Ann.Bull.RILP(1984)No.18, 1-4

AN X-RAY VIDEO-IMAGE ANALYSIS SYSTEM

Shigeru Kiritani, Akira Kawagoe* and Kiyoshi Hashimoto*

It has been generally the case that computer processing of the picture data requires elaborate large scale devices. However, in recent years, digitization and the computer storing of the video images has become very simple. In order to exploit this advantage for video images, we have converted several old x-ray motion pictures of articulation into video images and developed a system of the computer processing of video images for motion analysis.

The video image input system

In the present study a VAX 11/780 computer was used for the image processing. An image memory was connected to it via a DMA channel. The image memory can store the video image of a single frame in real time as 768(horizontal) x 512(vertical) pixels of data with an 8 bit gray level. In order to input to the computer a sequence of video images through an ordinary video tape recorder (i.e., without using a special device such as a video disk), following method was devised.

It is possible, in principle, to input to the computer the image of a selected frame by the stop motion reproduction of the selected frame. However, the use of stop motion generally incurs a degradation in the image quality. It is also impractical to locate individual frames by manual operation during the input of a sequence of images. In order to avoid these problems, a trigger pulse was recorded on an audio channel of the video tape at a selected time preceding the required portion of the video tape. The signal was led to one channel of the digital input of the computer. The frame pulse separated from the video signal was also led to another channel of the digital input.

The program for storing the video image in the computer remotely controls the video tape recorder and starts it. The program waits for the trigger pulse and, upon detecting the trigger pulse, starts counting the frame pulse. When a specified number of pulses is counted, the program sends out a command to the image memory to digitize and store a single frame of the video image. Then, the program rewinds the video tape recorder to the initial position, increases the number of frame pulses to be counted by one and repeats the above procedure until required number of images are stored.

* Electro Communication University

Measurement of the jaw position

As a preliminary step toward the automatic measurement of articulatory movements, an automatic identification of the position of the mandible was carried out. Identification of the position of the mandible was performed by matching to the individual image the template contour of the mandible around the mental process. The template contour was determined by tracing manually the edge of the mandible on the averaged image of several selected frames.

To measure the position of the mandible in a given image, the template contour was superimposed on to the image, and the sum of the absolute value of the gradient of the gray level at each pixel on the template contour was calculated. The calculation was carried out for various positions of the template contour within a selected range of x-and y-displacements and rotation, and the position which gave the maximum value for the above sum was determined.

In order to measure the movement of the mandible relative to the upper jaw, it is necessary to measure the movement of the head simultaneously. For this purpose, the contour of the anterior edge of the palatal bone and the nasal floor (which was enhanced in the present x-ray images by radio opaque material) was used. These template contours are shown in Fig. 1.

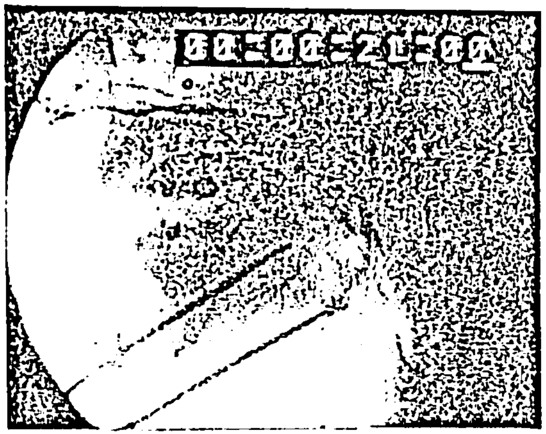
Fig. 2 shows the position of the mandible identified for the five images, each image being selected from the period of the production of one of the five Japanese vowels /i/, /e/, /a/, /o/ and /u/.

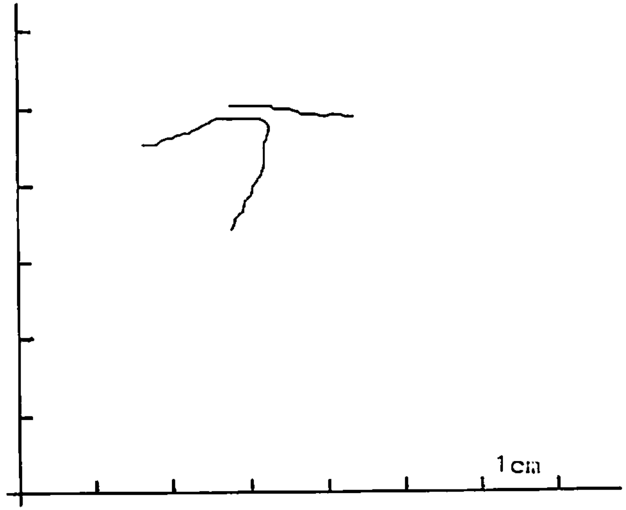
Fig. 3 shows a result of the automatic tracking of the movement of the mandible for the selected utterances. In this case, the range of the search of the mandible position in the initial frame was specified manually, and for the subsequent frames, the search was made within a pre-specified range around the position detected in the previous frame.

These results show that the present system offers a feasible means of automatically processing video images for motion analysis. Further study is being conducted on the automatic identification of tongue positions.

