# Quantitative Assessment of Tongue Shape and Movement Using Ultrasound Imaging

**Tim Bressmann**
**University of Toronto**

## 1. Introduction

In phonological research, it is often important to corroborate assumptions and models with speech data. While acoustic and perceptual evaluations are valuable, the motor basis of speech processes can only be understood fully if information about articulatory movement is included. The tongue is the most important organ for speech sound articulation. However, lingual movement is notoriously difficult to observe because the tongue is concealed in the oral cavity. Ultrasound imaging is rapidly gaining popularity as a tool for speech research in linguistics, speech sciences and speech-language pathology. Technological innovation and competition have lead to substantially lower costs for ultrasound machines, which are now becoming affordable to speech researchers. Ultrasound appeals to speech researchers because the method is biologically safe. The radiation levels that are generated by a medical ultrasound machine are very low and do not accumulate. Data acquisition is relatively easy, so that 'difficult' subjects such as children and patients can be studied. The real-time visualization of the tongue is visually pleasing and intuitive. This paper will provide a short introduction to ultrasound imaging for speech researchers. For more detailed information, the reader is referred to a recent special issue of *Clinical Linguistics and Phonetics* (Volume 19, issue 6-7, 2005), which focuses on ultrasound imaging in speech research. More extensive introductions to the topic may be found in Stone (2005) or Bressmann et al. (2005a).

Diagnostic ultrasound has been used in phonetic research to describe static and dynamic aspects of tongue shape and speech motor control (Stone et al. 1987, 1988). Recent research has used ultrasound to compare articulatory gestures in different languages (Davidson 2005) and to describe language specific preparatory settings for speech (Gick et al. 2004). Gick et al. (2005) describe the use of portable, battery powered ultrasound machines for field research. Ultrasound has also been used in clinical speech research with patients with strokes (Wein et al. 1993), glossectomy (Schliephake et al. 1998, Bressmann et al. 2005b, 2005c, in press), and malocclusions (Kikyo et al. 1999, Cheng et al. 2002). The tongue shape can be displayed in real-time on the screen, which makes ultrasound attractive to Speech-Language Pathologists as a tool for biofeedback in speech therapy (Shawker & Sonies 1985, Bernhardt et al. 2003, 2005).

## 2. How ultrasound imaging works

When a pulse of acoustic energy is directed at an object with suitable conductivity, it puts the object into oscillation and elicits (faint) echoes. This mechanism is referred to as the pulse-echo principle. The echoes can be used to gather information about the object. An example of a master sonographer of the animal world is the bat, which uses ultrasound to navigate a flight path in the dark. Medical ultrasound machines use piezoelectric crystals to generate sound waves in the megahertz (MHz) range. The crystals are set into oscillation by an electrical current. When the current is switched off, the crystals are set into receiving mode for the echoes. By repeating this process hundreds of times every second, a two-dimensional image can be reconstructed by a computer. Commercially available ultrasound machines deliver a video frame rate of 30 frames per second. This corresponds to the National Television Standard Committee (NTSC) format that is used in video equipment in North America. The NTSC video frame rate is sufficiently fast to capture most aspects of tongue movement in speech. However, very quick articulatory gestures such as plosives or tongue tip trills may not

 Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology, ed. Laura Colantoni and Jeffrey Steele, 101-106. Somerville, MA: Cascadilla Proceedings Project.

always be accurately displayed. The spatial accuracy for ultrasound images is usually within 1 mm accuracy for the current generation of machines (Kaburagi & Honda 1994). However, ultrasound also has important technical limitations: when the signal passes into air or bone, the sound wave is lost and no echo is passed back to the transducer because the conductivity for the sound is either too low (bone) or too high (air). As a result, air and bone will appear as black ‘shadows’ on the ultrasound image.

Figure 1. Ultrasound machine with endocavity transducer.

Figure 2. Head anchor and transducer holder for sonographic examinations.

Figure 1 shows an ultrasound machine with the transducer (upper right-hand corner). There are many different types of specialized ultrasound transducers available for different medical applications. Speech researchers will usually find endocavity or intercostal transducers most suitable. These transducers combine wide view angles with small footprints so that the tongue can be visualized along its length while interfering as little as possible with mandibular movement. The desirable frequency range for the transducers is between 3 and 7 MHz for imaging the tongue. Lower frequencies provide coarser images while higher frequencies provide a better resolution and more image detail. However, a higher frequency will be attenuated more strongly in the tissue, limiting the depth of the view field.

