JSLHR, Volume 40, 1118–1121, October 1997

A System for Three-Dimensional Visualization of Human Jaw Motion in Speech

RESEARCH NOTE

Thierry Guiard-Marigny
Institut de la Communication Parlée
Grenoble, France

David J. Ostry
McGill University
Montreal, Quebec, Canada

With the development of precise three-dimensional motion measurement systems and powerful computers for three-dimensional graphical visualization, it is possible to record and fully reconstruct human jaw motion. In this paper, we describe a visualization system for displaying three-dimensional jaw movements in speech. The system is designed to take as input jaw motion data obtained from one or multi-dimensional recording systems. In the present application, kinematic records of jaw motion were recorded using an optoelectronic measurement system (Optotrak). The corresponding speech signal was recorded using an analog input channel. The three orientation angles and three positions that describe the motion of the jaw as a rigid skeletal structure were derived from the empirical measurements. These six kinematic variables, which in mechanical terms account fully for jaw motion kinematics, act as inputs that drive a real-time three-dimensional animation of a skeletal jaw and upper skull. The visualization software enables the user to view jaw motion from any orientation and to change the viewpoint during the course of an utterance. Selected portions of an utterance may be replayed and the speed of the visual display may be varied. The user may also display, along with the audio track, individual kinematic degrees of freedom or several degrees of freedom in combination. The system is presently being used as an educational tool and for research into audio-visual speech recognition. Interested researchers may obtain the software and source code free of charge from the authors.

KEY WORDS: visualization, jaw, speech production, motor control, kinematics

The motions of the orofacial structures in behaviors such as speech, mastication, and swallowing are as complex as any movements controlled by the human nervous system. Understanding the control and coordination of these motions is difficult, in part, because of the inaccessibility of these structures and, in part, because of the large number of articulators—each of which is capable of motion in multiple degrees of freedom. Moreover, because articulators such as the lips, tongue, and velum are formed wholly of soft tissue, these structures have an infinite number of mechanical degrees of freedom (see Goldstein, 1950) and, functionally, each is capable of several distinct motions (Harshman, Ladefoged, & Goldstein, 1977; Maeda, 1988; Sanguineti, Laboissière, & Payan, in press).

The development in recent years of precise two- and three-dimensional motion measurement systems and scientific workstations for three-dimensional visualization provides a useful combination of tools for the exploration of these complex motions. In this paper, we describe "Jaw

 1092-4388/97/4005-1118

Viewer," a system that combines three-dimensional motion measurement with three-dimensional visualization in order to reproduce human jaw motions in speech and other behaviors.

In the present paper, we describe the use of the Jaw Viewer software based on data provided by the Optotrak measurement system. However, the Jaw Viewer system may be used in conjunction with any jaw motion measurement system, even those with far more restricted data-capture capabilities than Optotrak. The software runs on all Silicon Graphics workstations. Since these are widely available at research facilities, we believe Jaw Viewer may provide a useful and general purpose data exploration tool for the speech research community. The software and associated source code for the Jaw Viewer program are available from the authors free of charge. (Information concerning data format and all instructions for running the program are provided in an easy-to-use online manual that comes as part of the package.)

Thus far, the Jaw Viewer system has been used as an instructional aid and for research in audio-visual speech recognition (Guiard-Marigny, Ostry, & Benoit, 1995). However, the system may prove to have more general applicability. Three-dimensional visualization systems like Jaw Viewer capitalize on the nervous system's pattern- recognition capabilities. They provide a resource for kinematic data that is not unlike that found in electrophysiological research where neural spike trains may be transformed into auditory signals in order to render characteristics of the time-varying pattern more easily recognizable. In this regard, the tool described here may aid in the exploration of complex orofacial patterns in both research and clinical settings.

