.26	5.19	6.23	34	2.00	1.25	0.86	0.48
15	3.43	3.85	5.27	6.39	35	2.25	1.60	0.84	0.47
16	2.97	3.26	5.35	6.19	36	2.36	1.83	0.86	0.44
17	2.57	2.67	5.03	6.01	37	1.97	1.76	1.15	0.30
18	2.00	2.00	4.49	5.99	38	1.41	1.35	1.41	0.24
19	1.33	1.34	3.93	5.87	39	0.83	0.77	0.92	0.22
20	0.75	1.06	3.38	5.37	40	0.40	0.32	0.30	0.12
					L	0.43	0.42	0.41	0.40