SPEECH MOVEMENT RESEARCH USING THE NEW X-RAY MICROBEAM SYSTEM

ROBERT D. NADLER

Waisman Center
University of Wisconsin
Madison, WI, USA 53705

JAMES H. ABBS

Depts. of Neurology and
Neurophysiology
Waisman Center
University of Wisconsin
Madison, WI, USA 53705

OSAMU FUJIMURA

AT&T Bell Laboratories
Murray Hill, NJ, USA 07974

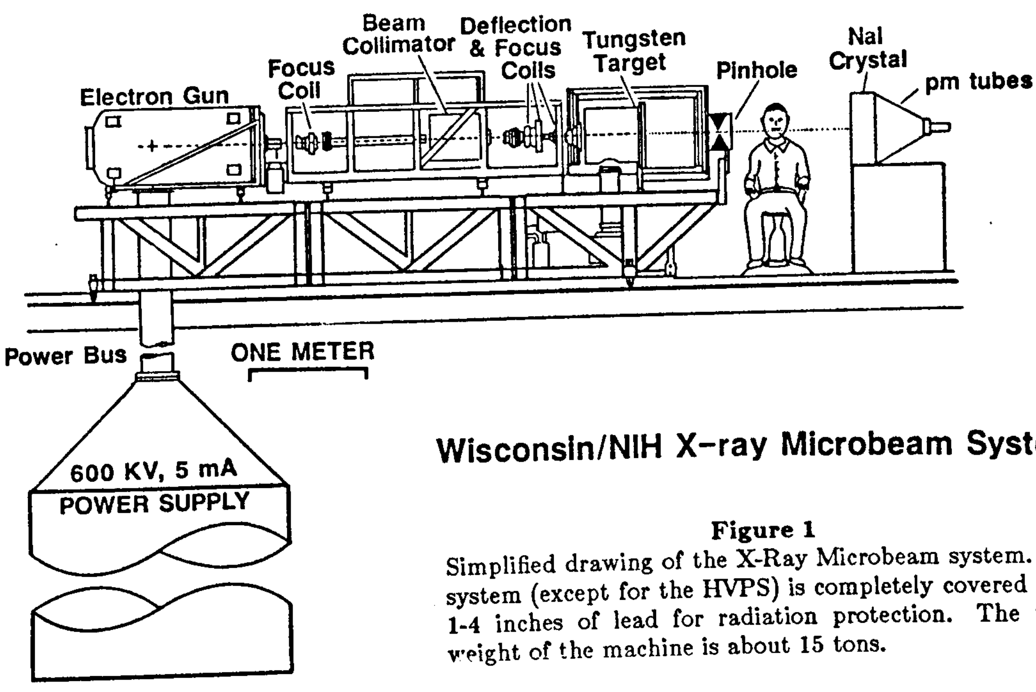
ABSTRACT

A new X-Ray Microbeam system for studying tongue movements and other articulatory gestures has been constructed to serve as the core instrument of a shared speech production research facility. Preliminary speech movement data has been obtained and this system is currently capable of tracking multiple articulatory pellets (up to 12) at aggregate sampling rates of about 1000 per second. Radiation exposures are very low due to the narrow x-ray beam and localized computer-controller scans used for tracking. The facility includes parallel capability for data display and analysis for multiple experimenters.

[Support for this facility is provided by NIH-NINCDS grant NS-16373.]

NEED FOR FACILITY

The progress of research into human speech production is severely limited by a number of factors inherent in the speech production process itself. In-depth physiological investigation of critical speech processes (e.g., neural activation, muscle activity, movement, *in vitro* biomechanics, etc.) cannot be conducted optimally because there appears to be no truly suitable animal model. Because the human speech apparatus has multiple overlapping functions (breathing, chewing, swallowing, *and* speech), functional inferences made from anatomical investigations provide only fragmentary and potentially confusing results. For example, while the masseter and temporalis muscles are capable anatomically of substantial jaw closing forces, they are generally inactive during movements of the jaw for speech production. These limitations require that many of the significant physiological issues surrounding human speech production *be addressed with human subjects under normal conditions of speech.*



Wisconsin/NIH X-ray Microbeam System

Figure 1
Simplified drawing of the X-Ray Microbeam system. The system (except for the HVPS) is completely covered with 1-4 inches of lead for radiation protection. The total weight of the machine is about 15 tons.

While speech production processes are inherently difficult to investigate, recent advances have demonstrated that the careful application of analysis and interpretation techniques adapted from systems physiology, biomedical engineering, and signal processing provide a means to exploit the limited available data. That is, in the last ten years, the methods for interpretation of multivariable speech events have become increasingly utilized by the speech physiologist.

