Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

On the problem of tracking mandibular movements

Cooker, H. S.

journal: STL-QPSR
volume: 14
number: 2-3
year: 1973
pages: 001-012

http://www.speech.kth.se/qpsr
Abstract

A rationale is presented for obtaining information on the movement of the articulators during the production of speech. Previous methods of tracking the articulators are discussed, their limitations are reviewed, and the importance to articulatory theory of obtaining artifact free measurements is emphasized. An articulatory tracking system that is based on the Whalen and DeHaan bonded wire strain gage technique is developed, and the output of the tracking system is assessed by two different experimental procedures. It is demonstrated that direct tracking of the mandible is the only technique that yields movement information that is uncontaminated by other articulator movements. Indirect tracking methods are shown to be useful for certain limited purposes, but it is concluded that any explanation of articulatory function in man may be suspect if it is based on contaminated measurements of articulatory movement.

Introduction

The production of the speech code is accomplished through a complex interlacing of various articulatory movements or gestures that are spatially and temporally variable. The neutral open shape of the vocal tract is modified by these gestures, and the resulting perturbations in the tract yield resonance transfer functions that are peculiar to the sounds of speech. A variety of acoustic analyses has been employed in attempts to unravel the fabric of speech and provide the investigators with rules or lawful relationships that would lead to a solution of the code. Certainly such analyses have led to valuable insights for a better understanding of the processes of speech. However, speech is so complex a code that these analyses have yielded only partial solutions. It would appear that valuable information might be obtained by observing speech activity at some point that is closer to the cerebral function that is generating the code. Although the speech signal at such a point would not be in an appropriate form for transmission to a listener's ear, it should be less highly encoded and, depending on its distance from the acoustic speech signal, relatable to known acoustic events.

Since speech sounds are laid down in a temporal stream by the intricate overlapping actions of the speaker's articulatory members, it seems plausible that careful observation of the movement of these articulators could reveal valuable information about the final movement-to-sound coding process. Furthermore, these movements could be related to the next higher level of activity, muscle contraction, and provide insights into the motor-acoustic relationships of speech.

Measurement of speech movements

A great deal of the early work on the measurement of the movements of speech was conducted by R.H. Stetson and his colleagues (1928, 1930, 1931, 1935). Much of this early work was reviewed and presented in the second edition of Motor Phonetics (Stetson, 1951). The methods employed in these investigations were based largely on a system of pneumatic transducers that purportedly were sensitive to very small changes in air pressure that occurred when the moving articulators exerted forces on the transducers. In at least one of these studies, (Stetson and Bouman, 1935) electromyographic procedures were employed to observe muscular contractions of the breathing muscles that were associated with the repetition of one-syllable words. Unfortunately, pneumatic systems are often non-linear in response and require careful and frequent calibration to insure accuracy of timing and amplitude excursions. Also, electromyographic records from surface locations in multiple muscle areas often are subject to widely divergent interpretations which are resolved only through repetitive measures from multiple electrode locations. Stetson and his co-workers produced a considerable number of registrations of "chest pulses" (chest wall movements), articulatory movements, and at least one electromyographic record of breathing muscle activity during the process of speaking, but the lack of calibration procedures and the inability of other investigators to confirm their findings has led to a general lack of faith in the validity of the methodology employed by the Stetson Group.

In an attempt to investigate Stetson's so-called chest pulses, Cooker (1963) employed a bonded wire strain gage device that measured the displacement of the chest wall that occurs during speech. The device first was developed for physiological research by Whalen and DeHaan (1955) as a way of measuring the breathing pattern of laboratory animals.
The sensing element consisted of a thin strip of spring steel to which bonded wire strain gages were cemented. This sandwich of a steel beam and two opposing gages was connected as one half of a Wheatstone bridge, and the output was fed to a high quality D.C. amplifier. As the beam of spring steel was bent, one strain gage was under compression, while the other was under tension. The result was a change in resistance in two arms of the bridge and a corresponding change in the voltage output of the D.C. amplifier (see Fig. I-A-1). These fluctuations in voltage were fed to strip chart recorders or other devices to provide a permanent record of the movements of interest. The basic design was modified by Cooker to provide a constant point of contact on the chest wall and an increased sensitivity that would permit the accurate measurement of displacements as small as three microinches.