Visualization of Jaw Motion

The jaw is a rigid skeletal structure whose motion is characterized by three orientation angles and three positions (see Vatikiotis-Bateson & Ostry, 1995, for details). Although most research on jaw motion in speech has focused on either rotation of the jaw in the midsagittal plane or upon the motion of individual tissue points, jaw movements are considerably more complex and, accordingly, our understanding of their organization may be aided by the ability to visualize complete patterns of motion. The need for these visualization tools is underscored by recent kinematic evidence (Ostry, Vatikiotis-Bateson, & Gribble, in press) that suggests that jaw movement in speech involves independent motion in four degrees of freedom—sagittal plane rotation (pitch), sagittal plane horizontal translation (protrusion/retraction), sagittal plane vertical translation, and yaw (coronal plane rotation). Jaw motions in mastication are likewise characterized by kinematic independence in these same four degrees of freedom.

The Jaw Viewer graphical display software allows the user to visualize jaw movements in the six kinematic dimensions that in mechanical terms account fully for jaw motion kinematics. The user is presented with a real-time three-dimensional computer animation of a skeletal jaw and upper skull that may be driven using empirical data, data derived from jaw motion mathematical simulations (Laboissière, Ostry, & Feldman, 1996) or controlled through the use of sliders. The user is able to view jaw motions from any orientation and to change viewpoints in the course of an utterance. Selected portions of an utterance or a complete utterance may be replayed and the speed of the visual display may be varied.

The motion of the jaw in its component six degrees of freedom may be explored using a "slider box" (Figure 1). This window contains six sliders, each corresponding to one of the six degrees of freedom of jaw motion. The user can visualize changes in each degree of freedom by changing the slider position or by typing a desired value for a specific degree of freedom.

The three-dimensional animation of the jaw and skull is typically presented along with a second graphics window called the "Curve Viewer" (Figure 2). The Curve Viewer displays time-varying kinematic and audio signals and provides the user with a menu bar and button boxes that control the display. The viewer can be used to display one or more data streams, to zoom in on parts of the signal, to select a zone for audiovisual playback, and to obtain numerical values in the data.

The kinematic records of jaw motion that drive the Jaw Viewer animation shown in Figure 2 have been recorded using Optotrak—an optoelectronic measurement system. However, other data may also be used, for example, jaw motion data derived from simulations (Laboissière, Ostry, & Feldman, 1996), hypothetical jaw motion trajectories, and, in particular, data obtained using other motion measurement systems that record jaw movement in one or more degrees of freedom. The Optotrak system used here records the motion of infrared emitting diodes (IREDs) that are attached to the jaw and the head. When using Optotrak, six IREDs are attached to a head-mounted plastic frame. The motion of these markers is used to correct for head motion by transforming the data to a head-based frame of reference. Four additional IREDs are used to track jaw motions. These are glued to a lightweight acrylic and metal dental appliance that is attached to the mandibular teeth using a dental adhesive. The appliance is seated bilaterally and is custom-molded for each subject to fit the contour of the buccal surface of the teeth.

Jaw motion data may be visualized with the Curve Viewer and simultaneously animated in the Jaw Viewer graphics window. The Curve Viewer offers a number of

JSLHR, Volume 40, 1118–1121, October 1997

Figure 1. Jaw Viewer graphical display and slider window.

