An interactive construction system of 3–D vocal tract shapes from tomograms

Tohru Yokoyama, Nobuhiro Miki, and Yoshihiko Ogawa

Graduate School of Engineering, Hokkaido University, Sapporo 060, Japan

Abstract: A new algorithm is proposed to construct the 3-D vocal tract shape from tomograms measured by MRI or X-ray CT, etc. for the purpose of modeling the speech production process in the human vocal tract including branching parts. An interactive system realizing the algorithm on the X-Window environment is developed and the re-constructed vocal tract shape is shown.

INTRODUCTION

The influence of physical shapes of human vocal tracts on speech production was investigated. Acquiring the 3-D geometrical information of the vocal tract during phonation is one of the most significant challenges for modeling the speech production mechanism(1,2). Tomographs, such as MRI or X-ray CT are powerful tools to obtain the steady-state geometry of the vocal tract with high resolution. Because of the complicated shape of the human vocal tract, however, it is difficult to get the 3–D surface figure from the slice images of tomogram. In this paper, an algorithm is proposed to construct the 3–D vocal tract shape with many triangular surfaces from 2–D tomogram images, and the interactive system realizing the algorithm is developed. On this system offering user-friendly environment with GUI on X-Window, a complicated vocal tract shape including branching parts such as the piriform sinus can be efficiently constructed with sufficient quality.

RECONSTRUCTION ALGORITHM OF 3–D VOCAL TRACT SHAPES

A diagram of the whole reconstruction algorithm of the 3-D surface shape of the vocal tract using tomogram images is shown in FIGURE 1. The overall process consists of three parts: selecting the tomogram images from the whole of the acquired images and allocating them into the 3–D space, extracting the air tract contained in the tomogram images, and constructing a wire-frame model of the vocal tract from the contours extracted.

As shown in FIGURE 1, the extraction of the contour on the selected image requires three operations beforehand: a noise reduction, a contrast enhancement, and a threshold process. The noise reduction is performed with $3 \times 3$ median filtering. Then pixel values of each image are transformed with a contrast enhancement process. Then a threshold process is performed by setting a relevant threshold value for transformation to only binary values from the tomogram image. After these three fundamental processes, the extraction of

\[
\begin{align*}
\sum d(i, j) \rightarrow \min, & \\
\text{where } d(i, j) = & \\
& d(P_i, P_{i+1}, Q_j) = w_d d_p(i, j) + w_a d_a(i, j) + w_n n(i, j) \\
& \text{OR} \\
& d(P_i, Q_j, Q_{j+1}) = w_d d_p(i, j) + w_a d_a(i, j) + w_n n(i, j)
\end{align*}
\]

FIGURE 1. A diagram of the reconstruction process.

FIGURE 2. The algorithm of wire-frame modeling between each pair of successive contours.
contours as a surface boundary of the vocal tract is performed by pointing inside of closed-boundaries; by
tracing the boundary between black and white pixels one after another.

A wire-frame model of the whole 3-D vocal tract shape is constructed by forming many triangular
facets between each two successive contours allocated in the 3-D space. The main idea is to determine
a combination of triangular facets which minimizes the total cost of criterions: each criterion
\[ d(i, j) = w_i \phi(i, j) + w_j \psi(i, j) + w_k \kappa(i, j) \]
is defined so as to represent the fitness of each triangular facet as an element
of the surface (shown in FIGURE 2). By computing the cost of criterions for all the possible triangular facets
between the two contours and searching the best combination of the facets which minimize the total cost
of criterions, the surface of the two successive contours is determined automatically. The branching problem
can be solved by dividing it into problems of two successive contours.

**INTERACTIVE SYSTEM FOR 3-D VOCAL TRACT RECONSTRUCTION**

A prototype of the interactive system is developed to realize the reconstruction algorithm proposed
above. This system is programmed with C++ language, with Motif for graphical user interface, and with
PEX for drawing 3-D shapes, and runs on the X-Window environment on a personal computer. FIGURE 3
shows an example of display screens of the interactive system. This system offers user-friendly interfaces to
easily extract a vocal tract contour from a selected tomogram image and to compute a set of triangular facets
between two successive contours. FIGURE 4 shows an example of the 3-D vocal tract shape reconstructed
on this system using MRI images. This result shows a natural surface of the complicated shape and agrees
with anatomical information of the vocal tract.

**FIGURE 3.** Display screens of the interactive system for constructing 3-D vocal tract shape.

**FIGURE 4.** The wire-frame model of the vocal tract shape (vowel /a/).

**CONCLUSION**

A new algorithm is proposed to construct the 3-D vocal tract shape from tomogram images. And a
prototype of an interactive system realizing the algorithm is developed. Showing an example of the vocal tract
shape reconstructed using the system, we showed the usefulness of our interactive system. Parameterization
of the 3-D geometrical information of vocal tracts should be implemented as a further work.

**ACKNOWLEDGMENTS**

The authors would like to thank Dr. Shinobu Masaki of ATR HIP Research Laboratories for giving us
an opportunity to acquire the MRI measurement data.

**REFERENCES**

2. Yang C.-S. and Kasuya H., "Dimensional differences in the vocal tract shapes measured from MR images across

1284