Sussman and Smith (1970a, 1970b) describe two transducer systems which they constructed to measure mandibular movement and lip movement during the production of speech. The sensing elements of these devices are essentially identical to those used by Whalen and DeHaan and by Cooker, but an important modification was introduced to permit the tracking of moving facial structures. The sensing element was anchored to the subject by a headband which provided a self-reference that reduces subject movement artifacts.

Other systems that utilize photographic or photoelectric procedures to track jaw and lip movements have been developed (Kozhevnikov and Chistovich, 1965; Lindblom and Bivner, 1966; Lindblom, 1967; Ohala, Hiki, Hubler, and Harshman, 1968). These systems require custom fitted tooth attachments, are subject to movement artifacts, or do not yield a convenient analogue of the articulatory movement being studied.

With the development of cineradiography, it was possible to measure the movements of articulators that normally are hidden from view; the tongue, velum, pharyngeal wall, and mandible. The mandible is included as a hidden articulator, for it is covered with an elaborate array of muscles of facial expression which tend to mask out the posture of the underlying boney structure. The general procedure for using cineradiography to track the articulators is rather laborious in that it involves frame-by-frame analysis in which the structure of interest is
traced and selected points recorded for each exposure that is produced. Since rapid tongue tip movements may require film speeds that approach 200 frames/sec for adequate sampling, it is apparent that the procedure is costly in both analysis time and in money.

In summary then, it would appear that the most convenient and least expensive method of tracking articulatory movement is through the use of the bonded wire strain gage in some variant form of the basic Whalen and DeHaan system. Of course, the usefulness of such a system would be severely restricted if ease and convenience of use were attained at the expense of increased error of measurement. Consequently, it was determined that an articulatory tracking system should be developed using bonded wire strain gage sensing elements and that the output of the system should be assessed carefully to establish the validity of the measurements obtained. Since one of the goals in observing speech movements is to relate these movements to controlling neural events, any measurement error could lead to serious misinterpretation of motor-acoustic relationships and result in erroneous conclusions as to the function of cerebral processes during speech.

Development of the articulatory tracking system

A sensing element was constructed of bonded-wire strain gages in the manner described previously, and several tracking probes were fabricated to provide for attachment of the sensing element to the subject. The complete system including probes had a natural frequency of 33 Hz which is well above the frequencies of interest, and the output was linear over a displacement range of \( \pm 40 \text{ mm} \), the widest range tested. The nature of the reference anchor, the location of the tracking point, and the method of attachment to the articulatory structures proved to be major considerations in the development of the system.

Reference Anchor: The anchoring of the sensing element to the subject poses certain problems in that the attachment system must be adaptable to a wide range of geometric configurations. The headband method of Sussman and Smith (1970a), though light, flexible, and easy to use, was discarded in favor of a more rigid system that eliminated the yaw that was a problem with our version of the headband mount. An adjustable semi-rigid plastic helmet was used as the basic head mount. It was appropriately modified through the addition of cinching straps to provide a
tight fit both for small children and for large adults. An aluminum frame was mounted firmly to the anterior of the helmet, and a three-dimensional adjustment system was attached to the frame. The weight of the frame and adjustment system were counterbalanced to reduce subject fatigue and insure long-time stability of the adjustment system. With the above system it is possible to support several of the displacement elements at one time. Thus, the method provides the capability of obtaining simultaneous displacement measures from more than one location.

Tracking Points: The most important factor in the selection of an articulatory tracking point is whether or not its movement is representative of the movement of the structure being studied. Other factors to be considered are excursion of movement, ease of attachment, subject comfort, long term stability. Lindblom (1967) demonstrated that the best tracking point for obtaining representative mandibular movement data is located at the lower teeth. Using a photographic procedure, Lindblom and Bivner (1966) previously showed that the movements of the skin over the chin were similar to lip movements and suggested that they did not represent pure mandible movement. However, the undersurface of the jaw is much more accessible than the teeth during speech, and it seemed plausible that, with the strain gage technique, a tracking point could be located that would reflect the movement of the underlying boney structure rather closely. After much preliminary experimentation, two points were selected as possible locations for the attachment of the tracking probes: the posterior undersurface of the chin directly over the center of the body of the mandible, and a point directly over the bone approximately two centimeters anterior to the angle of the mandible. These points are illustrated in Fig. I-A-2.

