1 Laboratory Techniques for Investigating Speech Articulation

MAUREEN STONE

This chapter discusses current laboratory techniques that measure the oral vocal tract during speech. The focus is on instruments that measure the articulators directly and indirectly. Indirect measurements come from instruments that are remote from the structures of interest such as imaging techniques. Direct measurements come from instruments that contact the structures of interest, such as, point-tracking devices and electropalatography. Although some references are made to current research using each instrument, to indicate its applications and strengths, the list of studies is not comprehensive as the goal is to explain the instrument.

Measuring the vocal tract is a challenging task because the articulators differ widely in location, shape, structural composition, and speed and complexity of movement. First, there are large differences in tissue consistency between soft tissue structures (tongue, lips, velum) and hard tissue structures (jaw, palate), which result in substantially different movement complexity. In other words, the fluid deformation of the soft structures and the rigid movements of the bones need different measurement strategies. Second, measurement strategies must differ between structures visible to superficial inspection, such as the lips, and structures deep within the oral cavity, such as the velum. Third, articulator rates of motion vary, so that an instrument with a frequency response appropriate for the slow-moving jaw will be too slow for the fast-moving tongue tip. The final and perhaps most important measurement complication is the interaction among articulators. Some articulatory behaviors are highly correlated, and distinguishing the contributions of each player can be quite difficult. The most dramatic example of this is the tongue–jaw system. It is clear that jaw height is a major factor in tongue tip height. However, the coupling of these two structures becomes progressively weaker as one moves posteriorly, until in the pharynx, tongue movement is only minimally coupled to jaw movement if at all. Thus, trying to measure the contribution of the jaw to tongue movement becomes a difficult task.

It is difficult to devise a transducer that can be inserted into the mouth, which will not in some way distort the speech event. Thus, the types of instruments

The Handbook of Phonetic Sciences, Second Edition. Edited by William J. Hardcastle, John Laver, and Fiona E. Gibbon.

used in the vocal tract need to be unobtrusive, such as by resting passively against a surface (e.g., electropalatography), by being small and positioned on noncontact surfaces (e.g., pellet tracking systems), or by not entering the vocal tract at all (e.g., imaging techniques).

Instruments that enter the oral cavity must meet certain criteria. They need to be unaffected by temperature change, moisture, or air pressure. Affixatives must be unaffected by moisture, nontoxic, able to stick to expandable, moist surfaces, and must be removable without tearing the surface tissue. Devising instruments that are noninvasive, unobtrusive, meet the above criteria, and still measure one or more components of the speech event is so difficult that most researchers prefer to study the speech wave and infer physiological events from it. However, since those inferences are based on, and refined by, physiological data, it is critical to add new physiological databases, lest models of the vocal tract and our understanding of speech production stagnate.

In recent times, physiological measurements have improved at an extraordinary pace. Imaging techniques are revolutionizing the way we view the vocal tract by providing recognizable images of structures deep within the pharynx. They also provide information on local tissue movement and control strategies. Point-tracking systems and palatographic measurements have transformed our ideas about coarticulation by revealing inter-articulator relationships that could only in the past be addressed theoretically. Applications to linguistics and rehabilitation are now ongoing. This chapter considers indirect measurements, that is, imaging techniques, and direct measurements such as point-tracking techniques, and tongue–palate measurement devices

1 Imaging Techniques

The internal structures of the vocal tract are difficult to measure without impinging upon normal movement patterns. Imaging techniques overcome that difficulty because they register internal movement without directly contacting the structures. Four well-known imaging techniques have been applied to speech research: X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. Imaging systems provide recordings of the entire structure, rather than single points on the structure.

1.1 X-ray

X-ray is the most well known of the imaging systems. It is important because it was the first widely used imaging system and most of our historical knowledge about the pharyngeal portion of the vocal tract came from X-ray data. To make a lateral X-ray image, an X-ray beam is projected from one side of the head through all the tissue, and recorded onto a plate on the other side. The resulting image shows the head from front to back and provides a lengthwise view of the tongue.

A frontal or anterior–posterior (AP) X-ray is made by projecting the X-ray beam from the front of the head through to the back of the head and recording the image on a plate behind the head. The resulting images provide a cross-sectional view of the oral cavity. Prior to the advent of MRI considerable research was done using X-ray imaging. More recent X-ray studies are based on archival databases.

X-ray data have contributed to many aspects of speech production research. Many vocal tract models are based on X-rays (cf. Fant, 1965; Mermelstein, 1973; Harshman et al., 1977; Wood, 1979; Hashimoto & Sasaki, 1982; Maeda, 1990). X-rays have also been used to study normal speech production (Kent & Netsell, 1971; Kent, 1972; Kent & Moll, 1972), nonspeech motions (Kokawa et al., 2006), motor control strategies (Lindblom et al., 2002; Iskarous, 2005), language differences (cf. Gick, 2002b; Gick et al., 2004), and speech disorders (Subtelny et al., 1989; Tye-Murray, 1991).

Usually soft tissue structures such as the tongue are difficult to measure with X-rays, because the beam records everything in its path including teeth, jaw, and vertebrae. These strongly imaged bony structures obscure the fainter soft tissue. Another limitation of X-ray is that unless a contrast medium is used to mark the midline of the tongue, it is difficult to tell if the visible edge is the midline surface of the tongue or a lateral edge. This is particularly problematic during speech, because the tongue is often grooved or arched. Finally, the potential hazards of overexposure have reduced the collection of large quantities of X-ray data. There is, however, public availability of archival X-ray databases for research use. One such database (Munhall et al., 1994a, 1994b) was compiled by Advanced Technologies Research Laboratories, Kyoto, and is available from http://psyc.queensu.ca/~munhallk/05_database.htm.

1.2 Tomography

Tomography is a fundamentally different imaging method from projection X-ray in that it records slices of tissue. Three tomographic techniques used in speech research are Computed Tomography, Magnetic Resonance Imaging, and Ultrasound Imaging. These slices are made by projecting a thin, flat beam through the tissue in one of four planes: sagittal, coronal, oblique, and transverse (see Figure 1.1). The mid-sagittal plane is a longitudinal slice, from top to bottom, down the median plane, or midline, of the body (dashed line – upper right). The para-sagittal plane is parallel to the midline of the body and off-center (not shown). The coronal plane is a longitudinal slice perpendicular to the median plane of the body. The oblique plane is inclined between the horizontal and vertical planes. Finally, the transverse plane lies perpendicular to the long axis of the body, and is often called the transaxial, or in MRI, the axial plane.

1.2.1 Computed Tomography (CT) Computed Tomography uses X-rays to image slices (sections) of the body as thin as 0.5 mm or less. Tomographic images

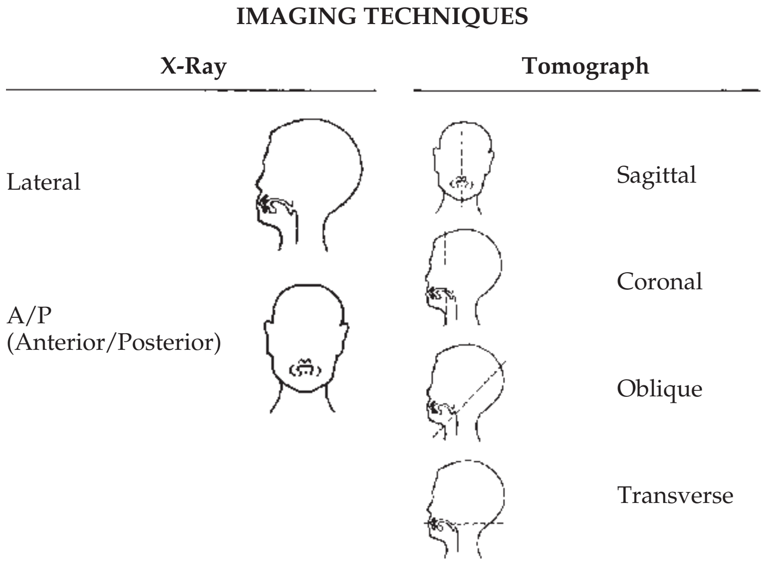

Figure 1.1 Scan types used in through-transmission and tomographic imaging. There are two X-ray angles contrasted with four tomographic scanning planes.

made in coronal planes are made by projecting very thin X-ray beams through a slice of tissue from multiple origins. The scanner rotates around the body taking these images and a computer creates a composite, including structures that are visible in some scans but obscured in others. Using this technique, tissue slices can be collected rapidly, 15 Hz or faster, and multiple slices can be collected simultaneously. CT images soft tissue more clearly than X-rays because it produces a composite X-ray. By digitally summing a series of scans, the composite section has sharper edges and more distinct tissue definition. From the multislice datasets, planar sections can be reconstructed in any direction. CT images can produce excellent resolution of soft and hard tissue structures. Figure 1.2, for example, is a reconstructed image of the midsagittal plane of the vocal tract. Bone appears bright white in the image, soft tissue structures are gray. In this figure, the junction of the velum and hard palate can be seen to be quite complex. The soft tissue below the hard palate widens before the velum emerges as a freestanding object. It is clear from this image that the shape of the palatine bone is not well reflected in the soft tissue. Measures of the palate bone made from an MRI or ultrasound image will differ from measurements made directly in the mouth or from dental impressions. Without this image, those differences would be hard to interpret.

Another method of CT data collection is Spiral CT. Spiral CT collects multiple slices at the same time by collecting a single spiral-shaped slice instead of multiple flat planar slices. In the mid 1980s, the cable and drum mechanism for

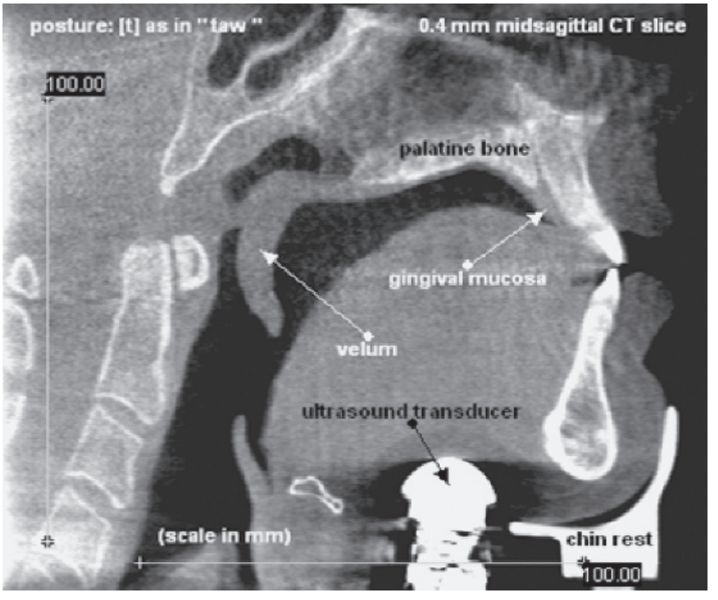

Figure 1.2 Midsagittal CT of vocal tract reconstructed from axial images. Bone is white; soft tissue is gray. (Reproduced courtesy of Ian Wilson)

powering the rotation of the CT machine was replaced with a slip ring. The slip ring allows the CT scanner to rotate continuously, creating a spiral image. Spiral CT scans have very high resolution, but currently take 20–30 seconds to create, and hence are too slow for imaging continuous speech, though excellent for static images (Lell et al., 2004).

