A head and transducer support system for making ultrasound images of tongue/jaw movement

Maureen Stone
Department of Electrical and Computer Engineering and Department of Cognitive Science, The Johns Hopkins University, Baltimore, Maryland 21218

Edward P. Davis
Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 27 December 1994; accepted for publication 31 July 1995)

A head and transducer support system (HATS) was developed for use in ultrasound imaging of tongue movement. Ultrasound is an imaging technique that captures tongue motion during speech and thus has great appeal as a tool for speech research. However, because ultrasound systems are designed for clinical use, the transducer is hand-held and it is almost impossible to hold it completely steady under the chin when collecting tongue data. A system was needed to fix the head and support the transducer under the chin in a known position without disturbing speech. The HATS system was designed, constructed, and modified to provide valid, reliable tongue movement data by (1) immobilizing the head and (2) positioning the ultrasound transducer in a known relationship to the head. © 1995 Acoustical Society of America.

PACS numbers: 43.70.Jt

INTRODUCTION

An underutilized, but very useful instrument for measuring tongue behavior is ultrasound imaging. Ultrasound images provide a unique and surprisingly complete representation of tongue movement during speech. An ultrasound image is a tomograph that is 2 mm thick. Sections of the tongue can be imaged in almost all planes of interest including mid- and parasagittal, coronal, and oblique. Moreover, the measurements are real-time; 30 or more scans per second. Usually, physiological recording techniques either provide an image of the tongue but are too slow to record movement, such as MRI (Baer et al., 1991), or provide rapid tracking of individual points on the tongue but do not record the entire structure, such as x-ray microbeam and electromagnetic articulometry (Perkell et al., 1992; Westbury, 1991). X-ray images are relatively rapid and record the entire tongue but are seldom used now because of the radiation hazard. Ultrasound also images the entire structure in one plane and is rapid enough to record movement (Stone et al., 1992; Stone and Vatikiotis-Bateson, 1995). In addition, 3-D reconstruction of the tongue surface from a series of ultrasound slices is already possible in static data (Stone and Lundberg, 1994, 1996) and soon will be possible in the time-varying case. Therefore, if ultrasound images can be captured with enough accuracy and reliability to ensure confidence in the recorded data, it will become an ideal tool for examining tongue movement.

It is challenging to accurately measure the tongue using ultrasound. In order to make tongue measurements, the ultrasound transducer is placed below the chin and the beam directed upward through the soft tissue of the floor of the mouth. The sound wave reflects back from the tissue–air interface at the upper surface of the tongue. Since ultrasound was developed for clinical purposes, the transducer is traditionally hand-held. This technique is not acceptable for research applications of ultrasound imaging because the slightest movement of the hand will cause rotation or translation of the transducer, resulting in an off-plane image. In addition, since many ultrasound transducers are curvilinear in shape, measurements in the coronal (cross-sectional) plane allow slippage on this curved surface. Furthermore, the natural movement of the jaw during speech makes steady positioning by hand very difficult. While these problems can be mitigated using careful positioning (cf. Stone et al., 1988), the accuracy and reliability of such data cannot be guaranteed.

Repeatability has been a problem with ultrasound imaging of the tongue. Assuming that the transducer is held steadily under the chin, either by hand or by a mechanical device, it has been virtually impossible to return to the same position for a repeat measurement. This is due to irregularity of head sizes and shapes and the inability to reposition the transducer in the same location across sessions. Although we are currently working to achieve high measurement repeatability, at the present time we cannot quantitatively compare measures made across sessions or subjects.

The problem of stably positioning the transducer under the chin has long been a limiting factor in ultrasound studies of the tongue (cf. Morrish et al., 1985) because of three problems: (1) immobilization of the head, (2) stabilization of the transducer, and (3) depression of soft tissue by the immobilized transducer.

I. HEAD IMMOBILIZATION

One way to control head movement is to immobilize the head completely. This has been done during brain surgery and involves piercing the skin and holding the skull in posi-
tion using three points. A similar approach uses immobile plugs fitted in the subject’s ears. This second approach is uncomfortable, however, and even painful. The invasiveness of these procedures makes them unsuitable for ordinary speech production research. Another head immobilization procedure is used occasionally in MRI scanning. A mold is made of the subject’s face and a face mask constructed with eye and nose openings. The mask is fitted to the standard MRI head rest to completely immobilize the head. This is unacceptable for speech research because the subject cannot open his/her jaw, and the mask feels very confining to the subject.

