2 Methods

2.1 The Cineradiographic Procedures

The 35 mm cineradiograph from which the data for this report have been taken was made by Dr. Sven Öhman and Professor Kenneth Stevens at the cineradiographic facility of the Wenner-Gren Research Laboratory at Norrtull's Hospital in Stockholm, Sweden. The facility was made available to Dr. Öhman and Professor Stevens by Dr. Carl Wegelius, director of the Wenner-Gren Laboratory. This cineradiographic equipment for studies of speech production had previously been assembled by Dr. H. M. Truby (1962). The equipment was modified principally to provide an improved acoustic environment for sound recordings and to provide an arrangement for controlling the positioning of the subject (Stevens and Öhman, 1963).

The general configuration of the facility is shown in Figure 2.1. The subject sat in a booth which was slightly larger than a telephone booth. Synchronizing pulses were generated by a switch in the camera indicating the instants at which the shutter opened. The speech signal and the synchronizing pulses were recorded on a two-channel tape recorder. The interfering noise from the camera, reduced to a relatively low level at the microphone, is not visible on the spectrograms made from the sound recordings. The spectrograms were made by allowing the synchronizing pulses to pass through a high-pass filter thus producing vertical lines at the top of the spectrograms. These marks, labeled with

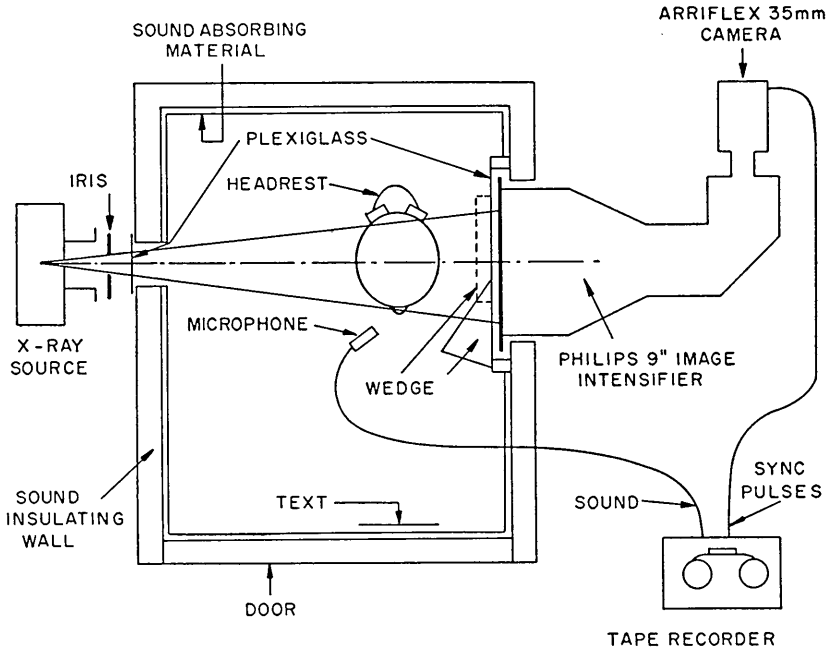

Figure 2.1 Schematic representation of the cineradiographic and sound-recording facility.

numbers corresponding to frame numbers, made it possible to align the film with the spectrograms. The frame rate was about 45 frames per second.

By accurately positioning the subject, it was possible to visualize the entire vocal tract including the vocal folds, the velum, and the lips. In order to minimize gross movements (due to gesticulation) the subject placed his head in a dental headrest. Soft tissue outlines were enhanced by painting a line of barium adhesive compound on the mid-line of the lips and the dorsum of the tongue. In addition, lead pellets were attached to the corner of the mouth, and tongue tip, and the dorsum of the tongue in the velar region. Contrast was improved by limiting the size of the X-ray beam with a rectangular diaphragm or iris to minimize the amount of scattered radiation, and by mounting wedgelike structures between the subject and the image-intensifier surface. The wedges served to reduce the intensity of the portions of the X-ray beam that passed through the lips and pharynx where the tissues are relatively radiolucent.