Electron Beam CT was developed to measure calcium deposits around coronary arteries. Its principles are similar to CT, but it uses an electron "gun" instead of regular X-ray. EBCT collects a set of parallel images that are reconstructed as a 3D volume. EBCT is a fast acquisition technique and therefore has been used to collect vocal tract images for datasets requiring short acquisition times. For example, Tom et al. (2001) scanned the entire vocal tract in under 90 seconds, to compare vocal tract shapes during falsetto and chest registers.

Although CT has been used to image the vocal tract, it is not the instrument of choice for speech research because of radiation exposure and because MRI provides much the same information, albeit at a lower spatial and temporal resolution. In fact, the major limitation of CT is that it has more radiation exposure than traditional X-ray, because it images thinner slices, and each slice is scanned several times to collect multiple images. Another limitation is that the subject is supine or prone, so gravitational effects on the subject differ from upright. On the positive side, 3D reconstructions can be made and sliced in any plane, and images are clear and easy to measure.

1.2.2 Magnetic Resonance Imaging (MRI) Another tomographic technique is Magnetic Resonance Imaging, which uses a magnetic field and radio waves rather than X-rays to image a section of tissue. There are a number of MRI procedures that yield a variety of information: high-resolution MRI, cine MRI, tagged-snapshot MRI, tagged-cine MRI, diffusion tensor MRI, and functional MRI. All of these use identical hardware: typically 1.5 or 3 Tesla machines. The differences lie in the software algorithms, which are designed to exploit different features of the relationship between the hydrogen proton, magnetic fields, and radio waves.

An MRI scanner consists of electromagnets that surround the body and create a magnetic field. MRI scanning detects the presence of hydrogen atoms, which occur in abundance in water and, therefore, in human soft tissue. Figure 1.3 depicts the MRI process. Picture (a) represents hydrogen protons spinning about

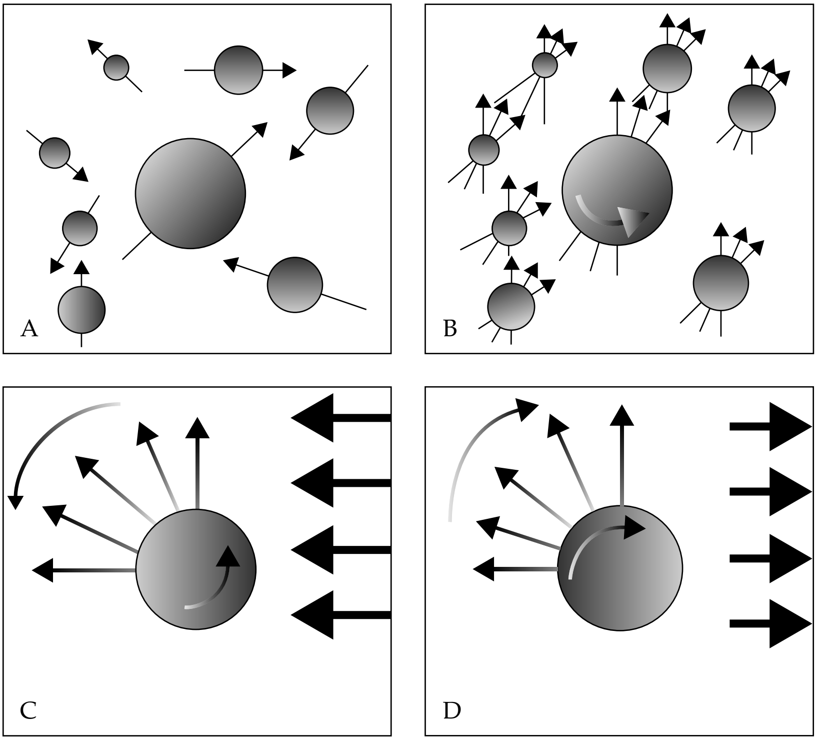

Figure 1.3 MRI recording of the amount of hydrogen in tissue. Hydrogen protons spin about axes that are oriented randomly (A). The MRI magnet causes them to align to the long axis of the body, but with a small precession (wobble) (B). A radio-frequency pulse knocks them out of alignment (C). As the protons realign to the magnet (D) they emit a radio pulse that is read by the scanner.

their axes, which are oriented randomly. (b) shows what happens when a magnetic field is introduced. The protons' axes align along the direction of the field's poles. Even when aligned, however, the protons wobble, or precess. In (c) a short-lived radio pulse, vibrating at the same frequency as the precession, is introduced. This momentarily knocks the proton out of alignment. (d) shows the proton realigning, within milliseconds, to the magnetic field. As the proton realigns, it emits a weak radio signal of its own. Period (d) is when the MR image is "read." The radio signals are summed until the protons return to position (b). The resulting data are constructed into an image that reflects the hydrogen content (i.e., the amount of water or fat) of the different tissues. Because the proton emissions are weak, the process is repeated many times and the data are summed into a single image. If the process is repeated for several minutes, while the subject holds still, high-resolution images result.

MRI measurement of oral structures has replaced X-ray for many research applications. MR images have been used to detail developmental vocal tract anatomy and function (Xue & Hao, 2003; Vorperian et al., 2005). MRI also has provided quite accurate extraction of vocal tract surfaces (Story et al., 1996). These surfaces have been used to calculate 3D vocal tract volumes for modeling geometry to acoustic relationships (Tameem & Mehta, 2004; Story, 2005). Extracted edges have also been used to model 3D structures within the vocal tract. Serrurier and Badin (2005) modeled velar position for French vowels from MRI and CT images. Engwall (2003) modeled tongue position for Swedish vowels from MRI, Electromagnetic Articulography (EMA), and Electropalatography (EPG). Story et al. (1996) modeled vocal tract airway shapes for 18 English phonemes from MRI. MRI is very good at characterizing different types of soft tissue and therefore is quite successful in identifying tumors and soft tissue pathology. For example, Lenz et al. (2000) used MRI and CT together to stage oral tumors and Lam et al. (2004) had good success using MRI T1 and T2 weighted images to determine tumor thickness.

Two types of MRI are used particularly to characterize tissue: high-resolution MRI (hMRI) and diffusion tensor MRI (DTI). Figure 1.4 shows a high-resolution sagittal MRI image of the vocal tract at rest. The vocal tract appears black, as do the teeth, since neither contains water. Water and fat, both of which are high in hydrogen, are found in marrow, seen in the palate and mandible. Muscles are visible in the tongue, velum, and lips. The other method of characterizing soft tissue is diffusion tensor MRI (DTI), which measures 3D fiber direction, typically in *ex-vivo* structures. DTI, developed in the early 1990s, visualizes fiber direction by measuring random thermal displacement of water molecules in the tissue. The direction of greatest molecular diffusion parallels the local fiber direction. DTI has virtually microscopic spatial resolution and distinguishes tissue fibers with their orientations for any muscles. A fiber map can be drawn and superimposed on an MRI structural image. The fiber map is 3D and can differentiate among nerve fiber pathways and detail anatomical structures based on their fiber architecture. There are limitations of this technique that impede the measurement of oropharyngeal structures. First, when fiber directions cross within a

Figure 1.4 High-resolution MRI (hMRI) of the midsagittal vocal tract at rest.

single voxel (3D pixel), visualization of the underlying fiber structure is reduced. Fiber interdigitation is typical in oral musculature, especially the tongue, lips, and velum. Second, DTI is sensitive to motion and the structure must remain immobile for several minutes to record a volumetric scan. Using long collection times, DTI has been used to study the excised tongues of animals (Wedeen et al., 2001) and humans (Gilbert & Napadow, 2005). In addition, DTI can be used *in vivo* with cooperative subjects to collect data in as little as 3–5 minutes. Figure 1.5 shows a fiber map indicating the fan-like fibers of the genioglossus muscle, which run from superior–inferior to anterior–posterior in direction. This image was taken from an *in vivo* human tongue at rest (Shinagawa et al., 2008, 2009).

When measuring vocal tract motion, Cine-MRI is of particular interest. Cine-MRI is similar to other cine techniques, such as videofluoroscopy or movies, in that it divides a moving event into a number of still frames. Because MRI sums proton emissions over time, it typically takes a long time to reconstruct a single image, and collecting data during speech motion is challenging. Cine-MRI is often done by having the subject repeat a task multiple times and summing data from each frame across repetitions, similar to ensemble averaging. This technique has been used to compare vocal tract behaviors during speech production (Magen et al., 2003), especially vowel production (Hasegawa-Johnson et al., 2003; Story, 2005; McGowan, 2006). However, the subject must produce the repetitions very

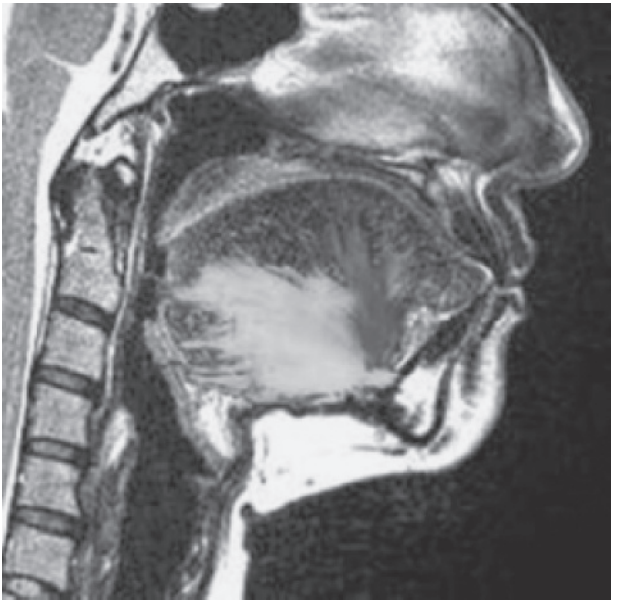

Figure 1.5 Diffusion Tensor Image shows fiber directions of the genioglossus muscle (light gray). Fibers are overlaid on a high resolution image of the head.

precisely to prevent image blurring. It is also possible to compare MRI vocal tract data to speech acoustics, either by collecting the speech wave independently from the (quite noisy) MRI data collection, or by using subtraction microphones within the MRI machine itself (cf., NessAiver et al., 2006). Some Cine-MRI data have been collected with single repetitions of a task (Mády et al., 2002; Narayanan et al., 2004). These images usually have a reduced frame rate or a reduced image quality, but typically have sufficient spatial resolution to extract surfaces of vocal tract structures. Faster frame rates and good image quality require several repetitions per slice (Stone et al., 2001a). Rapid collection of MRI frames using a single repetition has been reported using spiral MRI (Narayanan et al., 2004). This method acquires images in interleaved spirals rather than planes. Although Cine-MRI does not produce the quality of anatomy seen in high resolution (hMRI) images, the speed of pellet-tracking systems, or the level of muscle detail seen in DTI, it is able to answer many speech questions that were previously unstudied. Therefore, Cine-MRI is very popular in the study of speech.