A better strategy for stabilizing the head during ultrasound imaging is to stabilize the skull while allowing the jaw free movement. One method of stabilizing the skull but not the jaw employs a helmet. Keller and Ostry (1983) used this approach to measure /ka/ sounds using a small A-mode ultrasound transducer resting in the crook-of-the-neck. The A-mode transducer measures edges at a single point in a plane. They concluded that jaw opening and upward pressure from the transducer were not critical problems, provided the speech sample was limited to velar consonants. This technique is not suitable for more recently developed ultrasound instrumentation because the new transducers, which scan a slice of tissue, are larger and heavier than the A modes. In addition, in a preliminary system, we found a helmet-based holder caused claustrophobia, discomfort, and inadequate head stabilization.

An alternative method of stabilizing the skull is wrapping a fitted band around it, or offering resistance at specific locations. Such a system was devised previously for ultrasound data collection (Stone et al., 1988). Although this device stabilized the head and discouraged large movements, over time the head had a tendency to drift and the band to loosen. Electromagnetic articulometers (EMAs) also stabilize the head with a forehead band. The head is surrounded by a large apparatus suspended from above which houses transmit sound coils and also stabilizes the head with respect to the recording device. A band, within the suspended apparatus, wraps around the forehead and back of the head, and may be tightened inward from the sides (Perkell et al., 1992; Alfonso et al., 1993). However, head immobility even within the device is less assured in the front-to-back dimension than the side-to-side one. Therefore, a headband stabilizer is not adequate for use with ultrasound.

II. TRANSDUCER STABILIZATION

Holding the transducer in a known relationship relative to the head is necessary so that the tongue/jaw data will be aligned with the head. This is important in ultrasound data collection because it allows precise measurement of where the ultrasound beam is pointing and assures consistent data measurements. Some instruments, such as point tracking measures, use a subtraction method to align tongue/jaw data with head position. In this procedure the head is allowed to move freely and its movements are tracked using the same device that tracks articulatory motion. Head movements are mathematically subtracted out of the tongue/jaw data during data analysis (Vatikiotis-Bateson and Ostry, 1995; Westbury, 1991). This strategy would not work for ultrasound, however, because the same system does not collect both the head and tongue movements. The ultrasound transducer measures the tongue image, but head position must be measured using some other system. Head and transducer position can both be tracked using a point tracking device. This would allow later detection of slippage in transducer position. However, the slippage could not be corrected mathematically because the ultrasound images collected during the slippage would be in the wrong plane. Even a half-second off plane would produce 15 error frames, which is too many to interpolate. Therefore, the desired images would be nonexistent and could not be recovered through mathematical calculation.

The method used in our previous transducer holder (Stone et al., 1988) was to attach a series of rods and beams to a dental chair base. The rods supported a housing that positioned the ultrasound transducer under the chin. The transducer was positioned, and then locked in place. The entire housing moved up and down with the jaw, on springs. The head was held in position with a headband. However, this system did not adequately stabilize the head and did not position the transducer with enough accuracy and precision.

III. TISSUE DEPRESSION

A concern with using a fixed transducer position was that when the jaw opened downward, the soft tissue contacting the transducer would be pushed upward by the fixed transducer, causing artificial tongue movement. To avoid this possibility, a 1/2-in.-thick acoustic standoff was used to distance the transducer from the soft tissue of the chin (see Figs. 1 and 2). The acoustic standoff (Kitecko, 3M) was a gelatinous substance made from polymerized mineral oil. The standoff compressed more readily than the soft tissue of the tongue and floor of the mouth. This allowed the jaw to move freely, while maintaining clean transmission of the beam into the skin, and preventing tissue compression from the upward pressure of the transducer. (See below for validation.)
of the head and the transducer as well as adequate subject comfort and calibration.

A. Table and superstructure

In the chosen design, a table was used as a base for the apparatus. A U-shaped cut allowed it to fit around a subject seated in a chair. A superstructure, to support the headholder, was fixed directly to the table.

The superstructure was fitted above the cut out section of the table and contained an H-shaped frame that supported the head holder (Fig. 1). The structure was bolted together, to the table, and gusseted with small pieces of angle iron to make it stronger, stiffer, and resistant to vibration.