To advance this important work, we have developed an x-ray system with increased capabilities to obtain large samples of the relevant movement, EMG, aerodynamic, and acoustic data *simultaneously*, data with which to elaborate and refine preliminary models and further test their viability. The X-Ray Microbeam system that previously existed at the University of Tokyo [1] provided preliminary studies with suggestive information demonstrating the strength of this method [2,3,4,5].

X-RAY SYSTEM

Figure 1 is a simplified drawing of the x-ray generator. Its major components include a 600 kV, 5 mA power supply, a source electron gun and accelerating column, beam-line components for electron beam focusing and deflection, a thin (900 micron) water-cooled Tungstun target for photon generation, x-ray pinhole, and NaI detector.

Sampling of the x-ray detector output, control of the beam deflection and x-ray scanning, and implementation of the pellet spherical pattern recognition/background subtraction is done with specialized digital hardware controlled by two fast microsequencers which in turn are remotely controlled by the main computer processor (VAX 11/750). All of these processors communicate through a shared four-port memory. This implementation has been designed (in particular, the 600 kV acceleration voltage) to allow us to track pellets in the presence of common tooth filling amalgams. Development of the specialized algorithms and support computational hardware to provide this capability is currently under way.

As shown in Figure 2, small (2-3 mm) gold pellets are placed, for example, on the tongue, lips, maxillary and mandibular teeth, and velum. The x-ray generator emits a narrow x-ray beam whose two-dimensional position on the object field of the subject's head is computer controlled. The x-ray beam passes through the object field and is detected by a scintillation counter. The output of the scintillation counter reflects the relative radiopacity of the image field through which it passed.

Based upon the previous pellet positions or the pellet position determined during an initial scan, the main computer transmits a set of initial X and Y coordinates to the digital scan controller. The computer system locates the position of a pellet by first obtaining the x-ray image generated using a locally restricted raster scan (as shown in Figure 2). This image, after background subtraction, is subjected to a global template recognition algorithm to determine the pellet location within the scan area. The predicted pellet position is determined using an algorithm in the main computer that utilizes previous pellet displacement, velocity, and acceleration. Each of the other pellets are scanned in turn with associated image processing to provide time-motion tracking of all the pellets.

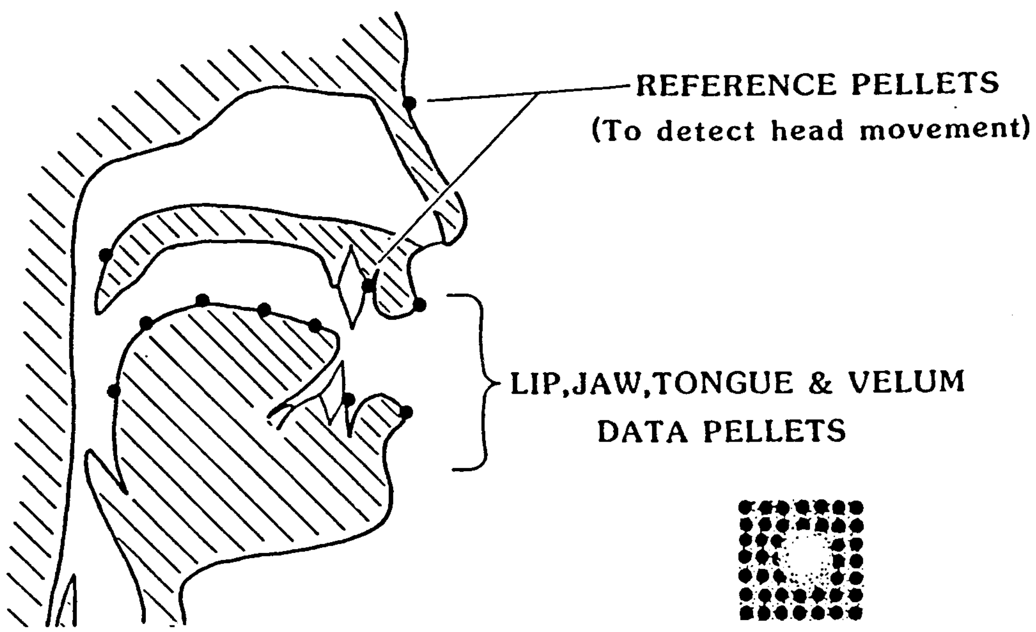


Figure 2
Mid-sagittal view of speech articulators with attached gold pellets. Small computer generated scan showing image of a gold pellet.