Concurrent with the development of Cine-MRI, was the development of Tagged Snapshot MRI. Tags are created by applying a spatial gradient to the tissue, which demodulates the spinning protons in alternating planes. In the demodulated planes the protons spin out of phase with the rest of the tissue. The demodulated protons are invisible to the machine when the image is read, thus the invisible proton planes appear as black stripes on the image (see Figure 1.6). After tags are

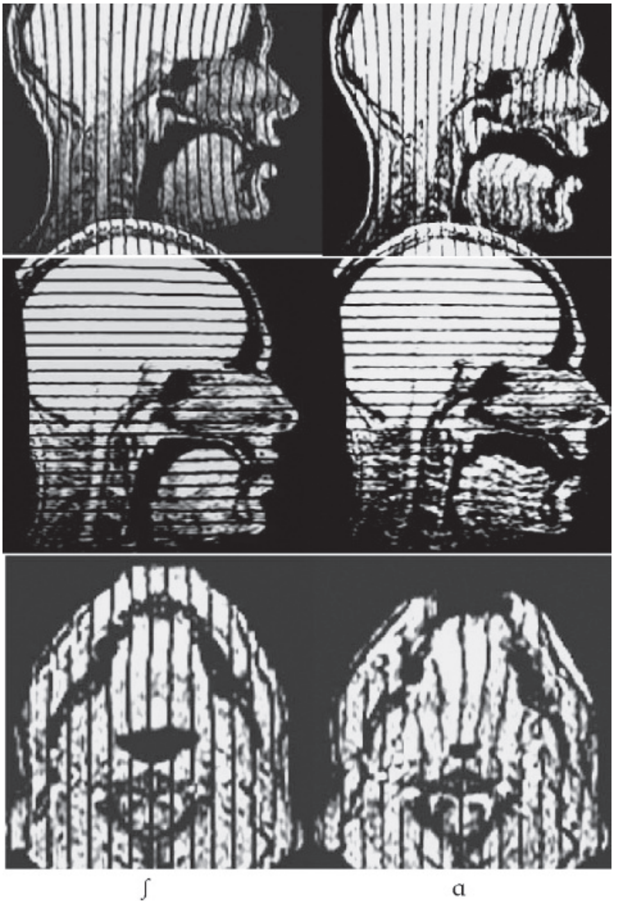

Figure 1.6 Tagged MR images in three planes. The left column shows the reference state of the tongue, just after the tags were applied during the sound /ʃ/ The right column shows the tongue in the deformed state, after the tongue has moved into /ɑ/ position.

applied to the tissue, the object of interest, such as the tongue or lips, is moved. Because magnetization stays with the tissue, the tags deform to exactly the same extent as the tissue. Figure 1.6 shows reference frames in three planes taken during /ʃ/ and deformed frames taken during the /ɑ/ in "sha." In Tagged Snapshot MRI only two images are read. One is before the motion (reference) and the other is during or after the motion (deformed). A grid of tags is created by applying tags in horizontal and vertical planes in immediate succession prior to reading the image, or by combining two datasets of orthogonal tags collected separately. Niitsu et al. (1994) examined tag positions from rest (before) and a vowel position (after) to derive the direction of the movement. Napadow et al. (1999a, 1999b)

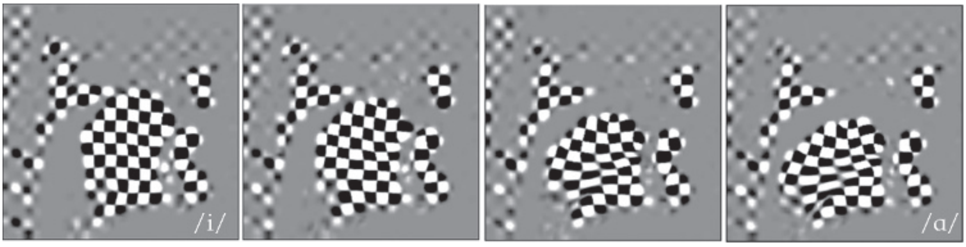

Figure 1.7 Tagged Cine-MRI sequence shows motion from /i/ to /ɑ/. Checkerboard deformation shows local expansion, compression, and shear.

studied swallowing and nonspeech tongue motions by having subjects repeat the task multiple times and each time collecting a deformed image at a later time. These images were then put into a pseudo-motion sequence reflecting the movement.

Cine-MRI and Tagged Snapshot MRI can be combined to form Tagged Cine-MRI (tMRI). tMRI captures internal tissue motion over time during the performance of a task. The first applications of tMRI were studies of the heart's motion and internal tissue characteristics (Zerhouni et al., 1988; Axel and Dougherty, 1989). Continuous tongue deformation during speech has also been measured using tMRI. Figure 1.7 shows deformation of the midsagittal tongue between the two vowels /i/ and /ɑ/. The black and white squares are a visualization device to better depict the deformations. The change in their shapes over time demonstrates features of local tissue deformation. From these images tags can be tracked to directly measure positions and motion of all tissue points in the tongue (Parthasarathy et al., 2007). From the tissue point motions one can calculate displacement, velocity, and local strain (compressions and expansions).

Functional MRI (fMRI) is a method of MRI scanning that measures changes in blood flow in the brain. Increased blood flow characterizes increased uptake of blood in the active region of the brain. Because blood is high in hydrogen, the local increase in activity can be measured by MRI. The premise is that spatially distinct, distributed areas of the brain are connected into functional networks organized to produce specific tasks. If these networks can be imaged, they can be mapped geographically to detail brain function during various behaviors. Increased neural activity in a region causes increased demand for oxygen, and that oxygen is brought to the region by blood. Replacement of deoxygenated blood with oxygenated blood produces a more uniform magnetic environment, which increases the longevity of the MRI signal. fMRI signals are snapshots that are collected at 10–20 Hz. Multislice recording of the brain and multiple repetitions can be recorded. Some limitations exist in the use of fMRI for speech research. First, the visible effects of the increase in oxygen occur some time after the event itself (0.5–8.0 seconds), which results in poor temporal resolution. Second, signals can vary even with no change in brain state. To overcome the latter problem,

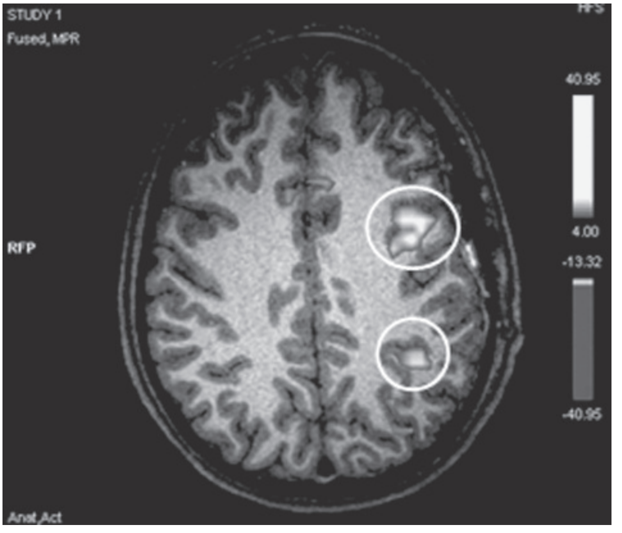

Figure 1.8 Axial fMRI scan showing regions of the brain that are active while subject thinks of words. The white regions (circled) indicate activity in the left hemisphere. (Photo courtesy of Rao Gullapalli, University of Maryland Medical School)

fMRI usually compares sets of images before and after the task. Despite these difficulties, research on speech and language is being done including studies of cortical aspects of speech production (Gracco et al., 2005), speech perception (Specht et al., 2005; Pulvermuller et al., 2006; Uppenkamp et al., 2006), and speech disorders (Ackermann & Riecker, 2004; Bonilha et al., 2006). Figure 1.8 shows an axial fMRI scan of the brain during a speech task. The task is to think of as many words as possible that use a specific letter, in a 24-second period. The task is repeated several times and modeled to bring out the contrast between the on and off states. Active regions are circled.

As with all instruments, MRI has several drawbacks. The first is the slow capture rate that results from summing the weak radio signals emitted by each proton. Thus, high-resolution images and DTI require long periods of immobility for a good image, and fMRI has a slow response time. Cine- and tMRI require summation of multiple, very precise repetitions for optimal images. A second drawback is the width of the section. Whereas CT sections are less than 1 mm wide, and ultrasound sections are less than 2 mm, MRI sections are usually 5–10 mm wide. A tomographic scan compresses a three-dimensional volume into two dimensions, which is like flattening a cylinder into a circle. For example, in a slice that is 5 mm wide, items that are actually 5 mm apart will appear to be in the same plane. Thus, in the transverse plane, the hyoid bone and epiglottis might appear in the same slice even though one is several millimeters below the other. Narrower widths in MRI sections are possible, but require longer exposure time. A third drawback is that many subjects, as many as 30 percent, experience claustrophobia and cannot tolerate the procedure. Fourth, metal clamps, tooth crowns, and steel implants quench the signal creating a diffuse dark spot surrounding the metal. Final drawbacks for MRI are that the subject must be lying supine or prone, which changes the location of gravity with respect to the oral structures and normal agonist–antagonist muscle relationships. Despite these nontrivial drawbacks, MRI's strengths make it an important instrument in speech production research.

1.2.3 Ultrasound Ultrasound produces an image by using the reflective properties of sound waves. A piezoelectric crystal stimulated by an electric current emits an ultra high-frequency sound wave. The crystal both emits a sound wave and receives the reflected echo. The sound wave travels through the soft tissue and reflects back when it reaches an interface with tissue of a different density, like bone, or when it reaches air. The best reflections are perpendicular to the beam (see Figure 1.9). In order to see a section of tissue rather than a single point, one needs an array transducer. In an array transducer, up to 128 crystals fire sequentially, imaging a section of tissue that is rectangular or wedge-shaped. The image size is proportional to the size of the transducer and frequency of the crystals, the wedge angle may be up to 140 degrees. The returning echoes are processed by an internal computer and displayed as a video image.