The H-frame was made from 2.5-in.-square, 1/4-in.-thick steel tubing, with all welded construction for high strength and rigidity. The H-frame had four reinforcement plates brazed on. In these reinforced areas of the tube, holes were drilled and precision reamed to support the vertical bars of the head-holder. This reinforcement insured high strength, stiffness, precision, and wear resistance.

The headholder was suspended from the H-frame’s cross bars by two vertical bars. Locking knobs on each side of the H-frame supported and locked the vertical bars of the headholder in place by pressing the bars against the reamed holes ensuring a tight lock.

B. Head holder

The first consideration in designing the HATS system was immobilizing the head. During ultrasound data collection, the head must be fixed in relation to the ultrasound transducer for several reasons. First, movement of the head during an experiment will lead to inaccurate measurements. Second, it is desirable to be able to direct the ultrasound beam to the right area of the vocal tract; this requires reliable positioning of the transducer with respect to the head. Third, the transducer must be in contact with the acoustic standoff, and the standoff must be in contact with the skin to ensure a good image.

In determining the optimal method for steadying the head, considerable research was done to learn about head sizes and shapes (Pheasant, 1986). Regarding size, the holder was designed to accommodate the smallest and largest head sizes—a difference of 2 in. in length and 2 in. in breadth. Clamps contacted the head at four locations and their supporting shafts were easily lengthened or shortened to accommodate variation in head size and length/breadth ratio.

To accommodate varying head shapes, several aspects of head anatomy were considered. Based on the location of the flat portions of the skull, and the potential rotation, translation and tilt during head movement, the following features were considered essential in designing the headholder. (1) Opposition to left/right rotation and tilt of the head had to occur at the flat portion of the temples, just above and in front of the ear. Opposition in this area would minimize slippage of the bone beneath the skin. The temples are slightly anterior to the center of the head in the A/P dimension. This anteriority would prevent the clamps from acting as a fulcrum that allowed head rotation. (2) Opposition to front/back rotation and translation was best accomplished with resis-
tance at the forehead and base of the skull. (3) The posterior head clamp had to angulate to accommodate a variety of head shapes. (4) Well-padded clamps which had large contact areas and would conform to local head shapes were used to ensure the subject's comfort. (5) Finally, all clamps needed to be attached to their shafts using swivel feet. In this way the shaft angles could be positioned normal to the direction of head movement while the padded clamps conformed to local head shape.

The headholder system was made primarily from lightweight aluminum. Its largest component was a circular ring (Fig. 1) made from a welded rectangular aluminum bar. The ring was suspended from the H-frame by two vertical bars welded to the sides of the ring. The vertical bars were fitted through holes drilled through the H-frame to allow up and down adjustment of the ring. Each vertical bar had a flat milled surface on one side for maximum fixation against the H-frame with set screws. Five 1.5-, 1.5-, 1.0-in. aluminum blocks were welded around the sides of the circular ring. One was positioned on the front, two on each side, and two on the back of the ring.

The front and side blocks each had a horizontal hole drilled and threaded through the center which extended through the ring to support a horizontal \( \frac{3}{4} \) diameter screw. Two knobs were fitted on the distal end of each screw, one to permit rotation of the screw and the other to lock it in place. On the proximal end of each screw was a clamp attached to the screw with a swivel joint to permit maximum rotation of the clamp for optimal positioning against the subject's head. The 4-, 1-, 0.06-in. forehead clamp was curved to a diameter to maximize contact area with the front of the subject's head. The 2.75-, 1.25-, 0.06-in. side clamps were flat.

The two remaining blocks had unthreaded vertical holes and were welded to the back of the ring to support the posterior head rest. The head rest consisted of four unthreaded \( \frac{1}{4} \) diameter bars, a 6-, 2-, 1-in. rectangular support block, and a 6-, 4-, 0.06-in. rectangular clamp. It was designed to permit maximum adjustment against the back of the subject's head. Two of the bars were positioned vertically and fitted on one end through the small blocks on the ring and on the other end through two unthreaded vertical holes in the support block. The other two bars were positioned horizontally and fitted through unthreaded horizontal holes in the support block. All of the fittings were secured with set screws. The clamp was attached to the proximal ends of the horizontal rods with pins so as to allow free up and down rotation. This clamp was slightly curved (3.56-in. radius), like the forehead clamp, to fit the natural curve of the back of the head. Also like the other clamps, it was covered with a foam pad to maximize subject comfort.

