

PALATRON: A TECHNIQUE FOR ALIGNING ULTRASOUND IMAGES OF THE TONGUE AND PALATE*

Jeff Mielke, Adam Baker, Diana Archangeli, and Sumayya Racy
University of Arizona

This paper describes a technique for aligning ultrasound images of the tongue to images of the palate, providing both a fixed point of reference and a means to consider the tongue in the context of passive articulators, thereby addressing two major challenges presented by ultrasound as a means of articulatory imaging, without compromising its portability.

1. Background

A variety of articulatory imaging methods have been employed in the study of language, all with their own advantages and disadvantages. This paper describes a technique for addressing some of the serious challenges presented by ultrasound imaging, namely that ultrasound images offer no fixed point of reference, and that passive articulators such as the palate and velum are invisible under normal circumstances. Ultrasound is advantageous in many ways over other imaging methods, so addressing its challenges is a worthwhile pursuit.

1.1 Why ultrasound

Several different imaging methods are available for linguistic research. Each has its own set of advantages and disadvantages. X-ray movies also show tongue movements in real time, but the equipment is expensive, immobile, and the radiation presents a danger to subjects. A safer means of generating detailed images of the tongue is MRI, but MRI imaging also requires expensive and immobile equipment which is also noisy, requires a subject to lie inside a machine, and suffers from poor temporal resolution. Electromagnetic midsagittal articulometry (EMMA) allows points on the tongue to be tracked safely and in real time, but is expensive, invasive, and is impossible to affix pellets (necessary for imaging) to the back of the tongue. Static palatography and electropalatography show areas of contact between the tongue and palate, but also have difficulty with the back of tongue, and show little or no information about vowels.

Ultrasound imaging employs high-frequency sound waves to generate images of objects, relying on echoes caused by abrupt changes in density. The

* Thanks to James S. McDonnell Foundation grant #220020045 BBMB to Diana Archangeli and to Dr. Bryan Gick, UBC, Bryan Meadows, Jason Packer, Peter Richtsmeier, and Yolanda Vazquez Alvarez.

tongue-air interface is strongly echogenic (because of the large distance in density between air and muscle); therefore an ultrasound transducer placed beneath the chin can produce a real-time movie of the full length of the tongue surface. The ultrasound unit is small and portable, and relatively inexpensive, and the imaging technique is quiet, non-invasive, and non-toxic. These benefits set ultrasound apart from other articulatory imaging methods.

1.2 Challenges of ultrasound

Ultrasound imaging is not without challenges. Ultrasound images are grainy compared to x-ray movies and MRI images. Most significantly, ultrasound images offer no fixed point of reference, and passive articulators such as the palate and velum cannot be imaged at the same time as the tongue. Thus, just like other imaging methods, ultrasound has serious complications. The difference is that unlike danger and expense and the other drawbacks discussed above, ultrasound's complications are not inherent to the technology, and so can be remedied.

The fan-shape of an ultrasound image (Figure 1a) is like the beam of a flashlight (Figure 1b) in that it illuminates the points of the vocal tract at which it is directed. Changing the position or orientation of the ultrasound transducer with respect to the object causes the position and orientation of the object to appear to move, while the fan stays motionless, even though the opposite may be true. The result of the method described in this paper is to determine the position of the tongue (Figure 1c, 1d) within a fixed frame of reference, that of the hard palate. This is important because the tongue and palate cannot be imaged simultaneously, two images are required. Simply overlaying the two fan-shaped areas gives an incorrect view if there has been any movement on the part of the transducer or the head (Figure 1e). What is needed is a method for combining the two images so that they are aligned with respect to the objects rather than the transducer (Figure 1f).

1.3 Prior solutions

Previous solutions to the fixed reference/palate problem in ultrasound imaging of the tongue have involved head and transducer immobilization, e.g. HATS (head and transducer support system, Stone and Davis 1995) and the use of an optoelectric motion measurement system (Optotrak), e.g. HOCUS (Haskins optically-corrected ultrasound system, Whalen *et al.* 2004). The method described in this paper requires comparatively little in the way of head immobilization or additional and costly equipment.