Ultrasound imaging requires a careful positioning of the transducer. The ultrasound image is acquired by holding the transducer under the chin of a speaker. If the transducer is held manually, the examiner’s hand may move or change the angle of the transducer. The subject’s head and shoulders may move, which can affect the coupling or the angle of the transducer. This is not necessarily a problem when one wants to obtain qualitative data or use the ultrasound display for real-time biofeedback. However, for quantitative measurements, many researchers use head and transducer holders to reduce movement artefacts (Stone & Davis 1995, Bressmann et al. 2005a). An example of such a head stabilizer is shown in Figure 2.

## 3. Ultrasound images and 2D data

The ultrasound image is usually displayed as a brightness scan (B-mode). The borders between different structures and layers of tissue are displayed as grey values. The interface between the tongue and the air is visible as a bright white band. Figure 3 shows a series of midsagittal ultrasound images of cardinal vowels. The midsagittal plane is the most common plane for ultrasound imaging because the image is most intuitive and can be compared between different speakers. However, ultrasound also has the great advantage of enabling a researcher to observe intrinsic lingual deformation (Stone et al. 1988).

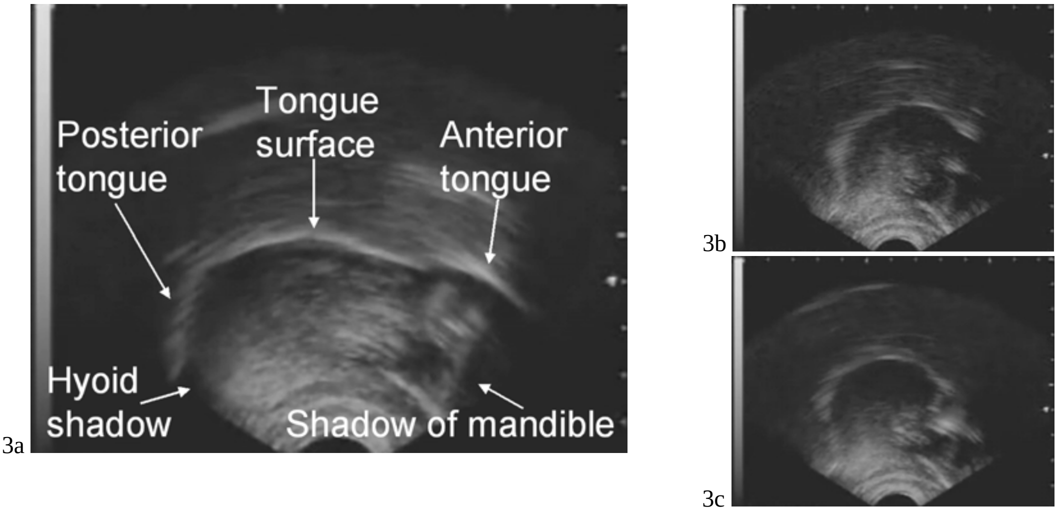


Figure 3. Midsagittal images of sustained cardinal vowels. The anterior tongue is towards the right side in all images; (a) sustained /ɑ:/; (b) sustained /i:/; sustained /u:/.

Figure 4 displays coronal images with and without midsagittal grooving. The video image of the ultrasound machine can be captured and digitized for quantitative analysis. However, the ultrasound images are inherently blurry. Consequently, the extraction of the tongue contour from a movement sequence is a complicated image processing problem. Figure 5 shows a screenshot of an image analysis program used to extract and measure tongue contours (Gu et al. 2004, Bressmann et al. 2005a). Figure 6 shows a waterfall display of Ultra-CATS measurements of midsagittal tongue movement during the sentence ‘It rang a lot’. It is a problem of the ultrasound data that a researcher will not be able to track individual flesh points in space. Instead, the tongue movement is measured in a transducer centred space (Rastadmehr et al. 2005).