1. **Head-holder validation**

The success of the headholder in stabilizing head position was tested using lateral videotapes. During the recording session the subject's head was recorded with a video camera positioned on the right side. This video image was inserted into one corner of the ultrasound image for later use in aligning the ultrasound image with head position. Two subjects were recorded for sessions of 20 min in length. Frames were taken from the videotape at the beginning and the end of the session. Two easily identified features were measured on each subject: the location of the outer canthus (corner) of the eye, and the lowest anterior most portion of the ear canal. For one subject, the two features were in the identical location on both images. For the second subject both features were 1 pixel lower in the y dimension and identical in the x dimension for the later image. One pixel on these video images was 0.56 mm. This difference indicated virtually no head movement after 20 min of recording.

**C. Transducer holder**

The major considerations in designing the transducer holder were: (1) positioning the transducer in an easily accomplished manner, (2) determining transducer position in arbitrary head space, and (3) not depressing the soft tissue of the floor of the mouth.

The transducer-holder immobilized, positioned, and tilted an ultrasound transducer under the subject's chin (Fig. 2). The holder was secured firmly to the same table as the frame supporting the head holder. Hence accurate positioning of the transducer with respect to the head was possible.

In preliminary transducer holder designs, the transducer was positioned so that it moved up and down with the jaw. That feature was eliminated in the final design because of the way that tongue motion is coupled to jaw movement. Near the front of the mouth, tongue movement is very sensitive to jaw movement. This is less true in the pharynx. Hence, if transducer movement were linked to jaw movement, observations of the back of the tongue would contain irrelevant movement of the jaw. It was decided, therefore, to keep the transducer immobile, and use the acoustic standoff to maintain contact between the transducer and the jaw. In this case the measurements are of tongue/jaw movement, not tongue movement alone.

1. **Positioning the transducer**

Forward and backward movement of the transducer was made along a fore-aft rail attached to the table on the subject's left (see Fig. 1). A horizontal locking carriage slid along the rail and held an upright rail 90 deg to the horizontal one. Upward and downward movement of the transducer was made along the upright rail. A locking vertical carriage slid along the upright rail and supported a round metal bar horizontally. This bar slid left and right to position the transducer on the subject and was then locked in place. The round bar rotated within the carriage support, allowing angular positioning of the transducer. A drum dial, marked off in degrees, was attached to the bar so it could be rotated in measured amounts. At the end of the bar was a square-edged U-shaped bracket. The distal arm of the U ended at the height of the bar's center of rotation. When the transducer was positioned correctly, the image center of the ultrasound beam was the center of rotation of the bar. This has several distinct advantages. First, the transducer could be easily and accurately re-angled by rotating the bar. Second, no translation (i.e., position change) occurred during angular readjustment between scans. Transducer rotation could be done in-
dependently of transducer translation. Third, the standoff stayed evenly distributed when the angle was adjusted.

2. Determining ultrasound beam angle

An important part of the transducer placement was determining the angle of the ultrasound beam relative to both arbitrary and human spatial coordinates. The most important considerations when attaching a transducer to the U bracket were to ensure that it was firmly secured with the center of the transducer surface at the center of rotation of the bar, and that the ultrasound beam was pointing straight up and not angled off-axis. Custom mounts were made for the most frequently used transducers.

To determine the ultrasound beam angle in arbitrary space, a miniature level was mounted on the flat horizontal beam portion of the U bracket. The level was used to set an angle measuring collar so that it read zero degrees when the ultrasound beam was vertical. When the horizontal bar was rotated to reangle the transducer, the collar indicated the angle of rotation of the ultrasound beam relative to vertical. The angle of rotation marked the plane at which the beam passed through the head and was used in subsequent image alignment and 3-D reconstruction. Once the head and transducer were positioned acceptably, calibration measurements were performed to: (1) determine that the transducer was placed in the midsagittal plane and the beam was directed vertically; (2) position the transducer a measured distance back from the mandibular symphysis (front of jaw); and (3) determine the ultrasound beam angle relative to the occlusal plane, the body of the mandible, and an experimental pair of glasses. The last measurement was done later on a full-screen video of the subject’s head, recorded immediately before and after the recording session (see Fig. 2).