Computer Image
of
Microbeam Scan Pattern

Radiation dosages to the subject are limited to very low levels due to a combination of many factors. They include the relatively small size of the x-ray beam (approximately 1 mm at the subject mid-sagittal plane), the limited time (pixel exposure is 10 microseconds maximum), the beam is allowed to expose any particular area of the oral cavity, tissue not in the immediate vicinity of a pellet is not exposed, and secondary photon scatter (Compton effect) is reduced by the high energy of the primary photons (photon energies below 100 KeV are filtered out). Multiple measures and estimates indicate that the system at an acceleration voltage of 400 kV and an electron beam current of 5 mA will yield average entrance radiation exposures of 0.56 mR/minute, or a total of 8.4 mR for 15 minutes of data acquisition. Other measures, such as the peak radiation exposure for a given small volume of tissue (1.0 cm^3) in the worst case, are also very low (11 mR/minute or 165 mR for 15 minutes of data). These exposure levels compare very favorably with exposure levels for other clinical and experimental procedures. For example, a single dental bitewing yields an entrance exposure of 650 mR while the same 15 minutes of speech data with cineradiography would require a prohibitive entrance exposure of 7.5 R. Given these measures and comparisons, we are confident that the radiation exposure as a result of this procedure will offer negligible risk if appropriate precautions are observed.

ACQUISITION OF OTHER SIGNALS

Instrumentation associated with this facility also allows for simultaneous detection/transduction and conditioning for the following speech production parameters: (1) the airborne speech acoustic signal, (2) an accelerometer/throat microphone signal, (3) up to four channels of aerodynamic signals or signals from other strain gage transducers and (4) up to ten channels of electromyography (EMG). A custom designed A/D subsystem which acquires this analog data at aggregate rates up to 125K samples/second (15 bits, isolated) has been built to provide this capability. This implementation provides for differential sampling across up to 64 channels that is time-synchronized with the x-ray system pellet movement data. A multi-channel D/A subsystem is also provided for audio playback of speech acoustic data as well as analog output of physiological data.

NETWORKING AND ANALYSIS CAPABILITIES

A general purpose networking system (consisting of Ethernet and Pronet local area networks) provides high bandwidth inter-computer communication capabilities for both data acquisition and analysis functions. Each processor on the network (data analysis graphics workstation or data acquisition CPU) has shared access to the central file servers. SUN graphics workstations provide a bit-mapped 1024 X 1024 monochrome display and are well suited for manipulation of multi-channel physiological and acoustic data. A custom designed data base has

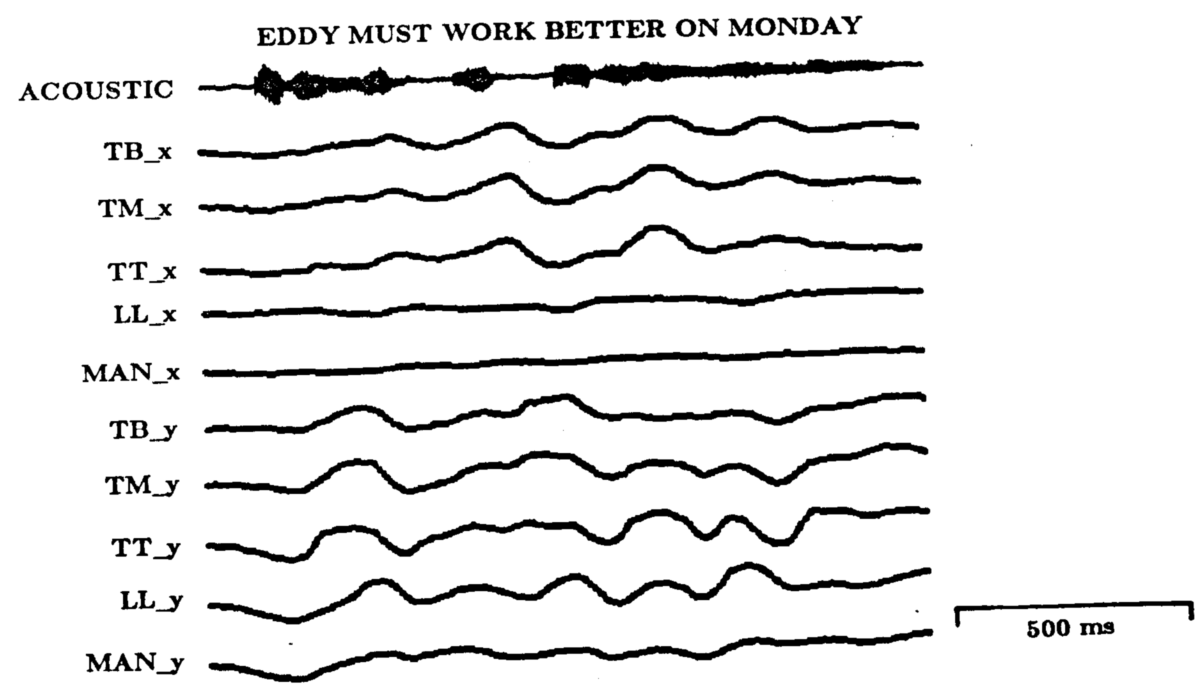


Figure 3
Simultaneously acquired acoustic speech signal and X- and Y-coordinate data from 5 pellets tracked with the x-ray microbeam system. See text for details.