A line of lead pellets spaced one centimeter apart was fixed to the surface of the image intensifier in one corner of the frame. In addition to giving an idea of the magnification, images that resulted provided a stationary reference from which all motions could be measured. To help align the film with the tape recording, a coin was dropped on a small platform fixed to the image-intensifier surface at the beginning of the film. The sound of the coin's contact with the platform could be correlated with a calculation of the exact time of impact from the trajectory of the falling coin's image in several successive frames of the film. (The stationary coin resting on the platform provided additional reference for measurement.) Frames on the film were numbered in sequence beginning with the frame corresponding to the time of impact of the coin.

In order to improve further the delineation of various structures, the original film was processed on a logEtronic printer (St. John and Craig, 1957). This device makes a contact print of the original film by exposing it with a scanning spot of light. The light spot is produced by a cathode ray tube, and is optically focused on the film. The intensity of the light spot is monitored as it passes through the two film layers by a photosensing device. The output of the phototube controls the intensity of the scanning beam through an inverse feedback filter, thereby increasing the exposure to the dense portions of the negative and decreasing the exposure to the thin portions. The result is a positive that has a decrease in over-all contrast and an increase in detail contrast with greater border definition.

> In an ordinary print, highlight and shadow regions within a roentgenogram would be printed at opposite extremes of the density scale, where the slope of the [density–log exposure] curve is flat (low contrast) and detail merges with the background. However, these same two regions in a logEtronic print are reproduced near mid-scale where the curve is steep (high contrast). Hence, by bringing the two regions closer together in average density, gross contrast is reduced while detail contrast is maximized, resulting in better visualization throughout [St. John and Craig, 1957, p. 127].

The film was processed on a logEtronic printer especially designed for use with 35 mm cine by the Satellite Tracking Service of the U.S. Weather Bureau, Washington, D.C. (The cooperation of Messrs. William Plew and Michael Petrick of the U.S. Weather Bureau is gratefully acknowledged.) A typical frame from the reprocessed film is shown in Figure 2.2.

Near the end of the film, a metal grid of one centimeter squares was put in place of the subject at a location corresponding with his

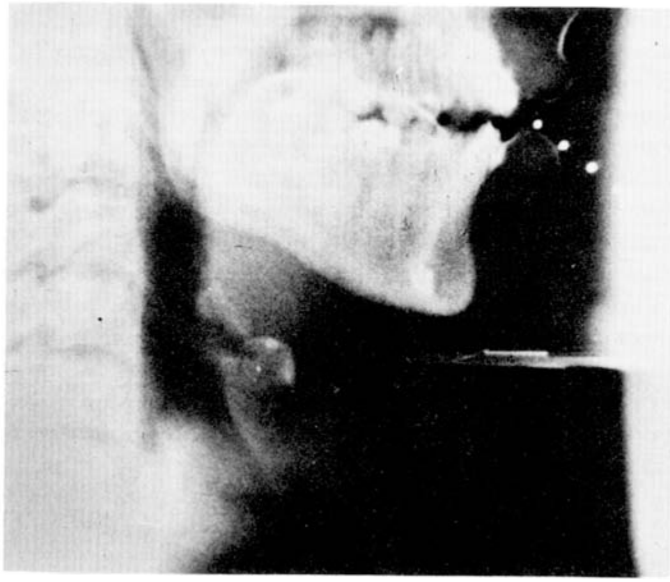

Figure 2.2 Frame number 100 from the logEtronic copy of the original cineradiograph.

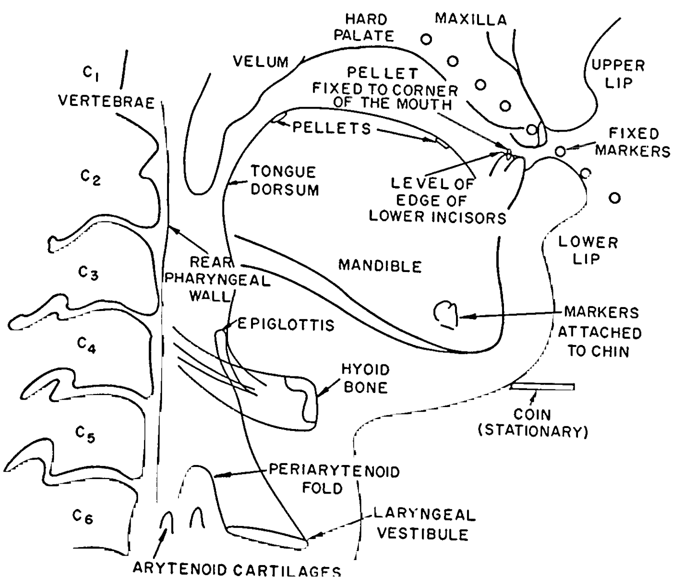

Figure 2.3 Tracing of frame number 100 with the template outlines drawn in and the vocal-tract structures labeled.