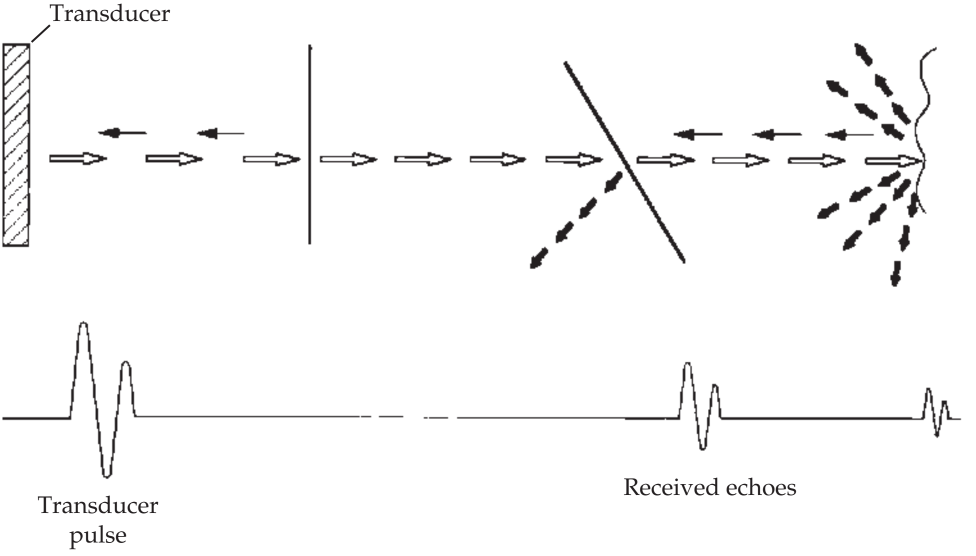

Figure 1.9 Schematic of ultrasound beam emitted from a transducer and reflected by surfaces of different angles. The best reflections are perpendicular to the beam. Soft tissue typically reflects and refracts sound in multiple directions.

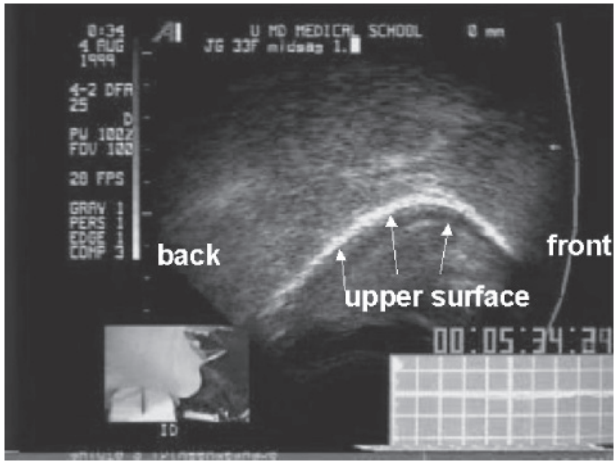

Figure 1.10 An ultrasound image of the sagittal (lengthwise) tongue. The white line is the upper surface of the tongue.

Figure 1.10 shows a sagittal image of the tongue in a 90-degree wedge-shaped scan. To create such an image, the transducer is placed below the chin and the beam passes upward through a 1.9 mm thick section of the tongue. When the sound reaches the air at the surface of the tongue, it reflects back creating a bright white line. The black area immediately below is the tongue body. The tongue surface is the lower edge of the white line. Interfaces within the tongue are also visible. Figure 1.11 depicts the tongue in coronal section. The tongue surface is thinner in cross-section and herein contains a small midsagittal depression. Measurement error on such images is at most 1 pixel (Stone et al., 1988). Although ultrasound is typically used to study tongue motion, it has been used on occasion to study other structures, such as the lateral pharyngeal wall (Parush & Ostry, 1993; Miller & Watkin, 1997), or vocal folds (Munhall & Ostry, 1985; Ueda et al., 1993).

The vocal folds may be imaged by placing the ultrasound transducer at the front of the neck, at the thyroid notch (Adam's apple), and pointing it directly back in the transverse plane. Glottal stops, tumors, and slow-moving behaviors can be seen this way. However, the vocal folds have a very fast vibration rate, at least 80 Hz, and usually much more. The sampling rate of the fastest ultrasound machines is about 90 Hz. Therefore, vibration may be seen during phonation, but it is undersampled, and in individual frames cannot be measured. Other instruments, such as Electroglottography, are more accurate methods for measuring vocal fold vibration.

Ultrasound presents specific challenges when measuring the tongue. The first challenge is that up to 1 cm of the tongue tip may not be captured in the image, because the ultrasound beam is reflected at the floor of the mouth and the sound wave doesn't enter the tongue tip. The tip may be imaged, however, if there is sufficient saliva in the mouth, if the tongue is resting against the floor, or if the

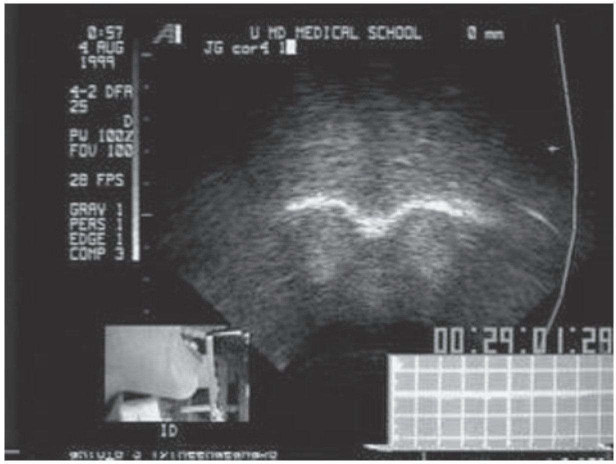

Figure 1.11 An ultrasound image of the coronal (crosswise) tongue. The upper surface has a midline groove.

transducer is posterior and angled forward (Stone, 2005). The second limitation is the inability to see beyond a tissue–air or tissue–bone interface. Since the tissue–air interface at the tongue's surface reflects the sound wave, the structures on the far side of the vocal tract, such as the palate and pharyngeal wall, cannot be imaged. Similarly, when ultrasound reaches a bone, the curved shape refracts the sound wave creating an acoustic shadow or dark area. Thus, the jaw and hyoid bones appear as shadows and their exact position cannot be reliably measured.

Despite these limitations, a large number of studies successfully use real-time ultrasound to study tongue movements. Normal speech production and exploration of tongue surface features are the most common applications, (cf. Davidson, 2006; Slud et al., 2002; Chiang et al., 2003). Ultrasound has also been the basis of tongue surface models in the sagittal plane (Green & Wang, 2003), coronal plane (Slud et al., 2002), and in 3D (Watkin & Rubin, 1989; Yang Stone & Lundberg, 1996; & Stone, 2002; Bressmann et al., 2005b). Because ultrasound is well tolerated by subjects it is well suited to the study of disorders (Bressmann et al., 2005a; Schmelzeisen et al., 1996) and to studies of children (Ueda et al., 1993). Ultrasound is also an excellent tool to study swallowing because it is noninvasive and does not affect the swallow (Miller & Watkin, 1997; Chi-Fishman et al., 1998; Watkin, 1999; Peng et al., 2000; Soder & Miller, 2002). Moreover, it is now possible to align tongue position with the hard palate, if the transducer and head are held still (Epstein & Stone, 2005). This alignment allows better interpretation of the ultrasound data. Finally, applications of ultrasound to linguistics are increasing because portable machines allow ultrasound to be used in fieldwork (cf. Gick, 2002a). Ultrasound provides large quantities of time–motion data with a single repetition, thin slices, and a noninvasive method. Many machines are now digital and can collect 90 or more scans per second, which is fast enough to measure most tongue motions, though not fast enough to measure vocal fold vibration. A second advantage is that ultrasound involves no known biological hazards, since the transduction process involves only sound waves.

Although most ultrasound machines scan a single tissue slice, some transducers scan multiple slices and reconstruct 3D volumes. Visualizing 3D tongue movements facilitates understanding of tongue surface motion. This is important because the tongue is anisotropic from front to back and medial to lateral. Therefore no single slice fully represents its motion. Some commercial ultrasound machines have 4D transducers that create 3D ultrasound volume sequences. These transducers contain a standard 2D array transducer, which sequentially scans a series of tissue slices by mechanically stepping through a series of angles, scanning a slice at each step. If the stepping is slow, the resulting scans and steps are combined into a static 3D volume. A 3D movie is created when the scanning, stepping, and combining are done very quickly, at multiple times per second. This technology is very promising even though at present the scan rates are fairly slow. To capture the entire tongue, a large volume must be imaged, and the scan rate is usually limited to about 6 Hz for a volume large enough to capture the entire tongue. Alternatively, multiple 2D ultrasound scans can be collected with a faster frame rate, and combined into a 3D tongue surface reconstruction as seen in Figure 1.12 (Stone & Lundberg, 1996; Bressmann et al., 2005b).

Ultrasound, like the other imaging techniques, captures the inaccessible parts of the tract (e.g., the pharynx), and measures planes rather than tissue points to provide comprehensive and detailed information about vocal tract structures.

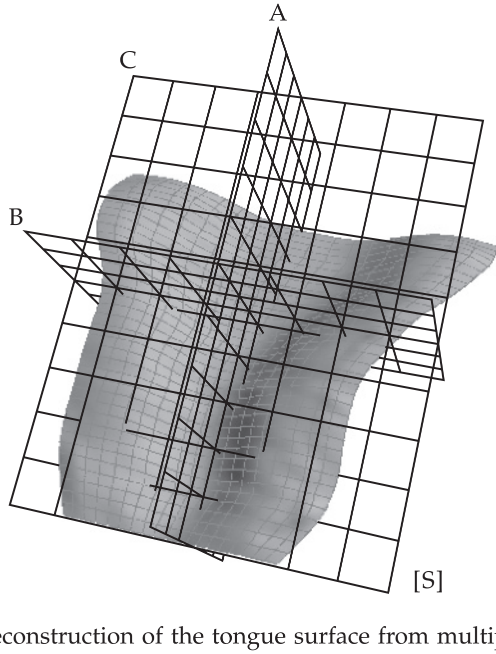

Figure 1.12 A 3D reconstruction of the tongue surface from multiple ultrasound slices.