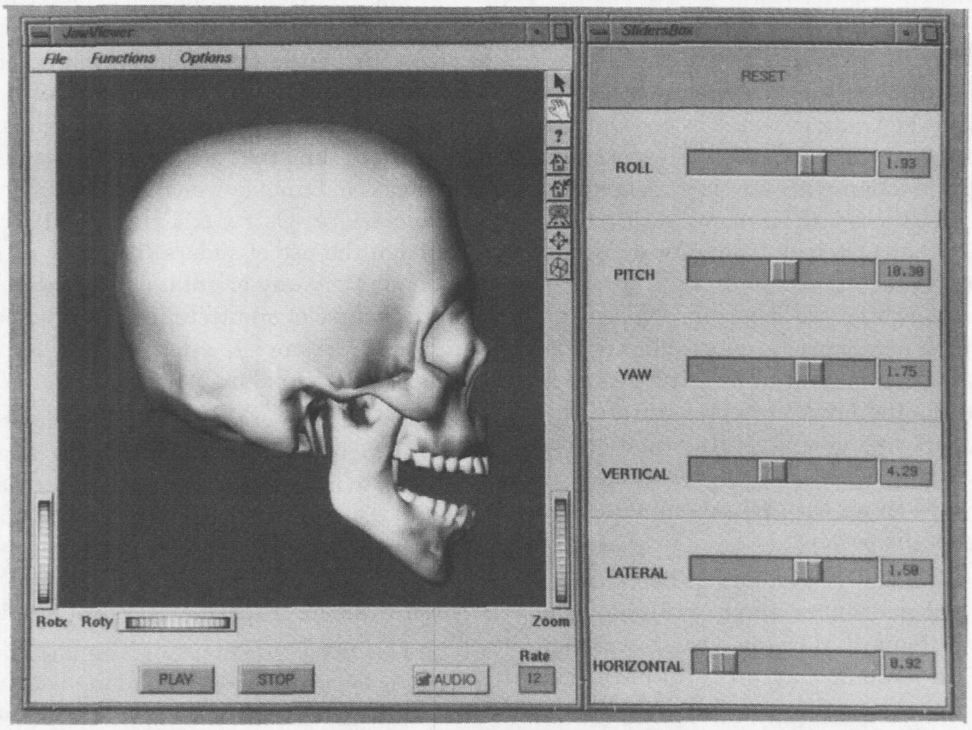
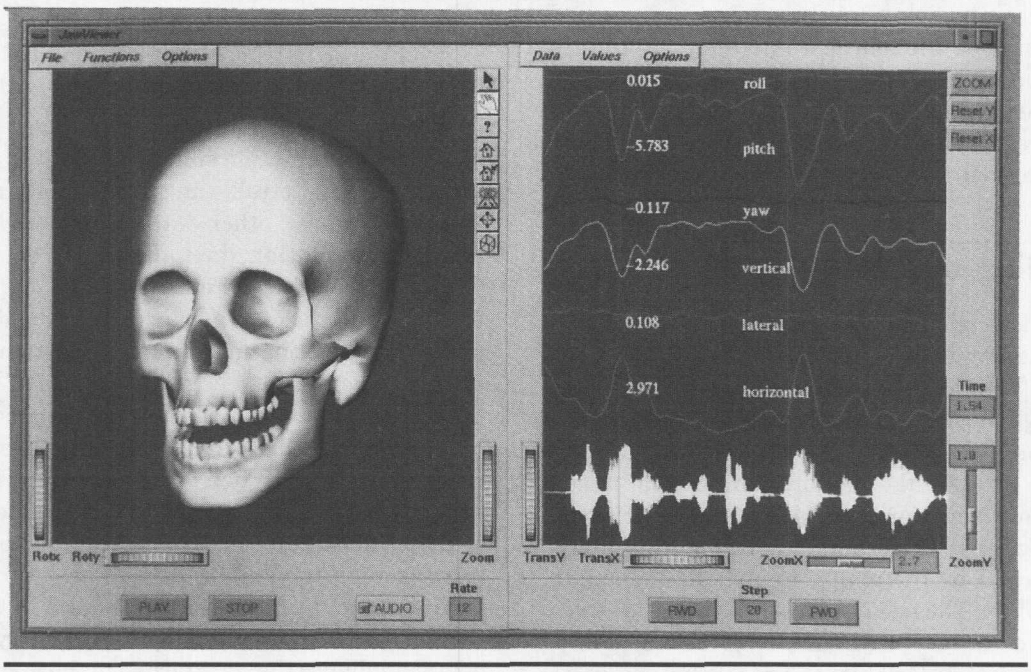


Figure 2. Jaw Viewer graphical display and Curve Viewer window.



features for data exploration. The viewer uses two cursors. One allows the user to select a point on the displayed data streams and view the resulting static jaw position. By dragging the cursor along the data stream, jaw position may be visualized over time. A second cursor is used to select the end of a region either to zoom in on or for playback. In the playback mode, jaw motion and speech acoustics are presented synchronously in real-time.

A menu bar in the Curve Viewer allows the user to select the kinematic degrees of freedom that are to be displayed. Only the selected data streams are used in the animation and data streams may be displayed either alone or in combination with others. This is a particularly useful feature as it allows the user to visualize the effect of either individual data streams—corresponding to individual degrees of freedom—or the form of the movement with a particular stream removed.