midsaggital plane. The image on the film shows the distortion resulting from the divergent X-rays striking the curved surface of the image intensifier. As a result of the distortion, the magnification from the midsaggital plane to the tracing (made using a projector as described in Section 2.3) varies from 1.35 at the center to 1.45 at the edge. An average magnification factor of 1.4 was used in calculating actual distances from the tracings. Because the measurements at different locations were not compared quantitatively, and because the variation was not too great, the distortion has not been taken into account.

The utterances were spoken with a short pause between each one. Each of the utterances is approximately 0.5 to 0.7 sec long and covers about 30 frames on the film. The subject was Kenneth Stevens, a speaker of General American English.

2.2 The Speech Material

The speech material used for the film consisted primarily of a series of nonsense utterances constructed to illustrate the production of specific phonetic segments in various phonetic environments. The material also included two sentences. (See Appendix A for a complete list of the utterances.) Thirteen utterances, each composed of an unstressed syllable followed by a stressed syllable (of the form /hə′CV/), were chosen for this study. Seven of the thirteen have the form /hə′tV/, in which the different vowels are /i, ɪ, ɛ, æ, u, ʊ, ɑ/. It should be noted that the lax vowels /ɪ, ɛ, ʊ/ (and /æ/) do not occur in open syllables in English, and there is a possibility that some of the resulting vocal-tract behavior for the utterances containing these vowels is not natural for the speaker. However, in the discussion of the results it has been assumed that most of the observed vocal-tract behavior for the utterances containing the lax vowels in open syllables is natural, and would be found for these vowels in closed syllables. The six remaining utterances have the form /hə′Cɛ/; the different consonants are /d, s, z, n, k, p/.

2.3 Tracing and Measuring Techniques

Frame-by-frame tracings were made for the thirteen utterances. The image was projected onto the rear of a sheet of high quality velum paper taped to a clear glass plate on a Tage-Arnø analyzing projector. The tracing was made directly on the sheet of velum.

In order to facilitate the tracing and measuring procedures, templates corresponding to the outlines of the vertebrae, maxilla, and mandible were made. By overlaying tracings of these structures from several randomly selected frames, it was found that the images of the hard structures varied slightly from one frame to the next. The inconsistency is attributed to slight shifts in the positions of the structures, to slight movements of the vertebrae relative to one another, and to a lack of constancy in the X-ray beam and the resulting image. The templates were therefore made by averaging the outlines of the structures from several frames.

It was difficult to see the midsaggital outline of the hard palate in the cineradiograph. To obtain an accurate representation of this outline, a dental cast of the subject's maxilla was made. The case was sectioned saggitally and the resulting outline was magnified and superimposed on the maxillary template. This procedure has not proved to be as accurate as would be desired, because on many of the tracings made of the postdental consonants, the tongue-tip outline overlaps the template outline.

The templates corresponding to the outlines of vertebrae, maxilla, and mandible were reproduced on a transparent photographic film. The position of each structure could then be indicated on each tracing by placing the template with the best possible fit over the image and making marks on the tracing paper through two small holes cut in the template. Thus the location of the vertebrae, maxilla, and mandible could each be marked with two points on each tracing. In order to have fixed reference points on each tracing, the outlines of the pellets and coin fixed to the image intensifier were traced. The following could be seen with varying degrees of clarity and reproducibility: the outlines of the upper and lower lips, the velum, the rear wall of the pharynx, both arytenoid cartilages, the vestibule of the larynx, the epiglottis, the hyoid bone, the entire dorsal mid-line of the tongue and the pellets at the tongue tip, the corner of the mouth, and the velar region of the tongue. All of these structures were traced whenever possible. A typical tracing with these structures labeled and the template outlines drawn in is shown in Figure 2.3.