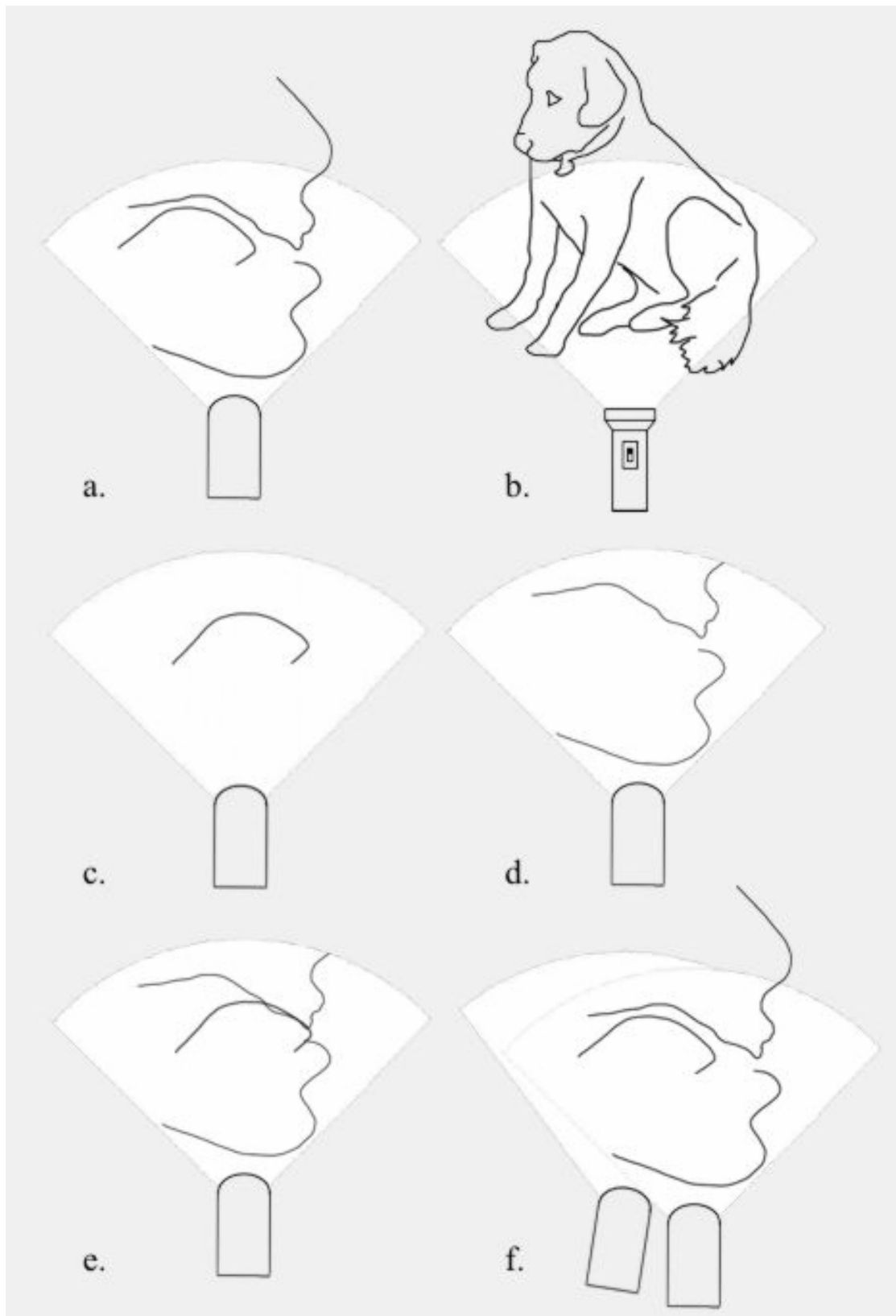


Figure 1. Aligning images

2. Methods

Ultrasound imaging of the palate is possible when the mouth is filled with a medium whose density is similar to that of the tongue (water, gelatin (Vasquez Alvarez p.c.), yogurt, etc.). Whereas a typical ultrasound image shows the sharp density drop where the tongue ends and air begins, filling the mouth with an appropriate medium allows imaging the palate due to the change in density where the medium ends and the roof of the mouth begins.