Although Jaw Viewer is a powerful aid to scientific visualization, its editing and output capabilities are limited to selecting regions within the dataset for playback or zoom and to the visual display of the numerical magnitudes of points in data streams. In this regard, quantitative examination of jaw motion data is left to other specialized software designed for analysis rather than visualization purposes.

Jaw Viewer is written in C++ and so has a modular, object-oriented format. The graphical interface is written in Motif, a library that is widely used by windowing applications. The three-dimensional graphics are based on the Open Inventor library, which provides an easy way to manipulate, animate, and render three-dimensional objects such as the jaw and upper skull. An online manual prepared in Show Case accompanies the Jaw Viewer program.

The jaw and skull used in Jaw Viewer program were purchased from Viewpoint DataLabs. The skull has approximately 5600 polygons. The video playback rate is dependent upon the computer, but it is automatically adjusted to maintain synchrony between the audio and video displays. This permits the program to be used on various machines with different hardware capabilities. The audio playback rate is similarly dependent on the audio card used in the workstation. Audio rates range from 8 kHz to rates compatible with CDs and DAT tapes.

The audio signals are managed with the use of two Silicon Graphics audio libraries. One library manipulates the audio files and provides tools to deal with properties such as format and sampling rate. The format for this application is AIFF (Audio Interchange File Format), a format used both by Apple and SGI. A second library is used to manage the audio port and produce audible signals. This library also provides functions that synchronize the audio signal and the jaw animation.

As a tool for the presentation of complex patterns of three-dimensional motion, Jaw Viewer has a number of potentially valuable applications in the fields of orofacial research. As noted above, the program has thus far been used both as an instructional aid and as a research tool to assess the contribution of jaw motion to audiovisual speech perception (Guiard-Marigny, Ostry, & Benoit, 1995). In the latter context, it was found that the intelligibility of speech in noise improved when jaw motion animated using Jaw Viewer was combined with a visual display of lip movement. However, we believe the program may have more general applicability as an aid to scientific visualization. Because it provides an integrated visual reconstruction of complex time-varying motion patterns, it may prove useful as an exploratory tool, for example, in identifying different modes of coordinated motion and also in assessing changes to intra-articulator coarticulation (which are not readily detectable from an examination of individual data streams). There is also the possibility of clinical applications, for example, as an assessment tool for evaluation of temporomandibular dysfunction and as a technique to aid in the evaluation of various therapeutic interventions. Moreover, thanks to its object-oriented structure, the software can be easily modified to provide a variety of new functions. For example, animation output could be recorded onto a VCR or converted to a compressed movie format such as Quicktime or MPEG.

Acknowledgments

This research was supported by NIH grant DC-00594 from the National Institute on Deafness and Other Communication Disorders, NSERC-Canada, and FCAR-Québec. The authors thank P. L. Gribble and K. G. Munhall for comments.

References

Goldstein, H. (1950). *Classical mechanics*. Addison-Wesley.

Guiard-Marigny, T., Ostry, D. J., & Benoit, C. (1995). Speech intelligibility of synthetic lips and jaw. *Proceedings of the 13th International Congress of Phonetic Sciences, 3*, 222–225.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America, 62*, 693–707.

Laboissière, R., Ostry, D. J., & Feldman, A. G. (1996). Control of multi-muscle systems: Human jaw and hyoid movements. *Biological Cybernetics, 74*, 373–384.

Maeda, S. (1988). Improved articulatory model. *Journal of the Acoustical Society of America, 81*(S1), S146.

Ostry, D. J., Vatikiotis-Bateson, E., & Gribble, P. L. (in press). An examination of the degrees of freedom of human jaw motion in speech and mastication. *Journal of Speech, Language, and Hearing Research.*

Sanguineti, V., Laboissière, R., & Payan, Y. (in press). A control model of human tongue movements in speech. *Biological Cybernetics.*

Vatikiotis-Bateson, E., & Ostry, D. J. (1995). An analysis of the dimensionality of jaw movement in speech. *Journal of Phonetics, 23*, 101–117.

Received October 16, 1996

Accepted January 31, 1997

Contact author: D. J. Ostry, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal QC Canada H3A 1B1. Email: ostry@motion.psych.mcgill.ca