Attachment of Tracking Probes: The greatest potential source of error in the measurement of articulatory movement is at the junction between the sensing element and the structure of interest. Not only is the point of contact critical but also the method of attachment is extremely important if one is to obtain registrations that are free from artifacts. Several types of probes are illustrated in Fig. I-A-1. As can be seen in the illustration, the various probes are affixed to the end of the sensing element to provide different types of surface contact. Probes 1 and 2 include provisions for cementing the functional contact point to the surface of the skin. Probe 3 was designed for use with bearded subjects.
Fig. I-A-1. Sensing element and articulatory tracking probes for the displacement transducer.

Fig. I-A-2. Points of attachment for indirect tracking probes.
and presents a very small contact area. Both 2 and 3 may be inverted to provide two additional types of surface contacts. Probe 1 employs a low friction bearing system to permit the tracking of structures that are displaced in two directions.

Method of evaluation

Two different procedures were employed to evaluate the system. In the first procedure the tracking probes were attached either to the posterior undersurface of the chin or to a point near the angle of the mandible. The four male speakers who acted as subjects were asked to produce the vowels, /i/, /a/, and /u/ and the consonants, /p/, /b/, and /m/ in the linguistic frame, "Say /V CV CV/ again". Each sequence was repeated three times. At the end of each sequence, a neutral spacer was introduced between the molars to immobilize the jaw at an opening that was appropriate for the vowel that was to be produced and the sequence was repeated with fixed mandible. The neutral spacer was used in an attempt to eliminate any compensatory articulatory movements that might have been introduced if the teeth were clenched to achieve immobilization. The output of the movement transducer together with the voice signal were fed to two channels of a mingograph for permanent recording.

In a second procedure, one subject was fitted with a tooth cap-splint that was attached to the transducer system by a modification of a type 1 probe (Fig. I-A-I). A second probe was positioned on the undersurface of the chin. This arrangement permitted both direct and indirect tracking of the mandible simultaneously. The subject was asked to produce the same sequences as were produced by the four subjects in the first procedure including the immobilized jaw condition. The output from the two movement transducers and the voice signal were fed to three channels of a mingograph for permanent recording.

Results

For all consonant/vowel combinations, immobilization of the jaw failed to eliminate movement at either of the tracking probe locations used in the first procedure. Residual movement at the undersurface of the chin was almost always at least fifty percent and frequently eighty or ninety percent of the movement that was recorded when the jaw was unrestricted. Residual movement at the angle of the mandible was con-
considerably smaller. Generally, it was in the order of ten to fifteen percent of the unrestricted movement. A representative record of the restricted and unrestricted movements for the two different tracking probe locations is presented in Fig. I-A-3 and I-A-4.

The results of the second procedure clearly reveal both differences and similarities between direct and indirect tracking of mandibular action. In Fig. I-A-5 the amplitude and velocity of the excursions for the bilabial consonant occlusions are considerably greater for the indirect tracking probe located under the chin than for the direct tracking probe attached to the tooth cap-splint. The same observation can be made for Fig. I-A-6 in which the vowel /u/ is substituted for the vowel /i/. However, the movements associated with the linguistic frame, "say ... again", are very similar in both amplitude and velocity for both tracking methods. With the exception of the second bilabial occlusion in each production of the linguistic frame, all of the data for the vowel /a/ presented in Fig. I-A-7 indicate similar excursion patterns for the two tracking methods. The mean values of jaw displacement associated with the second vowel - bilabial consonant sequences are shown in Table I-A-1. A comparison of the direct and indirect measurements of mandible activity that occurs during these sequences reveals large differences in the values obtained. Invariably the indirect method yields values that are at least twice as large and often four or five times as large as the values obtained with the direct tracking probe.