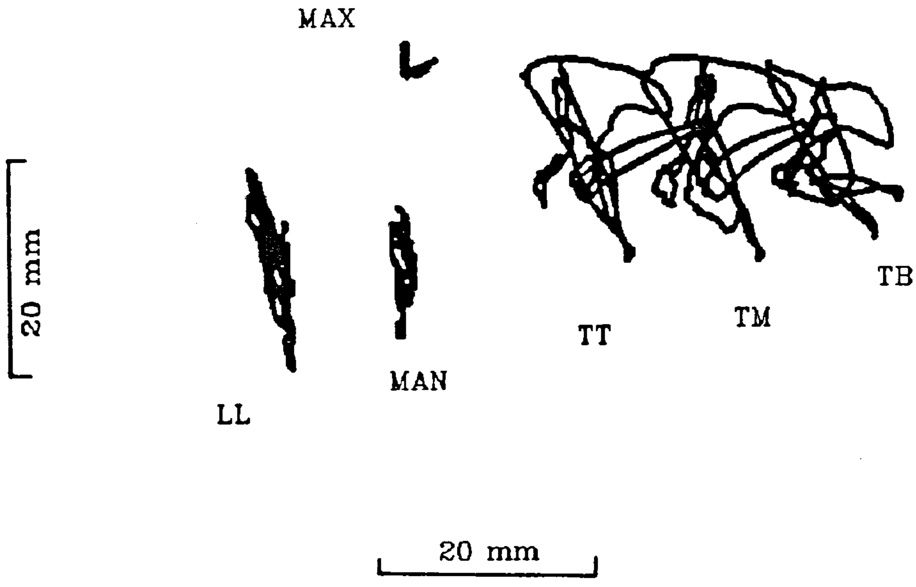


Figure 4
Cartesian coordinate plot of the same data from Figure 3. A head reference pellet is also included.

been constructed with binary descriptive and data files designed for optimum storage efficiency and access. In addition to the general windowed/mouse environment provided by the SUN workstation, graphics applications have been developed specifically for multiple data signal display, manipulation, and analysis. In addition, the 'S' statistical package (licensed from AT&T Bell Laboratories) provides an interactive computing environment and statistical data analysis language as well as a wide variety of specialized graphics capabilities.

EXPERIMENTAL RESULTS

Figures 3 and 4 present an example of typical data acquired using the x-ray microbeam system. This experiment used three tongue (TB, TM, and TT), a mandible (MAN), a lower lip (LL), and two maxillary reference pellets (MAX). Each of the data pellets was acquired at 100 samples/second, the reference pellets at 50 samples/second, along with a single channel of speech acoustic data at 10,000 samples/second. These data have not been digitally filtered or corrected for head movement. Normally, at least two head reference pellets are sampled so that articulatory pellet movement data can be corrected for head movement (translation and rotation in the mid-sagittal plane) prior to data analysis.

REFERENCES

[1] Fujimura, O., Kiritani, S., and Ishida, H. Computer controlled radiography for observation of movements of articulatory and other human organs. *Comput. Biol. Med.* 1973, *3*, 371-384.

[2] Kiritani, S. Articulatory studies by the X-ray microbeam system. In M. Sawashima and F. S. Cooper (Eds.), *Dynamic aspects of speech production.* Tokyo: University of Tokyo Press, 1977, 171-194.

[3] Harris, K. S., Tuller, B., Bell-Berti, F., and Kelso, J. A. S. *Some temporal rules in speech production.* Paper presented to the 105th meeting of the Acoustical Society of America, Cincinnati, Ohio, May 1983.

[4] Hirose, H., Kiritani, S., and Sawashima, M. Patterns of dysarthric movements in patients with amyotrophic lateral sclerosis and pseudobulbar palsy. *University of Tokyo Annual Bulletin,* 1980, *14*, 263-272.

[5] Fujimura, O. Relative Invariance of Articulatory Movements: An Iceberg Model. In J. S. Perkell and D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes,* London, Lawrence Erlbaum Ass., 1986, 226-234.