2.4 Correction of Hard Structure Template Location

The marks indicating the position of each template are shown in Figure 2.4. The location of each of the six marks (X_1, X_2, Y_1, Y_2, Z_1, Z_2) was determined using x and y coordinates with origins fixed in relation to the stationary markers on the image-intensifier surface. Plots of the x and y coordinates for each of the marks versus time were made. There

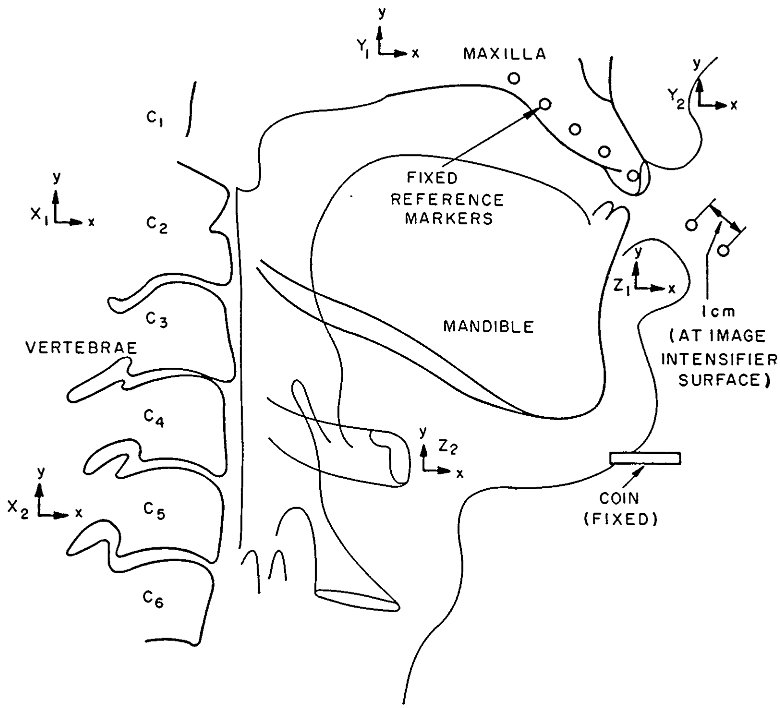

Figure 2.4 A typical tracing showing the outlines of the templates for the vertebrae, maxilla and the mandible with the points used to mark their location on the tracings (X_1, X_2, Y_1, Y_2, Z_1, Z_2), and the coordinate systems for locating the points with respect to the fixed reference markers.

was found to a 1–2 mm scatter in the location of the points on these plots. In order to position the templates more accurately, a smooth curve was drawn through the points for each plot of a coordinate versus time; and corrected values for the location of each mark were used to relocate the marks on the tracings indicating the template position. The gross movements of the maxilla, mandible, and vertebrae are discussed in Section 3.2.

2.5 Description of Measurements of Movements of the Various Organs of the Vocal Tract

From each tracing, measurements were made of the location of all the structures which could be seen and traced with a reasonable degree of accuracy. (It was found that the accuracy increased greatly with experience.) The axis for each measurement was fixed to one of the three templates; and the choice of the axes was based on observations of the film, measurements chosen by previous authors (Öhman and Stevens, 1963) and knowledge of the anatomy. Originally, some of the measurements of the location of a structure were made with respect to more than one template. Of the thirty-five measurements made on each tracing, only about twenty have proved to be of reasonable accuracy and to contain meaningful information.