Simply having the two images is insufficient however, unless there is a means for correctly aligning them. To be able to interpret a speaker's tongue images relative to a palate image for the same speaker, it is necessary either to completely immobilize the head and transducer or to compensate for any movements which do occur. The rest of this section describes a technique for adjusting for head movement, which is tracked with a video camera.

2.1 Equipment

The speaker's movements are recorded on two video channels: one for the ultrasound view inside the speaker's mouth, and another for the video camera's view of the profile of the speaker's face. While the ultrasound view records movements of the tongue relative to the ultrasound transducer, the camera view records externally-visible lip and jaw movements as well as movements of the speaker's head and the ultrasound transducer. Because both views contain the ultrasound transducer, they can be integrated with one another, allowing, among other things, for tongue position to be measured independently of any head or transducer movements.

The ultrasound view is generated by a SonoSite TITAN™ ultrasound unit with a C-11/7-4 11-mm broadband curved array transducer. The video view is generated by a Sony Mini-DV Digital Handycam®. In order to track head and transducer movements, the subject's head and the transducer are both outfitted with tracking devices. The subject wears a pair of sunglasses to which a strip of basswood marked with two pink circles has been attached. A similar strip of basswood with two pink circles is attached to the transducer. The position and orientation of the head and the transducer can be measured according to the position of the four circles, visible in the camera view. The backdrop to the camera view is blue construction paper, and apart from the circles, all equipment visible to the camera is covered with the same blue construction paper (see Figure 2). The two video signals are combined using a Videonics MXProDV digital video mixer. The color-separation overlay feature of the mixer allows the blue

background to be removed from the camera view, leaving only the speaker's face and the four pink circles. The speaker's face appears on the left side of the ultrasound image, while the four circles appear around the periphery (see Figure 3, below). This combined view contains enough information for head and transducer movements to be factored out and for different ultrasound images (e.g., tongues and palates) to be aligned correctly.