Although the sampling rates of imaging techniques tend to be slower than point-tracking systems, imaging systems are now widely available in hospitals and laboratories. Moreover, portable machines, used in fieldwork, provide expanded and varied linguistic datasets.

2 Point-Tracking Measurements of the Vocal Tract

Point-tracking systems have different strengths from imaging techniques; they measure individual fleshpoints by affixing pellets to the articulators and tracking their movement over time. Typically, multiple articulators are measured at the same time, and tracking speed is fast, so that inter-articulator timing measures are quite good. Well-known tracking systems can be external to the oral cavity, such as Optotrak or Vicon, or both external and internal, such as the articulometer and the X-ray microbeam. External tracking systems can directly measure markers on the face and lips using video or light emitting diode (LED) tracking. Tracking systems operating within the vocal tract use pellets or receivers tracked by magnetic fields or X-rays.

2.1 Electromagnetic Articulometer (EMA)

Several names refer to similar point-tracking devices; these are Electromagnetic Midsagittal Articulometer (EMMA), Electromagnetic Articulometer (EMA), and Articulograph. Several such instruments, using roughly similar principles are currently in use (Perkell et al., 1992; Van Lieshout et al., 1994; Zierdt et al., 1999). The instruments track tissue-point movement in and around the oral cavity by measuring the movement of small receiver coils through alternating magnetic fields.

In the 2D EMA systems, three transmitter coils form an equilateral triangle in the midsagittal plane. They are suspended around the subject's head using a clear plastic assembly. Each transmitter coil is driven at a different sinusoidal frequency to generate alternating magnetic fields of different frequencies. Small insulated receiver coils are attached with adhesive to oral and facial structures of interest at midline. The alternating magnetic fields induce an alternating signal in the receiver coils. The voltage of this signal is inversely related to the distance between the transmitter and the receiver coil. A computer algorithm uses these distances to calculate the location of the receiver coil as it moves in x–y space over time. The best resolution in the field space is found in the center, i.e., in the oropharyngeal region, where measurement resolution is calculated at less than 1 mm.

EMA's strength is its ability to measure multiple articulators simultaneously, at a rapid sampling rate, making it popular in the fields of linguistics, speech motor control, swallowing, and speech pathology. In linguistics, studies have focused on articulation and inter-articulator programming, among other topics. In English, Byrd et al. (2005) examined effects of sentence position on consonant articulation. In German, Fuchs et al. (2006) studied differences in the control of

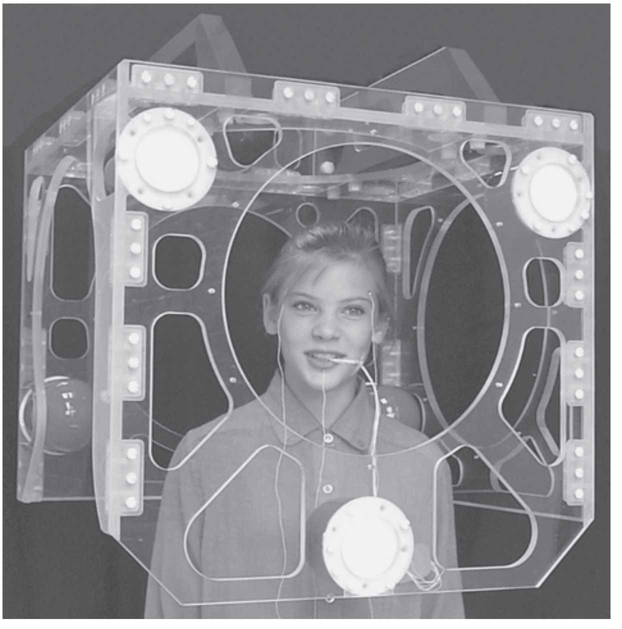

Figure 1.13 3D Electromagnetic articulograph tracks markers in the mouth using a magnetic field. (Photo courtesy of Carstens Medizinelektronik, Inc.)

fricatives and stops, and Kuhnert and Hoole (2004) found similar patterns of reduction in English and German. EMA studies have also effectively studied motor control (cf. Kaburagi & Honda, 1996; Van Lieshout & Moussa, 2000; Perkell & Zandipour, 2002; McClean & Tasko, 2003), swallowing (Nicosia et al., 2000; Steele & Van Lieshout, 2004), and speech disorders (Katz & Verma, 1994; Goozée, et al., 2000; Bose et al., 2003).

A 3D EMA instrument (the Articulograph AG500) is commercially available from Carstens, Inc., Munich, Germany (Zierdt et al., 2000; Hoole et al., 2003). In this 3D system, a clear acrylic cube surrounds the speaker's head (see Figure 1.13). The cube contains six transmitters, in a spherical configuration, each of which produces a magnetic field at a different frequency. Within the oral cavity sensors are affixed to the articulators of interest. Each sensor produces an alternating current that varies with its relative distance from the six transmitters. Sensor location is represented by five parameters: three positions (x, y, z), and two angles (azimuth and elevation). The angular parameters mean that tilt of the sensor no longer introduces artifact; instead tilt is incorporated into the two measurement angles. The position of each sensor is subtracted from a sensor affixed to the head. The 3D machine has a spatial resolution of 1 mm and an angular accuracy of one degree.

Two limitations of point-tracking systems are that sensors need to be affixed to the structures, potentially interfering with speech or swallowing. In addition, only points are measured, so the behavior of the entire articulator is largely inferred. This is most problematic for the soft tissue structures like the lips, tongue, and velum, whose movements are more fluid than rigid.

The biggest advantages of point-tracking systems are the rapid tracking rate, and the ability to track multiple articulators simultaneously. Because of these two features, interaction among the articulators can be measured, and questions about inter-articulator timing and programming can be answered. In addition, the 3D articulograph is one of the few instruments that does not impede head motion, allowing natural movement. Although there has been some concern over possible health consequences from exposure to magnetic fields, the articulometer poses minimal biological hazards as it uses short exposure times and low field strengths (cf. Hasegawa-Johnson, 1998). Nonetheless, its use is not recommended with pregnant women.

2.2 *X-ray Microbeam*

The X-ray Microbeam tracks tissue-points on the surface of the articulators, resulting in data similar to 2D EMA. The method is quite different, however, as the X-ray Microbeam uses a very thin X-ray beam to track the motion of small gold pellets. The pellets are affixed to one or more articulators using dental adhesive. The beam is 0.4 mm thick and the pellets are 2–3 mm in diameter. Gold pellets are used because gold is an inert metal and because, as the X-ray dosage is very small, only a very dense metal can be detected. The system was designed to reduce radiation dosage to well below that of a dental X-ray, to avoid radiosensitive areas, such as the eyes, and to reduce secondary photon scatter. The X-ray beam focuses primarily on the pellets so that the surrounding tissue receives only minimal radiation.

The direction of the beam is computer controlled. The X-ray beam originates at one side of the subject, passes through the subject's head, and is detected by a scintillation counter on the far side. Up to 1,000 pellet positions are sampled per second. Thus, if 10 pellets are used, they can each be sampled 100 times per second. Differential sampling rates are also possible. The pellets are sampled initially in rest position to determine baseline x–y displacement. A computer algorithm causes the beam to scan the area in which the pellet is predicted to move. The prediction is based on the pellet's previous displacement, velocity, and acceleration. Each pellet is scanned in order, and position accuracy of the beam is 62 microns.

Many studies have collected X-ray Microbeam data, or used data available in an archival database from the University of Wisconsin, Madison (Westbury, 1994). X-ray Microbeam data have been used to study swallowing (cf. Tasko et al., 2002), articulator interaction (Westbury et al., 2002), motor control (Tasko & Westbury, 2002, 2004), and speech disorders (Weismer et al., 2003).

The strengths of the X-ray Microbeam are rapid sampling rate and accuracy of tracking, which make this an excellent system for examining timing related coarticulatory effects, kinematic parameters such as velocity and acceleration, and the intercoordination of the articulators. In addition, the technique is unobtrusive and the low radiation dosage allows for reasonably large data sets to be collected on each subject.

In addition to the limitations found in EMA, the X-ray Microbeam is further limited by being unique, and therefore difficult to access for most investigators and their subjects. An additional disadvantage is that although the radiation dosage is low, an X-ray based system does contains some biological hazard. In addition, only two-dimensional data are collected, and movements off-plane are lost or induce error. Finally, a problem common to all pellet-tracking devices is the near impossibility of affixing pellets to the pharynx, velum, and pharyngeal tongue due to the gag reflex and poor accessibility, thus limiting its use with many subjects. The X-ray Microbeam archival database is available at www.medsch.wisc.edu/~milenkvc/pdf/ubdbman.pdf.

2.3 *Optotrak*

It is easier to track points on the face than those within the oral cavity, because the markers maintain visual contact with the sensors at all times. Point-tracking systems, such as Optotrak (Northern Digital, Waterloo, ON, Canada) and Vicon (Vicon Motion Systems Inc., Lake Forest, CA, USA) use optical measurements of LED markers in three-dimensional (*x–y–z*) space. Thus, they measure left-to-right movement as well as anterior-to-posterior and superior-to-inferior. The system consists of (1) markers placed on surface structures, (2) sensors that track their position, and (3) a unit that controls the timing of marker emissions and sensor processing. The markers are tracked at high sampling frequencies; for example, Optotrack samples at 100 kHz divided by the number of markers. Embedded in the center of each Optotrak marker is a small semiconductor chip that emits an infrared signal. A strip of three camera-like sensors tracks the movement of the infrared emissions of the markers. Spatial resolution is also excellent, positional accuracy is 0.1 mm on the *x*- and *y*-axes and 1.5 mm on the *z*-axis at 2.25 meters. The sensors are pre-aligned on a single unit so that they measure the position of each marker in 3D space.

These external tracking instruments track lip, jaw, and face motion in 3D and are well-suited to examining the complex relationships between them, such as the cues provided by facial motions in hyperarticulated versus normal speech (Maeda & Toda, 2003). The instruments are noninvasive and are convenient to use. Their disadvantage is that they are limited to external use, unlike the X-ray Microbeam and the articulometer, because the sensors must maintain "visual" contact with the markers. Therefore, Optotrak and Vicon cannot be used inside the mouth and so do not reveal structures within the vocal tract. Two more limitations are common to all point-tracking systems. First, only tissue-points are tracked, not the entire structure. For rigid structures such as the jaw, this is not a problem and the entire structure can be reconstructed. However, the flexible deformation of soft tissue structures, such as the tongue, lips, and velum, is incompletely measured and represented. It is also possible that the markers or their attached wires may interfere, at least minimally, with truly natural speech.