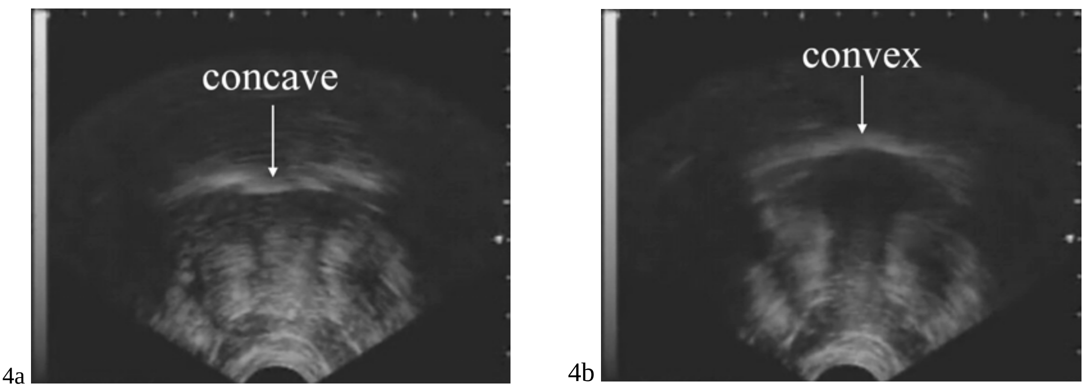


Figure 4. Coronal view of the posterior dorsum of the tongue; (a) midsagittal concavity during the production of /i:/; (b) midsagittal convexity during the production of /u:/.

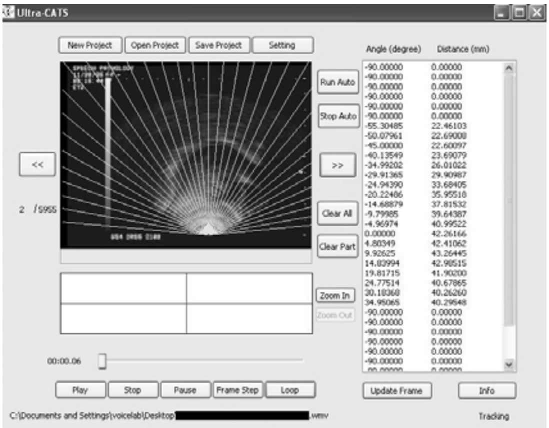


Figure 5. Screenshot of the Ultra-CATS software (Gu et al. 2004) with measurement grid superimposed on a midsagittal tongue image.

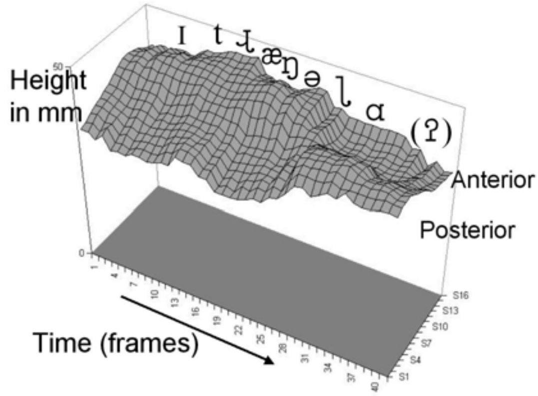


Figure 6. Midsagittal tongue movement during the production of the sentence ‘It rang a lot’, with transcription.

## 4. Static and dynamic 3D ultrasound imaging

The tongue is a three-dimensional structure. For some research applications, it is desirable to collect 3D information, rather than just two-dimensional cross sections. Static 3D imaging of the tongue can be accomplished by collecting multiple two-dimensional slices from a sustained speech sound and assembling them into a 3D volume. This can be done by either moving the transducer through a set of predefined positions (Stone & Lundberg 1996) or by tracking the movement of the transducer in space (Bressmann et al. 2005b, 2005c, accepted). Figure 7 shows an image of a reconstructed three-dimensional ultrasound volume.

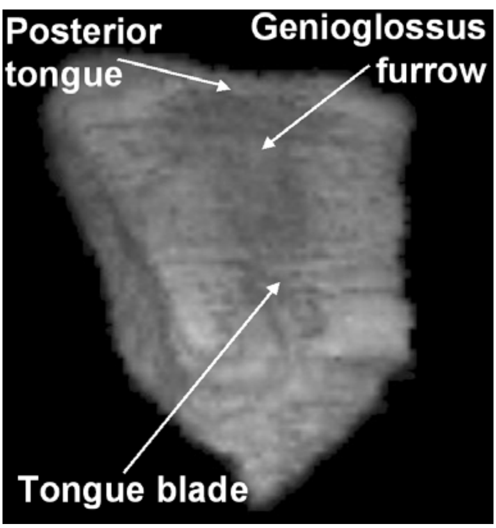


Figure 7. 3D tongue volume of a sustained /a:/.