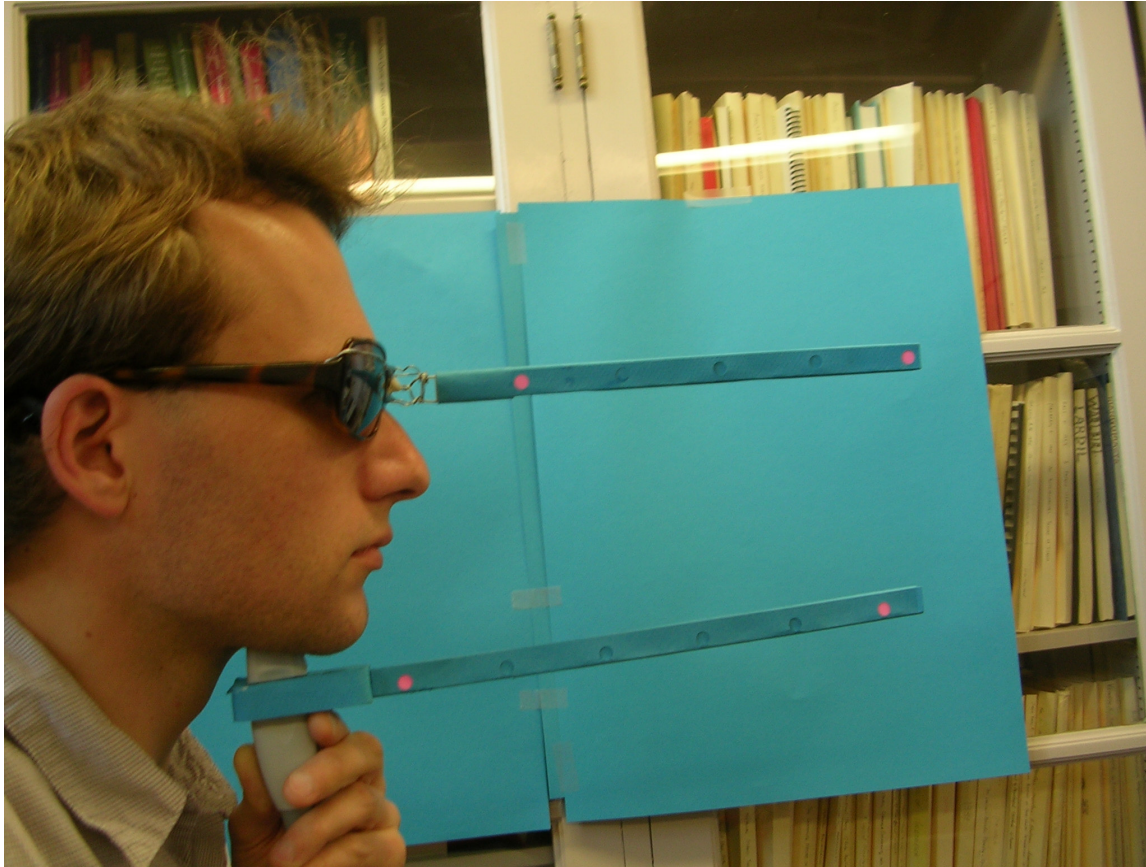


Figure 2. Equipment to track head and transducer position and orientation

2.2 The algorithm

The speaker is recorded while holding a gelatinous, bubble-free substance such as jello or yogurt in her or his mouth.¹ The resulting image of the palate is traced, and then available to be combined with any images of the tongue made during the same session. The position of the two circles attached to the glasses is fixed with respect to the palate. Similarly, the position of the two circles attached to the transducer is fixed with respect to the transducer. Therefore, once the palate has been imaged,

¹ We use an Eden Foods® agar agar solution (1 teaspoon per cup of fruit juice (approx. 11 g/L)) about 1/3 the strength recommended on the package), so that the substance remains gelatinous but also liquid; further it is compatible with both vegetarian and vegan diets.

its location within the ultrasound view can be tracked via the four circles which indicate the position and orientation of the head and transducer. There are three parts to aligning the images, (i) matching the scale of the two images, (ii) aligning the location of the transducer tip in the two images, and (iii) aligning the rotation angle in the two images.

2.2.1 Aligning the scales

The ultrasound and video views are combined in a 720×533 image, with pixel coordinates measured from the top-left corner of the image (Figure 3). The two views are not to scale, and so the first step is to reconcile the two scales. The ultrasound scale is calculated as in (1) on the basis of the vertical scale which appears on the right side of the ultrasound output. In this case the ends of the scale are 80 mm apart.