3 Measurement of Tongue–Palate Interaction

3.1 Tongue–palate contact

Electropalatography (EPG) directly measures tongue–palate contact in real time during the motions of speech and swallowing. EPG data are quite different from both point-tracking and imaged data. Small metal electrodes (about 0.5–1.5 mm diameter) are embedded into an acrylic pseudopalate and are activated when contacted by the tongue; this contact completes an electric circuit in the body. The electrodes measure on/off contact, and though the activation threshold can be changed, subtle changes in pressure are not recorded. Thin wires (approx. 42 gauge) attached to each electrode, exit the palate by winding behind the back molars bilaterally, and running forward along the outer surface of the teeth and out through the corners of the lips. A ground electrode completes the circuit. The pseudopalate is electrically isolated from the computer that drives it, and the wires are driven by an AC current of less than five microamps. When the tongue contacts an electrode, the circuit is completed and the contact registered. EPG is high-dimensional as it dynamically records multiple contact points between two structures (tongue and palate) from the entire palate surface. A recent system, WinEPG (Articulate Instruments, Edinburgh), uses 62 electrodes embedded into a very thin (0.5 mm) acrylic pseudo-palate, which is molded to fit the palate of the speaker (Figure 1.14). The electrodes are sampled at 100–200 Hz (Wrench, 2007).

Data from EPG provide a unique perspective on the interaction of the tongue and palate, and their relationship to vocal tract shape in the palatal area (Scobbie et al., 2004). EPG was used initially, over three decades ago, to study linguistics (Hardcastle, 1972) and speech disorders (Fletcher et al., 1975). Since then it has been used as a clinical tool for on-site and remote speech therapy (Gibbon et al., 1998), to study cleft palate (Gibbon et al., 2001), cerebral palsy (Gibbon & Wood, 2003), Parkinson's Disease (McAuliffe et al., 2006), open bite (Suzuki et al., 1981), Traumatic Brain Injury (Hartelius et al., 2005), glossectomy (Suzuki, 1989), and apraxic speech (Hardcastle & Edwards, 1992).

EPG has also been used to answer linguistic questions, often with EMA, as the two methods provide complementary information about tongue motion and tongue–palate contact. One application that is well established across languages is the documentation of tongue–palate contact patterns for alveolar sounds. These have been studied extensively in such languages as Japanese (Miyawaki et al., 1974), Norwegian (Moen et al., 2004), Italian (Farnetani & Recasens, 1993), Catalan (Recasens & Pallarès, 2001), German (Kohler, 1976), and English (Ellis & Hardcastle, 2002). Velar sounds are less studied since the traditional pseudo-palate usually ends before the soft palate (Hardcastle, 1985). However, some velar patterns have been studied (Suzuki & Michi, 1986) as has the behavior of the tongue during labials (Gibbon et al., 2007). Coarticulation is observed well in EPG because of its high spatial resolution and fast sampling rates. Studies have examined locus equations (Tabain, 1998), assimilation in consonant sequences (Ellis & Hardcastle, 2002), and the interference of prosody in vowel-to-vowel coarticulation (Fletcher, 2004).

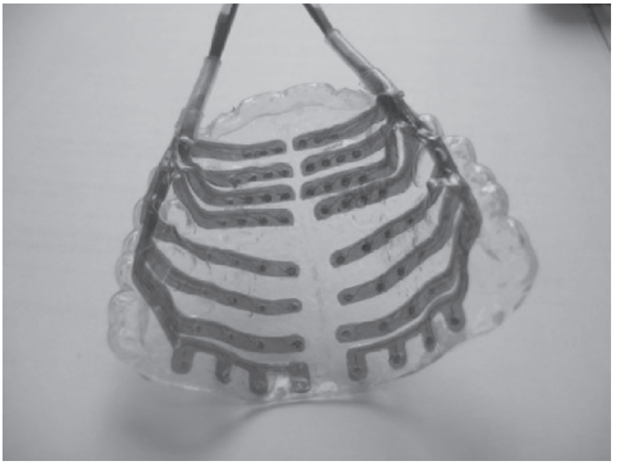

Figure 1.14 A pseudopalate containing 64 electrodes that record tongue–palate contact. (Photo courtesy of Alan Wrench)

The drawback of EPG is that data are collected only when the tongue touches the palate. Thus information is lost when the jaw lowers the tongue away from the domain of the palate, as occurs during mid and low vowels. However, during continuous speech the tongue is in fairly continuous contact with the palate at one location or another. EPG thus contributes greatly to the study of speech, especially the study of lingual consonants, constriction shapes and sizes, increasing its value in studies of language and disorders.

3.2 Tongue–palate pressure

Pressure palatography (PPG) measures moment-to-moment changes in tongue–palate pressure. The transducers embedded in the acrylic palate measure pressure instead of contact. Several methods have been used to transduce this pressure. The earliest and most commonly used transducers were strain gauges. A strain gauge is a transducer that is mounted on an object, such as a beam or diaphragm, that is capable of deformation. As the object deforms the gauge also deforms. As the gauge deforms, or strains, the strain produces a change in resistance, which is converted into a change in voltage. The change in voltage is an analog waveform that is amplified by a wheatstone bridge and recorded in analog or digital form.

Miniature strain gauges were first used to measure tongue–palate pressure over 30 years ago. McGlone and Proffit (1972) inserted two strain gauges into an artificial palate. Changes in pressure due to tongue–palate contact were measured. Despite their success in delineating pressure differences in /t/ versus /d/, tongue–palate pressure studies were not published again for 20 years. In recent years, strain gauge pressure sensors have been successfully used for the assessment of the tongue in oral functions, such as deglutition (Chiba et al., 2003; Ono et al., 2004) and articulation (Tsuga et al., 2003). PPG allows the measurement of tongue force changes over time, including time between the onset of linguopalatal contact to the time of maximum pressure, the maximum pressure and the position–force ratio. Pressure transducers have been used to determine hemispheric dominance (Shinagawa et al., 2003), and to record tongue behavior in patients who have had glossectomy (Yoshioka et al., 2004) and rapid palatal expansion (Kucukkeles & Ceylanoglu, 2003).

In addition to strain gauges other methods of measuring pressure changes have been used experimentally. Wakumoto and colleagues (1998) devised a pressure recording "sheet" composed of two polyester layers separated by regions of air and pressure-sensing cells. A pressure-sensing cell has a layer of "pressure-sensitive ink" embedded between two electrodes that measure its resistance to pressure. A cell is 3 mm diameter with a sensitivity of 173–2734 Pa (Wakumoto et al., 1998). The sheets of electrodes are 0.1 mm thick and are glued to an acrylic pseudopalate of 0.5 mm. This work has used up to 16 sensors per palate and has sensitivity enough to distinguish the lingual pressures of /t/ and /d/. Murdoch et al. (2004) developed pressure sensors that contain a magnet and a Hall effect transducer (HET). HETs produce a voltage when in the presence of a magnetic field that is proportional to the magnetic field. When the tongue touches one of these sensors a cantilever beam is deflected, which changes the distance between the magnet and the HET. This varying distance results in proportionally varied voltage output. One HET sensor is 4.4 mm x 6.2 mm, and five were embedded in the prototype pseudopalate. A single subject trial was inconclusive because, as with all the PPG transducers, the limited number of sensors reduces sensitivity to subtle tongue-pressure changes. Continued work is expected to improve these techniques.

Instrumental studies of physiology are challenging and, no single instrument provides total vocal tract information. However, the importance of these instruments lies in the critical role they play in cutting-edge studies of speech motor control, speech disorders, phonetics, phonology, and even speech processing. The data acquired by these instruments is in great demand for two reasons. First, the data keep increasing our knowledge of speech physiology, speech disorders, and coarticulation strategies. They reveal how articulation and rhythm are organized and controlled, what aspects of speech are common among languages, and the nature of differences between speakers, languages, and disorders. Second, the data are critical in testing current theories and models. The facts and models of an era are supported or disproved by the data from new instrumentation. Forty years ago, the sound spectrograph destroyed the notion that speech sounds were independent, concatenated segments. Today, measures of speech physiology are challenging our ideas of how vocal tract constrictions are achieved and what features of them are acoustically salient. Similarly, we wish to find out what components of the articulatory gesture are carried in the speech wave and decoded by the listener. A better understanding of how the vocal tract produces speech can also improve synthetic speech and provide strategies for machine recognition of speech. These and other issues, which are in the forefront of speech research, can be addressed and perhaps resolved using the instruments described in this chapter. The exciting leaps in knowledge provided by physiological data far outweigh the difficulties associated with the use of these instruments.

REFERENCES

Ackermann, H. & Riecker, A. (2004) The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*, 89, 320–8.

Axel, L. & Dougherty, L. (1989) Heart wall motion: Improved method of spatial modulation of magnetization for MR imaging. *Radiology*, 172, 349–50.

Bonilha, L., Moser, D., Rorden, C., Baylis, G. C., & Fridriksson, J. (2006) Speech apraxia without oral apraxia: Can normal brain function explain the physiopathology? *Neuroreport*, 17, 1027–31.

Bose, A., Lieshout, P. H. H. M. van, & Square, P. A. (2003) Speech coordination in individuals with aphasia and normal speakers. *Brain and Language*, 87, 158–9.

Bressmann, T., Heng, C.-L., & Irish, J. C. (2005a) Applications of 2D and 3D ultrasound imaging in speech-language pathology. *Journal of Speech-Language Pathology and Audiology*, 29, 158–68.

Bressmann, T., Uyn, C., & Irish, J. C. (2005b) Analyzing normal and partial glossectomee tongues using ultrasound. *Clinical Linguistics and Phonetics*, 19, 35–52.

Byrd, D., Lee, S., Riggs, D., & Adams, J. (2005) Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America*, 118, 3860–73.

Chiang, Y. C., Lee, F. P., Peng, C. L., & Lin, C. T. (2003) Measurement of tongue movement during vowels production with computer-assisted B-mode and M-mode ultrasonography. *Otolaryngology-Head and Neck Surgery*, 128, 805–14.

Chiba, Y., Motoyoshi, M., & Namura, S. (2003) Tongue pressure on loop of transpalatal arch during deglutition. *American Journal of Orthodontics and Dentofacial Orthopedics*, 123, 29–34.

Chi-Fishman, G., Stone, M., & McCall, Gerald N. (1998) Lingual action in normal sequential swallowing. *Journal of Speech Language and Hearing Research*, 41, 771–85.

Davidson, L. (2006) Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120, 407–15.

Djamshidpey, H., Waneland, I., Krull, D., & Lindblom, B. (1998) X-ray analyses of speech: Methodological aspects. *Proceedings of Fonetik '98, 11th Swedish Phonetics Conference*, Stockholm, pp. 168–71.

Ellis, L. & Hardcastle, W. J. (2002) Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics*, 30, 373–96.