Immobilization of the jaw with a neutral spacer eliminated the movement of the direct tracking probe but merely attenuated the movement of the indirect probe. Inspection of Figs. I-A-8, I-A-9, and I-A-10, suggests that greater amplitude of residual movement is present for the /a/ and /i/ vowel contexts than for the /u/ context. However, a comparison of the movement traces presented in Figs. I-A-5 through I-A-9 with those presented in Figs. I-A-8 through I-A-10 shows that a large proportion of the movement measured indirectly remains under the neutral spacer conditions for all of the vowels studied. This interpretation is supported also by the data presented in Table I-A-1 for the /u/ vowel context which shows that about 80 percent of the free jaw movement is present at the location of the indirect probe after immobilization of the jaw.
Fig. I-A-3. Oscillographic records of the voice signal and jaw movements obtained from the undersurface of the chin during free and fixed jaw conditions.
Fig. I-A-4. Oscillographic records of the voice signal and jaw movements obtained from the angle of the mandible during free and fixed jaw conditions.
Fig. I-A-5. Oscillographic records of the voice signal, direct, and indirect measurement of free jaw movement for the vowel context /i/.
Fig. I-A-6. Oscillographic records of the voice signal, direct, and indirect measurement of free jaw movement for the vowel context /u/. 
Fig. I-A-7. Oscillographic records of the voice signal, direct, and indirect measurement of free jaw movement for the vowel context /a/.
Fig. 1-A-8. Oscillographic records of the voice signal, direct, and indirect measurement of restricted jaw movement for the vowel context /i/. 

NEUTRAL SPACER

TIME

VOICE

DIRECT

INDIRECT

say ipipi again

VOICE

DIRECT

DIRECT

ibibi

INDIRECT

VOICE

DIRECT

DIRECT

imimi

INDIRECT
Fig. 1-A-9. Oscillographic records of the voice signal, direct, and indirect measurement of restricted jaw movement for the vowel context /u/.
Fig. I-A-10. Oscillographic records of the voice signal, direct, and indirect measurement of restricted jaw movement for the vowel context /a/.
TABLE I-A-I. Mean jaw displacement values in millimeters for the second VC movement in the linguistic frame, "Say /VCVCV/ again" measured directly, indirectly, and under neutral spacer conditions.

<table>
<thead>
<tr>
<th>Jaw free</th>
<th>Neutral spacer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct</td>
</tr>
<tr>
<td>i p</td>
<td>0.7</td>
</tr>
<tr>
<td>i b</td>
<td>0.5</td>
</tr>
<tr>
<td>i m</td>
<td>0.4</td>
</tr>
<tr>
<td>u p</td>
<td>0.7</td>
</tr>
<tr>
<td>u b</td>
<td>0.3</td>
</tr>
<tr>
<td>u m</td>
<td>0.3</td>
</tr>
<tr>
<td>a p</td>
<td>1.5</td>
</tr>
<tr>
<td>a b</td>
<td>1.8</td>
</tr>
<tr>
<td>a m</td>
<td>1.9</td>
</tr>
</tbody>
</table>

In summarizing the results of the several tracking procedures that were used, it seems clear that different methods produced quite different results. The extent of the movement recorded depended markedly on the selection of the tracking point location. Furthermore, the total elimination of mandibular movement that was achieved through the use of neutral spacers was reflected only in the output from the transducer that was connected directly to the mandible through the lower teeth. And finally, the text of the spoken material had an important effect on the agreement between direct and indirect measurements of mandible activity during speech.

Discussion

The articulatory tracking system that was developed proved to be highly satisfactory for observing the on-going movements of the mandible during the production of speech. When coupled directly to the mandible through a tooth cap-splint, the vertical movements of this boney structure were followed with a high degree of accuracy. Displacement, velocity, and time pattern information are readily measured through the use of this system.
Unfortunately, the more easily employed indirect tracking probes do not yield accurate information under all conditions. When attached to the undersurface of the chin, the probe follows not only the mandible but also the movements of the soft tissue associated with elevation and depression of the lower lip. Extensive experimentation with various probes on a large number of subjects led to the conclusion that it was not possible to eliminate this soft tissue movement. Whenever the speakers produced bi-labial sounds, the activity of the mentalis and depressor labii inferior acting upon the lower lip was transmitted to the probe under the chin and added movements to the records that were not caused by the mandible. In effect, then, the records obtained from this position are a complex combination of displacements and velocities from the action of the lower lip and mandible.