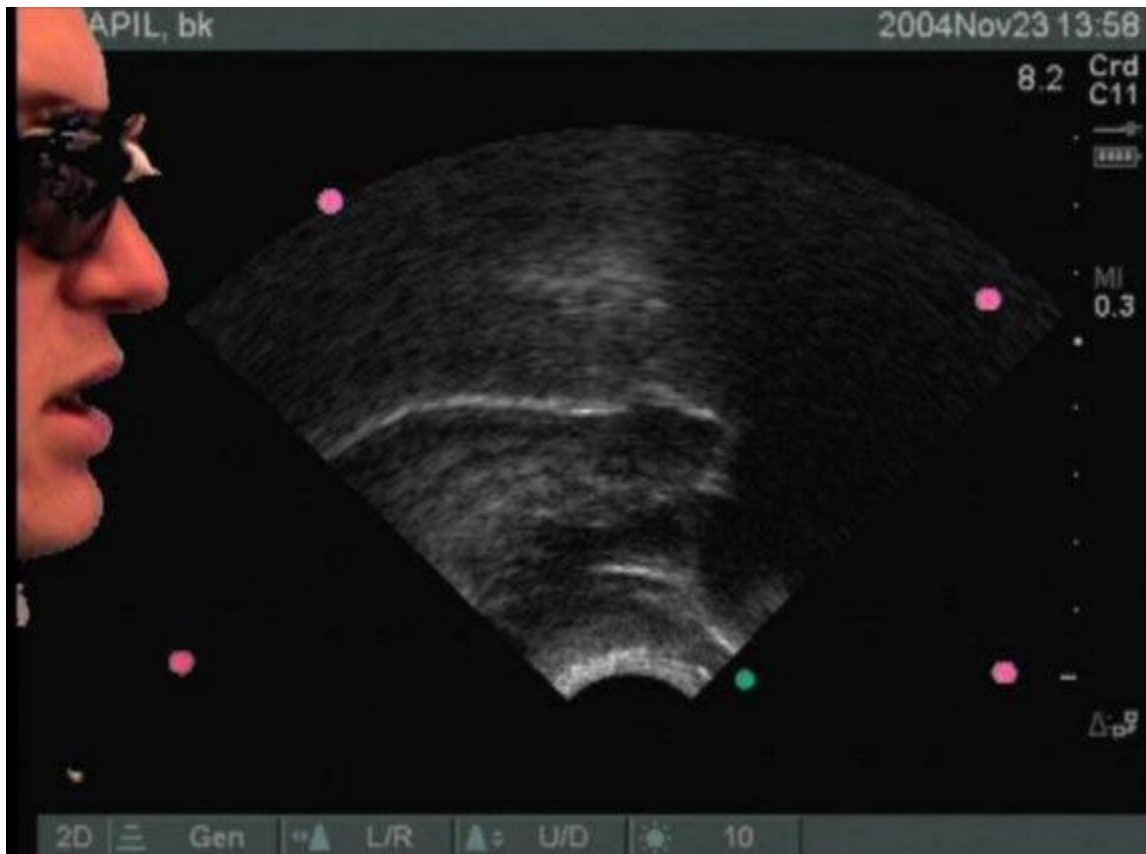


Figure 3. Combined ultrasound and video images

$$(1) \quad scale_{ultrasound} = \frac{80}{bottom_y - top_y} mm / pixel$$

The coordinates of the centers of the four pink circles are measured. The upper-left and upper-right circles, which mark the head, are referred to as *ul* and *ur*, respectively. Similarly, the lower-left and lower-right circles, which mark the transducer, are referred to as *ll* and *lr*. The center points of these two circles are 250.0 mm apart, and so the video scale is calculated on the basis of the distance between them in the image, as in (2).

$$(2) \quad scale_{video} = \frac{250}{\sqrt{(lr_x - ll_x)^2 + (lr_y - ll_y)^2}} mm / pixel$$

These calculations make it possible to adjust the two images to a single scale. We now turn to the other types of alignment, origin and rotation.

2.2.2 Aligning the “origin”

The midpoint of the curvature of the transducer (the “origin”), always appears in the ultrasound image at coordinates (390,416). The ultrasound view is contained in a fan-shaped region extending upward from the origin. The origin is crucial for reconciling the ultrasound and video images because it can be located in the video image as well; its location is predictable from the location of *ll* and *lr*. In this case, the distances from the transducer origin to *ll* and *lr*, respectively, are 72.5 mm and 314.0 mm (measured with a ruler). By the law of cosines, the angle from the *ll* point to the origin is 31.7° above the line between *ll* and *lr*, and the distance from *ll* to the origin in pixels (*orlldist*) is 72.5 divided by the video scale (3). This distance and angle allow the origin to be located with respect to *ll* and *lr*.

$$(3) \quad orlldist = \frac{72.5}{scale_{video}}$$

These calculations so far provide images of the same scale which can be aligned at a single point, the origin. We turn now to aligning the rotation of the two images, so that movements in the head and/or the transducer can be factored out.

2.2.3 Aligning the rotation

The purpose of the alignment algorithm is to align ultrasound images to a common standard, regardless of changes in head and transducer location and orientation. Therefore, the output of the algorithm includes both vertical and horizontal offset values (in pixels, for translation) and an angle (for rotation). All of these are based on the positions and angles of the head and transducer.