Engwall, O. (2003) Combining MRI, EMA & EPG measurements in a three-dimensional tongue model. *Speech Communication*, 41, 303–29.

Epstein, M. A. & Stone, M. (2005) The tongue stops here: Ultrasound imaging of the palate. *Journal of the Acoustical Society of America*, 118, 2128–31.

Fant, G. (1965) Formants and cavities. *Proceedings of the 5th International Congress of Phonetic Sciences* (pp. 120–41). Munster, 1964. Basel/New York: S. Karger.

Farnetani, E. & Recasens, D. (1993) Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech*, 36, 279–302.

Fletcher, J. (2004) An EMA/EPG study of vowel-to-vowel articulation across velars in Southern British English. *Clinical Linguistics and Phonetics*, 18, 577–92.

Fletcher, S. G., McCutcheon, M. J., & Wolf, M. B. (1975) Dynamic palatometry. *Journal of Speech and Hearing Research*, 18, 812–19.

Fuchs, S., Perrier, P., Geng, C., & Mooshammer, C. (2006) What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. In J. Harrington & M. Tabain (eds.), *Speech production: Models, Phonetic Processes and Techniques*. New York: Psychology Press, pp. 149–64.

Gibbon, F., Crampin, L., Hardcastle, W. J. et al. (1998) CleftNet Scotland: A network for the treatment of cleft palate speech using EPG. *International Journal of Language and Communication Disorders*, 33, supplement, 44–9.

Gibbon, F., Hardcastle, W. J., Crampin, L., Reynolds, B., Razzell, R., & Wilson, J. (2001) Visual feedback therapy using electropalatography (EPG) for articulation disorders associated with cleft palate. *Asia Pacific Journal of Speech, Language and Hearing*, 6, 53–8.

Gibbon, F., Lee, A., & Yuen, I. (2007) Tongue palate contact during bilabials in normal speech. *Cleft Palate-Craniofacial Journal*, 44, 87–91.

Gibbon, F. & Wood, S. (2003) Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: A case study. *Journal of Clinical Linguistics and Phonetics*, 17, 365–74.

Gick, B. (2002a) The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32, 113–22.

Gick, B. (2002b) An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*, 59, 38–48.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004) Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, 61, 220–33.

Gilbert, R. J. & Napadow, V. J. (2005) Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion tensor magnetic resonance imaging. *Dysphagia*, 20, 1–7.

Goozée, J., Murdoch, B. E., Theodoros, D. G., & Stokes, P. D. (2000) Kinematic analysis of tongue movements in dysarthria following traumatic brain injury using electromagnetic articulography. *Brain Injury*, 14, 153–74.

Gracco, V. L., Tremblay, P., & Pike, B. (2005) Imaging speech production using fMRI. *Neuroimage*, 26, 294–301.

Green, J. R. & Wang, Y.-T. (2003) Tongue surface movement patterns during speech and swallowing. *Journal of the Acoustical Society of America*, 113, 2820–33.

Hardcastle, W. J. (1972) The use of electropalatography in phonetic research. *Phonetica*, 25, 197–215.

Hardcastle, W. J. (1985) Some phonetic and syntactic constraints on lingual co-articulation during /kl/ sequences. *Speech Communication*, 4, 247–63.

Hardcastle, W. J. & Edwards, S. (1992) EPG-based descriptions of apraxic speech errors. In R. Kent (ed.), *Intelligibility in Speech Disorders: Theory, Measurement and Management*. Philadelphia: John Benjamins, 287–328.

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

Hartelius, L., Theodoros, D., & Murdoch, B. (2005) Use of electropalatography in the treatment of disordered articulation following traumatic brain injury: A case study. *Journal of Medical Speech–Language Pathology*, 13, 189–204.

Hasegawa-Johnson, M. (1998) Electromagnetic exposure safety of the Carstens Articulograph AG100. *Journal of the Acoustical Society of America*, 104, 2529–32.

Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., & Haker, K. (2003) Vowel category dependence of the relationship between palate height, tongue height, and oral area. *Journal of Speech, Language, and Hearing Research*, 46, 738–53.

Hashimoto, K. & Sasaki, K. (1982) On the relationship between the shape and position of the tongue for vowels. *Journal of Phonetics*, 10, 291–9.

Hoole, P., Zierdt, A., & Geng, C. (2003) Beyond 2D in articulatory data acquisition and analysis. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 265–8.
Iskarous, K. (2005) Patterns of tongue movement. *Phonetica*, 33, 363–82.
Kaburagi, T. & Honda, M. (1996) A model of articulator trajectory formation based on the motor tasks of vocal-tract shapes. *Journal of the Acoustical Society of America*, 99, 3154–70.
Katz, W. F. & Verma, S. (1994) Kinematic evidence for compensatory articulation by normal and nonfluent aphasic speakers. *Brain and Language*, 47, 357–60.
Kent, R. D. (1972) Some considerations in the cinefluorographic analysis of tongue movements during speech. *Phonetica*, 26, 16–32.
Kent, R. D. & Moll, K. L. (1972) Cinefluorographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research*, 15, 453–73.
Kent, R. D. & Netsell, R. (1971) Effects of stress contrasts on certain articulatory parameters. *Phonetica*, 24, 23–44.
Kohler, K. (1976) The instability of word-final alveolar plosives in German: An Electropalatographic investigation. *Phonetica*, 33, 1–30.
Kokawa, T., Saigusa, H., Aino, I., et al. (2006) Physiological studies of retrusive movements of the human tongue. *Journal of Voice*, 20, 414–22. Epub November 21, 2005.
Kucukkeles, N. & Ceylanoglu, C. (2003) Changes in lip, cheek, and tongue pressures after rapid maxillary expansion using a diaphragm pressure transducer. *Angle Orthodontist*, 73, 662–8.
Kühnert, B. & Hoole, P. (2004) Speaker-specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics and Phonetics*, 18, 559–75.
Lam, P., Au-Yeung, K. M., Cheng, P. W., et al. (2004) Correlating MRI and histologic tumor thickness in the assessment of oral tongue cancer. *American Journal of Roentgenology*, 182, 803–8.
Lell, M. M., Greess, H., Hothorn, T., Janka, R., Bautz, W. A., & Baum, U. (2004) Mutiplanar functional imaging of the larynx and hypopharynx with multislice spiral CT. *European Journal of Radiology*, 14, 2198–205.
Lenz, M., Greess, H., Baum, U., Dobritz, M., & Kersting-Sommerhoff, B. (2000) Oropharynx, oral cavity, floor of the mouth: CT and MRI. *European Journal of Radiology*, 33, 203–15.
Lieshout, P. H. H. M. van, Alfonso, P. J., Hulstijn, W., & Peters, H. F. M. (1994) Electromagnetic midsagittal articulography (EMMA). In F. J. Maarse, A. E. Akkerman, A. N. Brand, L. J. M. Mulder, & M. J. Van der Stelt (eds.), *Applications, Methods and Instrumentation* (pp. 62–76). Lisse, The Netherlands: Swets & Zeitlinger.
Lieshout, P. H. H. M. van & Moussa, W. (2000). The assessment of speech motor behaviors using electromagnetic articulography. *The Phonetician*, 81, 9–22.
Lindblom, B., Sussman, H., Modarresi, G., & Burlingame, E. (2002) The trough effect: Implications for Speech Motor Programming. *Phonetica*, 59, 245–62. *Using data collected in Stockholm by* Branderud, P, Lundburg, H, Lander, J.
Mády, K., Sader, R., Zimmermann, A., et al. (2002) Assessment of consonant articulation in glossectomee speech by dynamic MRI. *Proceedings of the 7th International Conference on Spoken Language Processing, ICSLP '2002*, Denver, USA. 961–4.
Maeda, S. (1990) Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W. L. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 131–49). Dordrecht: Kluwer.
Maeda, S. & Toda, M. (2003) Mechanical properties of lip movements: How to characterize different speaking styles? *The 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 189–92.

Magen, H. S., Kang, A. M., Tiede, M. K., & Whalen, D. H. (2003) Posterior pharyngeal wall position in the production of speech. *Journal of Speech, Language, and Hearing Research*, 46, 241–51.

McAuliffe, M. J., Ward, E. C., & Murdoch, B. E. (2006) Speech production in Parkinson's disease, I: An electropalatographic investigation of tongue–palate contact patterns. *Clinical Linguistics and Phonetics*, 20, 1–18.

McClean, M. D. & Tasko, S. M. (2003) Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech, Language, and Hearing Research*, 46, 1387–400.

McGlone, R. & Proffit, W. (1972) Correlation between functional lingual pressure and oral cavity size. *Cleft Palate Journal*, 9, 229–35.

McGowan, R. S. (2006) Perception of synthetic vowel exemplars of 4-year-old children and estimation of their corresponding vocal tract shapes. *Journal of the Acoustical Society of America*, 120, 2850–8.

Mermelstein, P. (1973) Articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, 53, 1070–82.

Miller, J. L. & Watkin, K. L. (1997) Lateral pharyngeal wall motion during swallowing using real time ultrasound. *Dysphagia*, 12, 125–32.

Miyawaki, K., Kiritani, S., Tatsumi, I. F., & Fujimura, O. (1974) Palatographic observation of VCV articulations in Japanese. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo*, 8, 51–7.

Moen, I., Simonsen, H. G., & Lindstad, A. M. (2004) An electronic database of Norwegian speech sounds: Clinical aspects. *Journal of Multilingual Communication Disorders*, 2, 43–9.

Munhall, K. G. & Ostry, D. J. (1985) Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control* (pp. 145–62), Denver: Denver Center for the Performing Arts.

Munhall, K., Vatikiotis-Bateson, E., & Tohkura, Y. (1994a) X-ray film database for speech research. *Journal of the Acoustical Society of America*, 98, 1222–4.

Munhall, K., Vatikiotis-Bateson, E., & Tohkura, Y. (1994b). X-ray film database for speech research. Technical report TR-H-116, ATR Human Information Processing Laboratories, Kyoto.

Murdoch, B. E., Goozée, J. V., Veidt, M., Scott, D. H., & Meyers, I. A. (2004) Introducing the pressure-sensing palatograp: The next frontier in electropalatography. *Clinical Linguistics and Phonetics*, 18, 433–45.

Napadow, V. J., Chen, Q., Wedeen, V. J., & Gilbert, R. J. (1999a) Biomechanical basis for lingual muscular deformation during swallowing. *American Journal of Physiology*, 277, G695–G701.

Napadow, V. J., Chen, Q., Wedeen, V. J., & Gilbert, R. J. (1999b) Intramural mechanics of the human tongue in association with physiological deformations. *Journal of Biomechanics*, 322, 1–12.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004) An approach to real-time magnetic resonance imaging for speech prodcution. *Journal of the Acoustical Society of America*, 115, 1771–6.