Acknowledgement

The test material for the x-ray images in the present study were selected frames from the x-ray motion picture "Production of Japanese sounds" produced by The National Language Research Insitute in 1967. We wish to acknowledge the courtesy of Prof. Y. Uemura and Dr. A. Takada in permitting the use of these x-ray images.





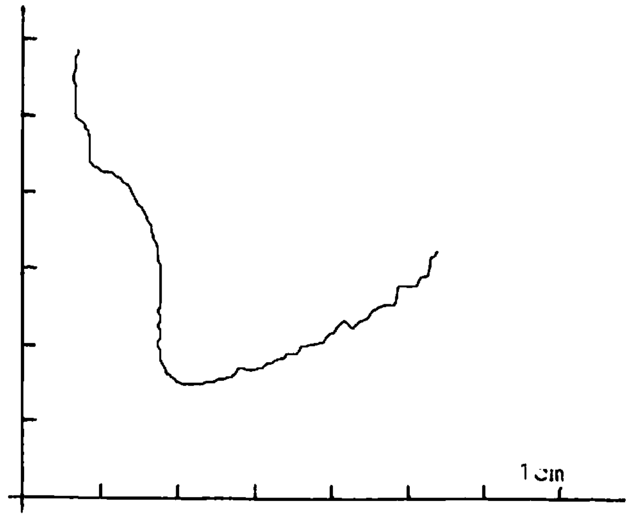


Fig. 1 An example of an x-ray image (above) and template contours of the mandible and the upper jaw (right).

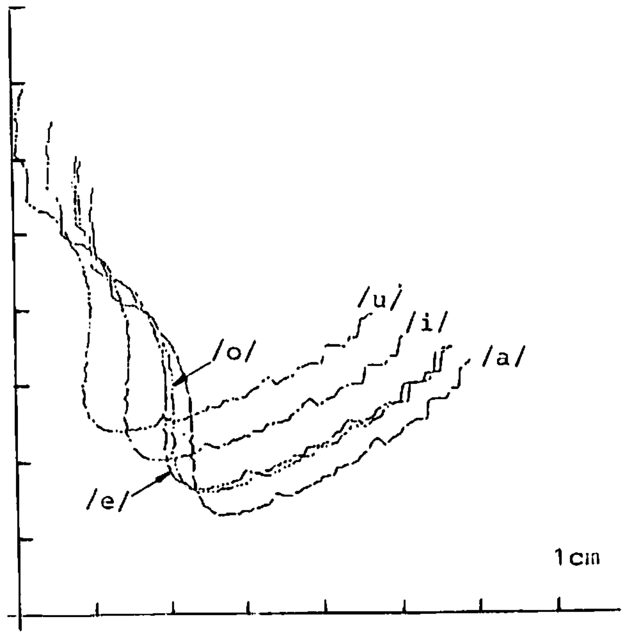


Fig. 2 The position of the mandible identified for five frames during the production of the five Japanese vowels.

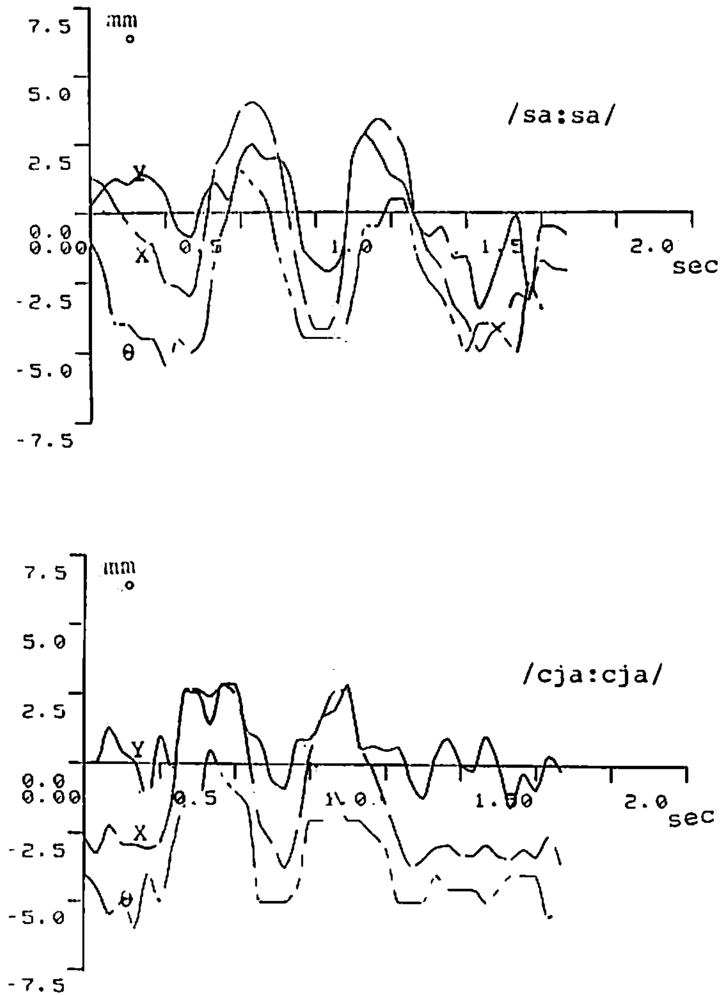


Fig. 3 Examples of the automatic tracking of the mandible position for selected utterances.