The transducer and head angles are computed as functions of the locations of the two points corresponding to each, as in (4-5).

$$(4) \quad tranangle = \arctan\left(\frac{ll_y - lr_y}{lr_x - ll_x}\right)$$

$$(5) \quad headangle = \arctan\left(\frac{ul_y - ur_y}{ur_x - ul_x}\right)$$

The actual coordinates of the origin in the video perspective is a function of the origin-*ll* distance (in pixels) and the angle of the transducer (6-7)

$$(6) \quad vidorigin_x = ll_x - (orlldist(\cos(31.7 - tranangle)))$$

$$(7) \quad vidorigin_y = ll_y - (orlldist(\sin(31.7 - tranangle)))$$

The next step is to locate the left head marker (*ul*) in the ultrasound perspective. The distance from the origin is based on the distance between the two points in the video perspective and the video and ultrasound scales (8). The angle is a function of the location of the two points in the video perspective and the transducer angle (9).

$$(8) \quad orheaddist = \frac{scale_{video}}{scale_{ultrasound}} \sqrt{(ul_x - vidorigin_x)^2 + (ul_y - vidorigin_y)^2}$$

$$(9) \quad orheadangle = \cos\left(\frac{ul_x - vidorigin_x}{\sqrt{(ul_x - vidorigin_x)^2 + (vidorigin_y - ul_y)^2}}\right) - tranangle$$

The vector from the origin to the ultrasound perspective version of *ul* (called *head*), is a function of the distance and angle (10-11).

$$(10) \quad head_x = 390 + (orheaddist)(\cos(orheadangle))$$

$$(11) \quad head_y = 416 - (orheaddist)(\cos(orheadangle))$$

In order to properly rotate around the center of the image, while translating the image as well, the vector from the center of the image to the ultrasound-perspective head is computed (12-13).

$$(12) \quad centerheaddist = \sqrt{(head_x - 360.5)^2 + (267.0 - head_y)^2}$$

$$(13) \quad centerheadangle = \arctan\left(\frac{head_x - 360.5}{267.0 - head_y}\right)$$

The horizontal and vertical offsets are functions of the distance from the center of the image (260.5, 267) to the ultrasound-perspective head, and the angles of the head and transducer (14-15). The arbitrary standard to which images are aligned is a situation in which the head has coordinates (750,-50), and the difference in angle between the head and transducer is zero. The angle by which the image is rotated is a function of the head and transducer angles, and the arbitrary standard (16).

$$(14) \quad offset_x = 750 - (360.5 + (centerheaddist)\sin(headangle - tranangle))$$

$$(15) \quad offset_y = -50 - (267.0 + (centerheaddist)\cos(headangle - tranangle))$$

$$(16) \quad offset_{angle} = headangle - tranangle - 0$$

A face outline can be constructed by applying a similar translation and resize to the face image shown in the camera view. As a result, any ultrasound image can be placed in the context of the palate and face outline of the speaker, as in Figure 4.