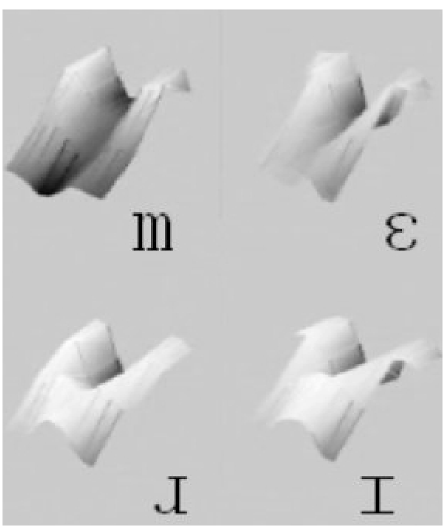


Figure 8. Reconstruction of 3D tongue movement during the word ‘merry’, using the multi-planar paced sonography method.

A number of newer ultrasound machines offer so-called 4D imaging, i.e., dynamic volume acquisition with a special transducer. While most of these machines do not offer volume acquisition rates that can match fast tongue movement during speech, the development of crystal matrix transducers and increased computational power will make dynamic volume acquisition possible in the future (see the discussion in Chi-Fishman 2005). In the meantime, 3D tongue movement can be approximated from repeated measurements in multiple viewplanes. Yang & Stone (2002) recorded repeated utterances of the same sentence in multiple parallel sagittal and coronal planes. The researchers then time-aligned the repetitions using a Dynamic Programming procedure and reconstructed the 3D movement data. Flowers et al. (2005) describe a similar method of reconstructing 3D tongue movement from a sparse set of coronal slices. The participants' speech was paced with a metronome (multi-planar paced sonography). Figure 8 shows a sequence of tongue surfaces reconstructed in this way. However, this method is demanding for the research participant so it is not suitable for clinical research populations.

## 5. Conclusions

Ultrasound offers exciting new opportunities for researchers in linguistics and speech-language pathology. It is safe, non-invasive, and cost-effective. The process of ultrasound data acquisition is relatively comfortable for the subject, and it is possible to acquire extensive amounts of tongue movement data. The direct visualization of the tongue shape on the ultrasound screen is often a strong motivation for research participants. The costs for ultrasound machines have dropped significantly over recent years, so it is likely that there will be continued interest in this technology for use in speech research.

## References


Bernhardt, Barbara, Gick, Brian, Bacsfalvi, Penelope, & Adler-Bock, Marcy. (2005). Ultrasound in speech therapy with adolescents and adults. Clinical Linguistics and Phonetics, 19, 605-617.

Bernhardt, Barbara, Gick, Brian, Bacsfalvi, Penelope, & Ashdown, Julie. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. Clinical Linguistics and Phonetics, 17, 199-216.

Bressmann, Tim, Uy, Catherine, & Irish, Jonathan. (2005a). Analysing normal and partial glossectomee tongues using ultrasound. Clinical Linguistics and Phonetics, 19, 35-52.

Bressmann, Tim, Thind, Parveen, Uy, Catherine, Bollig, Carmen, Gilbert, Ralph, & Irish, Jonathan. (2005b). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from 12 normal speakers and a partial glossectomee. Clinical Linguistics and Phonetics, 19, 573-588.

Bressmann, Tim, Heng, Chiang-Le, & Irish, Jonathan. (2005c). Applications of 2D and 3D ultrasound imaging in speech-language pathology. Journal of Speech-Language Pathology and Audiology, 29, 158-168.

Bressmann, Tim, Ackloo, Elizabeth, Heng, Chiang-Le, & Irish, Jonathan. (accepted). Quantitative 3D ultrasound imaging of partially resected tongues. Otolaryngology – Head and Neck Surgery.

Cheng, Chian-Fen, Peng, Chien-Lun, Chiou, Hung-Yi, & Tsai, Chi-Yang. (2002). Dentofacial morphology and tongue function during swallowing. American Journal of Orthodontics and Dentofacial Orthopedics, 122, 491-499.