NessAiver, M., Stone, M., Parthasarathy, V., Kahana, Y., Kots, A., & Paritsky, A. (2006) Recording high quality speech during tagged Cine MRI studies using a fiber optic microphone. *Journal of Magnetic Resonance Imaging*, 23, 92–7.

Nicosia, M. A., Hind, J. A., Roecker, E. B., et al. (2000) Age effects on the temporal evolution of isometric and swallowing pressure. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 55, M634–M640.

Niitsu, M., Kumada, M., Campeau, N. G., Niimi, S., Riederer, S. J., & Itai, Y. (1994) Tongue displacement: Visualization with rapid tagged magnetization-prepared MR imaging, *Radiology*, 191, 578–80.

Ono, T., Hori, K., & Nokubi, T. (2004) Pattern of tongue pressure on hard palate during swallowing. *Dysphagia*, 19, 259–64.

Parthasarathy, V., Prince, J. L., Stone, M., Murano, E., & NessAiver, M. (2007) Measuring tongue motion from tagged Cine-MRI using harmonic phase (HARP) processing. *Journal of the Acoustical Society of America*, 121, 1, 491–504.

Parush, A. & Ostry, D. J. (1993) Lower pharyngeal wall coarticulation in VCV syllables. *Journal of the Acoustical Society of America*, 94, 715–22.

Peng, C. L., Jost-Brinkmann, P. G., Miethke, R. R., & Lin, C. T. (2000) Ultrasonographic measurement of tongue movement during swallowing. *Journal of Ultrasound in Medicine*, 19, 15–20.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992) Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, *Journal of the Acoustical Society of America*, 92, 3078–96.

Perkell, J. & Zandipour, M. (2002) Economy of effort in different speaking conditions, II: Kinematic performance spaces for cyclical and speech movements. *Journal of the Acoustical Society of America*, 112, 1642–51.

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006) Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 7865–70.

Recasens, D. & Pallarès, M. D. (2001) Coarticulation, blending and assimilation in Catalan consonant clusters. *Journal of Phonetics*, 29, 273–301.

Schmelzeisen, R., Ptok, M., Schonweiler, R., Hacki, T., & Neukam, F. W. (1996) Re-establishment of speech and swallowing function following extensive tumour resections in the head and neck. *Laryngo-Rhino-Otologie*, 75, 231–8.

Scobbie, J. M., Wood, S. E., & Wrench, A. A. (2004) Advances in EPG for treatment and research: An illustrative case study. *Clinical Linguistics and Phonetics*, 18, 373–89.

Serrurier, A. & Badin, P. (2005) Towards a 3D articulatory model of velum based on MRI and CT images. *ZAS Papers in Linguistics*, 40, 195–211 (Zentrum für Allgemeine Sprachwissenschaft, Sprachwissenschaft, Typologie und Universalienforschung).

Shinagawa, H., Murano, E. Z., Zhuo, J., et al. (2008) Human tongue muscle fiber tracking during rest and tongue protrusion with oral appliance: A preliminary study with diffusion tensor imaging. *Acoustic Science and Technology*, 29, 291–4.

Shinagawa, H., Murano, E. Z., Zhuo, J., et al. (2009) Effect of oral appliances on genioglossus muscle tonicity seen with diffusion tensor imaging: A pilot study. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, 107, 57–63.

Shinagawa, H., Ono, T., Ishiwata, Y., et al. (2003) Hemispheric dominance of tongue control depends on the chewing-side preference. *Journal of Dental Research*, 82, 278–83.

Slud, E., Stone, M., Smith, P. J., & Goldstein, M. (2002) Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*, 59, 10.

Soder, N. & Miller, N. (2002) Using ultrasound to investigate intrapersonal variability in durational aspects of tongue movement during swallowing. *Dysphagia*, 17, 288–97.

Specht, K., Rimol, L. M., Reul, J., & Hugdahl, K. (2005) “Soundmorphing”: A new approach to studying speech perception in humans. *Neuroscience Letters*, 384, 60–5.

Steele, C. & Lieshout, P. H. H. M. van (2004) Use of electromagnetic midsagittal articulography in the study of swallowing. *Journal of Speech, Language, and Hearing Research*, 47, 342–52.

Stone, M. (2005) A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19, 455–502.

Stone, M., Davis, E., Douglas, A., et al. (2001a) Modeling the motion of the internal tongue from tagged cine-MRI images. *Journal of the Acoustical Society of America*, 109, 2974–82.

Stone, M., Davis, E., Douglas, A., et al. (2001b) Modeling tongue surface contours from cine-MRI images. *Journal of Speech, Language, and Hearing Research*, 44, 1026–40.

Stone, M. & Lundberg, A. (1996) Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America*, 99, 3728–37.

Stone, M., Shawker, T., Talbot, T., & Rich, A. (1988) Cross-sectional tongue shape during the production of vowels. *Journal of the Acoustical Society of America*, 83, 1586–96.

Story, B. (2005) Synergistic modes of vocal tract articulation for American English vowels, *Journal of the Acoustical Society of America*, 118, 3834–59.

Story, B., Titze, I., & Hoffman, E. (1996) Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, 100, 537–54.

Subtelny, J., Li, W., Whitehead, R., & Subtelny, J. D. (1989) Cephalometric and cineradiographic study of deviant resonance in hearing impaired speakers. *Journal of Speech and Hearing Disorders*, 54, 249–65.

Suzuki, N. (1989) Clinical applications of EPG to Japanese cleft palate and glossectomy patients. *Clinical Linguistics and Phonetics*, 3, 127–36.

Suzuki, N. & Michi, K. (1986) Dynamic velography. *Proceedings of the 20th Congress of the International Association of Logopedics and Phoniatrics*, Tokyo, 172–3.

Suzuki, N., Sakuma, T., Michi, K. I., & Ueno, T. (1981) The articulatory characteristics of the tongue in anterior openbite: Observation by use of dynamic palatography. *International Journal of Oral Surgery*, 10, 299–303.

Tabain, M. (1998) Coarticulation in CV syllables: A locus equation and EPG perspective. *Journal of the Acoustical Society of America*,103(5), 2980.

Tameem, H. & Mehta, B. (2004) Solid modeling of human vocal tract using magnetic resonance imaging and acoustic pharyngometer. *Proceedings of the 26th International Conference of the IEEE: Engineering in Medicine and Biology*, San Francisco, 2, 5115–8.

Tasko, S. M., Kent, R. D., & Westbury, J. R. (2002) Variability in tongue movement kinematics during normal liquid swallowing. *Dysphagia*, 17, 126–38.

Tasko, S. & Westbury, J. (2002) Defining and measuring speech movement events. *Journal of Speech, Language and Hearing Research*, 45, 127–42.

Tasko, S. & Westbury, J. (2004) Speed-curvature relations for speech related articulatory movement, *Journal of Phonetics*, 32, 65–80.

Tom, K., Titze, I. R., Hoffman, E. A., & Story, B. H. (2001) 3-D vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness, *Journal of the Acoustical Society of America*, 109, 742–7.

Tsuga, K., Hayashi, R., Sato, Y., & Akagawa, Y. (2003) Handy measurement for tongue motion and coordination with laryngeal elevation at swallowing. *Journal of Oral Rehabilitation*, 30, 985–9.

Tye-Murray, N. (1991) The establishment of open articulatory postures by deaf and hearing talkers. *Journal of Speech and Hearing Research*, 34, 453–9.

Ueda, D., Yano, K., & Okuno, A. (1993) Ultrasound imaging of the tongue, mouth, and vocal cords in normal children: Establishment of basic scanning positions. *Journal of Clinical Ultrasound*, 21, 431–9.

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006) Locating the initial stages of speech–sound processing in human temporal cortex. *Neuroimage*, 31, 1284–96. Epub February 28, 2006.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005) Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America*, 117, 338–50.

Wakumoto, M., Masaki, S., Honda, K., & Ohue, T. (1998) A pressure sensitive palatography: Application of new pressure sensitive sheet for measuring tongue–palatal contact pressure. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, 7, 3151–4. (Available online at: http://andosl.anu.edu.au.icslp98/main/html)

Watkin, K. L. (1999) Ultrasound and swallowing. *Folia Phoniatrica et Logopaedia*, 51, 183–98.

Watkin, K. L. & Rubin, J. M. (1989) Pseudo-three-dimensional reconstruction of ultrasonic images of the tongue. *Journal of the Acoustical Society of America*, 85, 496–9.

Wedeen, V. J., Reese, T. G., Napadow, V. J., & Gilbert, R. J. (2001) Demonstration of primary and secondary muscle fiber architecture of the bovine tongue by diffusion tensor magnetic resonance imaging. *Biophysics Journal*, 80, 1024–8.

Weismer, G., Yunusova, Y., & Westbury, J. R. (2003) Interarticulator coordination in dysarthria: An X-ray microbeam study. *Journal of Speech, Language, and Hearing Research*, 46, 1247–61.

Westbury, J. (1994) *X-ray Microbeam Speech Production Database User's Handbook*. Waisman Center, Madison: University of Wisconsin at Madison, 8–33.

Westbury, J., Lindstrom, M., & McClean, M. (2002) Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, 45, 651–62.

Wood, S. (1979) A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, 7, 25–43.

Wrench, A. A. (2007) Advances in EPG palate design. *Advances in Speech-Language Pathology*, 9, 3–12.

Xue, S. A. & Hao, G. J. (2003) Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *Journal of Speech, Language, and Hearing Research*, 46, 689–701.

Yang, C. S. & Stone, M. (2002) Dynamic programming method for temporal registration of three-dimensional tongue surface motion from multiple utterances. *Speech Communication*, 38, 201–9.

Yoshioka, F., Ozawa, S., Sumita, Y. I., Mukohyama, H., & Taniguchi, H. (2004) The pattern of tongue pressure against the palate during articulating glossal sounds in normal subjects and glossectomy patients. *Journal of Medical and Dental Sciences*, 51, 19–25.

Zerhouni, E. A., Parish, D. M., Rogers, W. J., Yang, A., & Shapiro, E. P. (1988) Human heart: Tagging with MR imaging – a method for noninvasive assessment of myocardial motion. *Radiology*, 169, 59–63.

Zierdt, A., Hoole, P., Honda, M., Kaburagi, T., & Tillman, H. (2000) Extracting tongues from moving heads. *Proceedings of the 5th Speech Production Seminar: Models and Data*, Munich, 313–16).

Zierdt, A., Hoole, P., & Tillmann, H. G. (1999) Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements. *Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS '99)*, San Francisco, August.