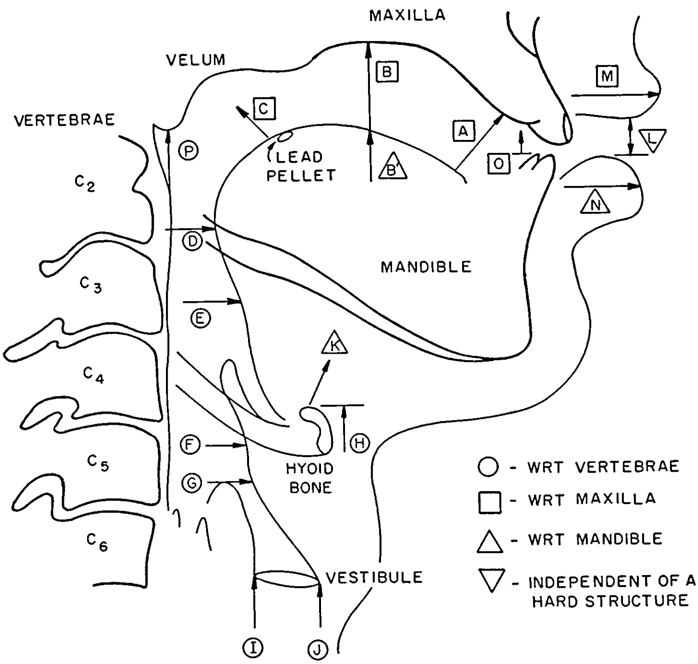

Figure 2.5 A typical tracing showing the measuring axes and the template outlines.

Figure 2.5 shows a typical tracing with the template outlines and measuring axes drawn in. The following measurements were made along axes fixed to the maxilla:

1. M, the horizontal location of the most anterior portion of the upper lip, or "protrusion of the upper lip."

2. O, the vertical location of a horizontal line tangent to the lower incisors, or "mandible height."

3. A, the intersection of the tongue tip contour with a line passing through the alveolar ridge. (This axis, and axes B, B′, C, and D and K are all radii of a circle which is tangent to the horizontal portion of the hard palate and to the vertical rear wall of the pharynx in the region of the second cervical vertebra, C_2.)

4. B, the intersection of the dorsum of the tongue contour with a vertical axis passing through the hard palate, or "tongue height."

5. C, the intersection of the dorsum of the tongue body with a line passing through the velum.

Axes fixed to the mandible were used as references for the following measurements:

1. N, the horizontal location of the most anterior portion of the lower lip, or "protrusion of the lower lip."

2. B′, the intersection of the tongue contour with a vertical line is approximately the same location as axis B, or "tongue height with respect to the mandible."

3. K, the distance between the hyoid bone and the mandible along a line passing through the body of the hyoid bone and the center of the circle defined by the hard palate and C_2.

The following measurements were made with respect to axes fixed to the vertebrae:

1. P, the vertical location of the lowermost portion of the uvula, or "velum height."

2. D, the horizontal location of the dorsum of the tongue in the pharynx close to the junction of C_2 and C_3, or "upper pharynx width."

3. E, the horizontal location of the dorsum of the tongue in the pharynx at a level closer to the glottis, or "lower pharynx width."

4. F, and

5. G, the horizontal locations of the dorsum of the epiglottis at two levels.

6. I, and

7. J, the vertical locations of the posterior and anterior ends of the larygeal vestibule, or "larynx height."

8. H, the vertical location of the uppermost part of the contour of the body of the hyoid bone, or "hyoid height."

The one measurement made independently of a hard structure is L, the vertical distance between the lowermost part of the upper lip contour and the uppermost part of the lower lip contour, or "height of the lip aperture."

The measurements were made by laying the tracing over a copy of each template with the axes drawn in and calibrated in millimeters.

The tracing was aligned with the template by superimposing the locating marks. These measurements were the main form of the data. The value of each of these measurements from each frame was plotted against time for each utterance. With the aid of the synchronizing pulses on the spectrograms, phonetic segment boundaries were marked on the plots. The graphs that resulted show the "behavior" or displacement versus time of the measurements for a particular utterance. The complete set of graphs along with the spectrogram and mid-phoneme tracings are given for each of the 13 utterances in Appendix B.

In order to compare behavior of a measurement among several utterances, the graphs of that measurement were overlaid and aligned at a time corresponding to consonant release. The criterion for "consonant release" for each consonant is also given in Appendix B. This alignment presents some problems (referred to in Chapter 3), but in general it is a convenient procedure to use for the purpose of this report. The results and interpretations of this type of comparison are presented in the remainder of this monograph.

Additional data are presented in the form of superimposed mid-phoneme tracings. In general, the comparisons have been made among the members of the groups of utterances in which either the consonant (/hə′Cɛ/) or the vowel (/hə′tV/) is different.