Chi-Fishman, Gloria. (2005). Quantitative lingual, pharyngeal and laryngeal ultrasonography in swallowing research: A technical review. Clinical Linguistics and Phonetics, 19, 589-604.

Davidson, Lisa. (2005). Addressing phonological questions with ultrasound. Clinical Linguistics and Phonetics, 19, 619-633.

Flowers, Heather, Bressmann, Tim, Carmichael, Brent, Heng, Chiang-Le, & Wong, Willy. (2005). Reconstructing 3D tongue movement trajectories from multiplanar paced ultrasound scans. Canadian Acoustics, 33, 50-51.

Gick, Brian, Wilson, Ian, Koch, Karsten, & Cook, Clare. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. Phonetica, 61, 220-233.

Gick, Brian, Bird, Sonya, & Wilson, Ian. (2005). Techniques for field application of lingual ultrasound imaging. Clinical Linguistics and Phonetics, 19, 503-514.

Gu, Jiayun, Bressmann, Tim, Cannons, Kevin, & Wong, Willy. (2004). The Ultrasonographic Contour Analyzer for Tongue Surfaces (Ultra-CATS). Toronto: University of Toronto.

Kaburagi, Tokihiko, & Honda, Masaaki. (1994). An ultrasonic method for monitoring tongue shape and the position of a fixed point on the tongue surface. Journal of the Acoustical Society of America, 95, 2268-2270.

Kikyo, Tomoaki, Saito, Miki, & Ishikawa, Masaaki. (1999). A study comparing ultrasound images of tongue movements between open bite children and normal children in the early mixed dentition period. Journal of Medical and Dental Sciences, 46, 127-137.

Rastadmehr, Orchid, Bressmann, Tim, Heng, Chiang-Le, & Irish, Jonathan. (2005). Acoustic and articulatory space before and after lateral tongue resections in oral cancer patients. Canadian Acoustics, 33, 82-83.

Schliephake, Henning, Schmelzeisen, Rainer, Schönweiler, Rainer, Schneller, Thomas, & Altenbernd, Christine. (1998). Speech, deglutition and life quality after intraoral tumour resection. A prospective study. International Journal of Oral and Maxillofacial Surgery, 27, 99-105.

Shawker, Thomas H., & Sonies, Barbara C. (1985). Ultrasound biofeedback for speech training. Instrumentation and preliminary results. Investigative Radiology, 20, 90-93.

Stone, Maureen (2005). A guide to analysing tongue motion from ultrasound images. Clinical Linguistics and Phonetics, 19, 455-501.

Stone, Maureen, & Davis, Edward P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. Journal of the Acoustical Society of America, 98, 3107-3112.

Stone, Maureen, & Lundberg, Andrew. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. Journal of the Acoustical Society of America, 99, 3728-3737.

Stone, Maureen, Morrish, Karl, Sonies, Barbara C., & Shawker, Thomas H. (1987). Tongue curvature: A model of shape during vowel production. Folia Phoniatrica, 39, 302-315.

Stone, Maureen, Shawker, Thomas H., Talbot, Thomas L., & Rich, Alan H. (1988). Cross-sectional tongue shape during the production of vowels. Journal of the Acoustical Society of America, 83, 1586-1596.

Wein, Berthold, Angerstein, Wolfgang, & Klajman, Stanislav. (1993). Suchbewegungen der Zunge bei einer Sprechapraxie: Darstellung mittels Ultraschall und Pseudo-3D-Abbildung. Nervenarzt, 64, 143-145.

Yang, Changsheng, & Stone, Maureen (2002). Dynamic programming method for temporal registration of three-dimensional tongue surface motion from multiple utterances. Speech Communication, 38, 199-207.

# Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology

edited by Laura Colantoni and Jeffrey Steele

Cascadilla Proceedings Project Somerville, MA 2008

## Copyright information



ISBN 1-57473-424-9 library binding



## Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

## Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Bressmann, Tim. 2008. Quantitative Assessment of Tongue Shape and Movement Using Ultrasound Imaging. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, ed. Laura Colantoni and Jeffrey Steele, 101-106. Somerville, MA: Cascadilla Proceedings Project.

or:

Bressmann, Tim. 2008. Quantitative Assessment of Tongue Shape and Movement Using Ultrasound Imaging. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, ed. Laura Colantoni and Jeffrey Steele, 101-106. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1717.