3. Adding noise to the image

One consideration when using an acoustic standoff was whether the standoff itself caused distortion or increased noise at the tongue surface. To test this, a small data set was collected with and without the standoff in place. The target vowel [a] spoken in both conditions is shown in Fig. 3. The tongue movements were unsmoothed. The standoff appeared slightly noisier, although smoothing reduced this effect. The slight difference in shape may be due to the inability to perfectly guarantee the same transducer position in both conditions.

4. Transducer holder validation

The HATS system was validated using videofluoroscopy, to ensure that the tissue under the jaw was not depressed during data collection, which would cause measurement error. Since the purpose of the standoff was to deform and allow natural soft tissue movement during speech, it was important to make sure that the transducer and standoff did not push upward on the tissue causing apparent tongue elevation.

Videofluoroscopic x rays were made of the tongue movements of a single subject (the second author) during speech with and without the transducer in place. The subject was seated in the HATS with his head in the headholder. The ultrasound transducer was put in place beneath the jaw using the transducer holder and standoff. Radio-opaque markers (1-mm dots) were placed on the subject’s cheek. The distance between these markers was measured to allow later normalization of measurements. The subject was asked to say “asassa, atata” twice each, while lateral x rays were taken and recorded on videotape. Next, the ultrasound transducer was removed, and the procedure was repeated.

The two sets of x rays were compared for differences due to the transducer/standoff arrangement. Sequences of x-ray frames corresponding to the sound [a] in [5] and [1] context, with and without the transducer below the chin, were selected for analysis. The sound [a] was chosen for examination because it requires the widest opening of the vocal tract, and hence had the highest likelihood of being disturbed by upward pressure from the transducer. The frame corresponding to the lowest position of the tongue in each of the two sets of frames was determined to be the “target” tongue position for [a] and was selected for measurement. These eight frames (2 repetitions × 2 contexts × 2 conditions) were imported onto a Macintosh computer. The lengthwise contour of the tongue was detected, smoothing was performed, and xy coordinates corresponding to the midsagittal tongue surface were recorded to a data file using a custom edge detection program (Unser and Stone, 1992). Additionally, the locations of the two radio-opaque markers were recorded for both images. These markers were left on the subject’s face during both the period with and the period without the transducer in place. They acted as constant points in space for alignment of the two images. The data file was imported to Microsoft Excel, where the data were aligned with each other using the positions of the radio-opaque markers as reference points.

After alignment, the two curves in each condition and context were averaged and the differences between the averaged curves were calculated for both contexts using an $L_1$ norm. An $L_1$ norm represents the average absolute difference between the two curves.
\[
\left( \frac{1}{n} \sum_{i=1}^{n} |a_i - b_i| \right) / n,
\]

where \(a\) and \(b\) are the two curves and \(n\) is the number of points per curve. The difference between the \([a]\)'s with and without the standoff in [ta] was 0.11 mm. The difference between the \([a]\)'s in [so] was 0.19 mm. These differences were well within our measurement error of 0.5–0.7 mm (Stone et al., 1988). Thus it appears that tissue depression did not have an appreciable effect on tongue shape using this system. An \(L_2\) norm (average squared difference) and \(L_{\infty}\) norm (largest absolute difference) were also calculated. However, these measures varied with the number of points chosen. The \(L_1\) norm was the most robust in that it provided identical results regardless of \(n\).

VI. CONCLUSION

Ultrasound is being developed as a technique for the investigation of the tongue during speech and swallowing. In order to scientifically study the tongue, precise, accurate 3-D tongue data are essential. The HATS system is able to hold the head steady and to prevent casual movement. It also allows rapid placement of the ultrasound transducer under the chin in a known position, and allows measurement of ultrasound beam angle relative to head space. In this way the tongue measurements can be understood in their relationship to head position.

The HATS system has been constructed and tested. X-ray validation has shown that the device does not disturb articulatory motion.

ACKNOWLEDGMENTS

The authors would like to thank Phillip E. Atkinson, Raulie Lo, Elisa Garvey, Dziugas Radzius, Joe Perkell, and Neil Rothman for their contributions to this and earlier versions of the HATS system. We would also like to thank Yongxing Zhang for her work on the \(L_1\) norms. This project was supported (in part) by Research Grant No. 7 R01 DC 01758-03 from the National Institute on Deafness and Other Communication Disorders, NIH; the Center for Language and Speech Processing, Johns Hopkins University; and Westinghouse Electric Corporation.