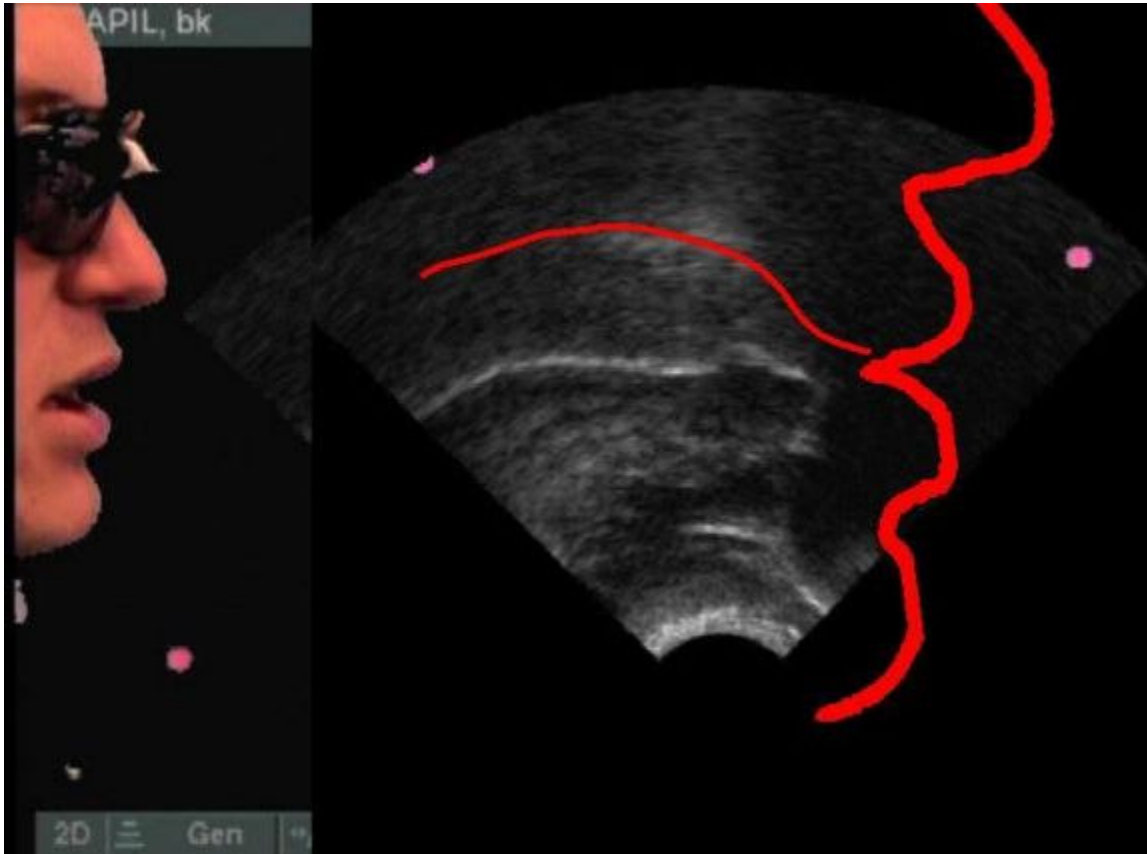


Figure 4. Ultrasound image placed in context of palate and face outline of speaker.

2.3 Automation

The algorithm described in the previous section is implemented as *Palatron*, a plug-in for Image-J, a public domain Java image processing program (Rasband 2004). The plug-in determines the scales of the ultrasound and video views by detecting scale ticks on the ultrasound output, extracts the coordinates of the four pink circles, and aligns ultrasound images with an overlay which includes outlines of the speaker's palate and face.

3. Results

The methods described here have been tested by examining images of stop consonants and confirming that the tongue position, combined with the projected palate location, is a complete closure.

4. Conclusion

Ultrasound has many advantages over other methods of imaging the tongue. But one of the major difficulties with using ultrasound for linguistic research has been the unreliability of the data due to images which were hard to quantify. Now, with our algorithm and *Palatron*, we have made an important step towards making ultrasound as useful for mainstream linguistic analysis as other methods. Because of the safe, non-invasive, portable, and relatively inexpensive nature of ultrasound, it is attractive to use for articulatory research. With the system presented here, we may now inexpensively collect reliable data. In addition, due to the portability of the method, we can collect data from populations from which it was traditionally much harder to collect verifiable lab data, such as children or native speakers in areas where there are no laboratories. These techniques open the door to the world of knowledge that can be gained from using ultrasound for linguistic research.

References

- Rasband, Wayne. 2004. Image-J. <http://rsb.info.nih.gov/ij/>.
- Stone, M., and E. P. Davis. 1995. Support system for ultrasound imaging. *J. Acoust. Soc. Am.* 98.6: 3107-12.
- Whalen, D. H., Khalil Iskarous, Mark K. Tiede, and David J. Ostry. 2004. HOCUS: The Haskins optically-corrected ultrasound system for measuring speech articulation. *J. Acoust. Soc. Am.* 115.5: 2632.

Jeff Mielke
Dept. of Linguistics
University of Arizona
Douglass Building, Room 200-E
PO Box 210028
Tucson, AZ 85721
mielke@u.arizona.edu

Adam Baker
Dept. of Linguistics
University of Arizona
Douglass Building, Room 200-E
PO Box 210028
Tucson, AZ 85721
tabaker@u.arizona.edu

Jeff Mielke, Adam Baker, Diana Archangeli, and Sumayya Racy

Diana Archangeli
Dept. of Linguistics
University of Arizona
Douglass Building, Room 200-E
PO Box 210028
Tucson, AZ 85721
dba@u.arizona.edu

Sumayya Racy
Dept. of Linguistics
University of Arizona
Douglass Building, Room 200-E
PO Box 210028
Tucson, AZ 85721
sracy@u.arizona.edu