Recently, Sussman, MacNeilage, and Hanson (1973) conducted an investigation of articulatory movement in which they examined the displacement and velocity of the mandible and upper and lower lips during the production of bilabial stop consonants. Their technique involved the use of an indirect tracking probe positioned under the chin to record the dynamics of the mandible. They assumed that these records were uncontaminated movement traces of mandibular activity and based a number of derived measures of articulatory activity upon these traces. The results of our investigation suggest that their assumption of uncontaminated mandibular activity was in error and may have led them to draw erroneous conclusions about the basic nature of articulatory function.

Perhaps the best location for the indirect tracking of the mandible is a point near the angle that is relatively free from contamination by facial muscle activity. Due to its proximity to the mandibular fulcrum, the magnitude of displacement is considerably less at this point than at the undersurface of the chin, but increased amplification can eliminate this minor problem. A more difficult problem is encountered with contamination from tongue movements. The high vowels are especially troublesome in that the soft tissues of the undersurface of the jaw tend to be pushed laterally when these vowels are produced, and this lateral movement may be sufficient to produce an apparent displacement of the
mandible in some subjects. However, in comparison with the undersurface of the chin, the degree of contamination is likely to be small.

If the material to be produced consists of vowels, diphthongs, or non labial sounds, the undersurface of the chin is a very desirable location. It is easily accessible, and a number of different types of tracking probes may be used with confidence. One good arrangement is to attach the probe with an adhesive to avoid migration of the point of contact. Such an arrangement is especially helpful when dealing with children and other active subjects. Furthermore, the positive attachment of the probe-end eliminates prestressing of the sensing element which increases the useful range of the system. There are other occasions when a fixed contact point is not as desirable as a sliding contact. For very large mandible excursions, as in the case of singers producing high pitched vowels, a sliding contact is useful in accommodating the extreme excursions encountered, and rather simple calibration procedures can account for the moving contact and insure the accuracy of the measurements.

Of course the undersurface of the chin may be used as the tracking point even though facial muscle action contaminates the movement data if the purposes of the experiment are appropriate. For instance, the timing characteristics of articulatory function for individual speakers or categories of speakers may be compared quite nicely by observing the temporal patterning of the movement trace during connected speech. In this way, the differences and similarities in articulatory function between child and adult, male and female, or normal and non-normal may come to light.

Conclusions

The results of this investigation suggest the following conclusions:

(1) The bonded-wire strain gage technique provides the investigator with an efficient tool for measuring articulatory movements during on-going speech activity.

(2) The location of the tracking point and the attachment of the tracking probe are the major considerations in obtaining measurements that are free from artifacts.

(3) Accurate tracking of the mandible during connected speech can be accomplished only through direct connection of the tracking probe to the boney structure by a tooth cap-splint or similar means.
(4) When attached to the posterior undersurface of the chin, a tracking probe describes a movement pattern that is produced by the actions of both the mandible and the lower lip.

(5) Locating the tracking probe near the angle of the mandible reduces facial muscle contamination of the movements but increases the likely-hood of contamination from tongue movements.

(6) Conclusions about the basic nature of articulatory function in man which are drawn from measurements of lip-contaminated mandibular activity are suspect.

(7) The posterior undersurface of the chin is a good location for tracking the mandible if the speech activity of interest includes no labial sounds. It is also a good location if the intent is to compare timing patterns of articulation between subjects or groups of subjects.

Acknowledgments

The author wishes to thank G. Fant who created a fertile and inspiring research atmosphere, made available the specialized facilities of the Department of Speech Communication, and provided encouragement during the investigation and constructive comments on the manuscript. Thanks are extended to J. Lindqvist, L. Andersson, and T. Murray for obtaining the special materials needed to construct the system, and to S. Jansson for help in the construction. A special note of gratitude goes to P. Branderud for designing the amplification system for the strain gage bridge. Finally, the author is indebted to J. Sundberg and B. Lindblom for their valuable discussions on the movements of the jaw.

References


