Applications of 2D and 3D Ultrasound Imaging in Speech-Language Pathology

Utilisation de l'échographie en 2D et 3D en orthophonie

Tim Bressmann, Chiang-Le Heng and Jonathan C. Irish

Abstract

Tongue motion in speech and swallowing is difficult to image because the tongue is concealed in the oral cavity. It is even more difficult to assess the extent of lingual motion quantitatively. Ultrasound imaging of the tongue in speech production and swallowing allows for a safe and non-invasive data acquisition. The paper describes the potentials and methodological problems of conducting ultrasound speech research using dynamic two-dimensional, static three-dimensional and dynamic three-dimensional ultrasound imaging.

Abrégé

Le mouvement de la langue pour la parole et la déglutition est difficile à mettre en image parce que cet organe est dissimulé dans la cavité buccale. Il est encore plus difficile d'évaluer l'ampleur du mouvement de la langue de manière quantitative. L'échographie de la langue lors de la production de la parole et de la déglutition constitue une méthode sécuritaire et non invasive d'obtenir des données. Le présent article décrit les possibilités et les problèmes méthodologiques pour mener des recherches sur l'utilisation de l'échographie dynamique en deux dimensions, statique en trois dimensions et dynamique en trois dimensions dans le domaine de l'orthophonie.

Key Words: Glossectomy, cancer, tongue paralysis, speech, articulation, swallowing, tongue, 3D ultrasound, 2D ultrasound

Tim Bressmann, Ph.D.
Graduate Department of Speech-Language Pathology
University of Toronto
Toronto, ON Canada

Chiang-Le Heng, B.Sc.
Graduate Department of Speech-Language Pathology
University of Toronto
Toronto, ON Canada

Jonathan C. Irish, M.D., M.Sc., F.R.C.S.C., F.A.C.S.
Departments of Otolaryngology and Surgical Oncology
Princess Margaret Hospital
University Health Network and Toronto General Hospital, University Health Network
Toronto, ON Canada

Introduction

Ultrasound imaging of the tongue is currently gaining popularity as a research tool in speech-language pathology and speech science. Until ten years ago, ultrasound machines were out of reach for most speech researchers because they were very expensive. In recent years, advances in computer technology and increased competition between the different manufacturers of ultrasound machines have helped to bring down the costs considerably. As a consequence, more speech researchers are now able to purchase ultrasound machines for their laboratories. Also, more physicians are buying machines for their hospital or private practice. This in turn may potentially give speech-language pathologists who are affiliated with hospitals or physicians in private practice access to ultrasound machines.

The main advantages of ultrasound imaging over other methods in phonetic research lie in low cost, bio-safety and ease of image acquisition. After an ultrasound machine has been purchased, associated costs to collect data are negligible. The radiation levels that are generated by a medical ultrasound machine are extremely low and do not accumulate so it is biologically safe to make extended recording sessions or examine patients repeatedly. This is an advantage over x-ray based imaging methods such as videofluoroscopy. Ultrasound imaging is non-invasive for the patient because it is not necessary to glue transducer coils to the tongue (like in electromagnetic midsagittal articulography). The ultrasound data acquisition is reasonably comfortable for the participant so that clinical populations can be studied. It also becomes easier to make recordings with notoriously 'difficult' research populations such as children.

In this paper, we will give an introduction to a number of different applications that we see for ultrasound imaging in speech-language pathology and demonstrate research findings from different projects that were undertaken in the Voice and Resonance Laboratory at the University of Toronto.

Two-Dimensional Ultrasound

Diagnostic ultrasound makes use of the pulse-echo principle: Any physical impulse into an environment puts objects into oscillation and results in echoes. In the ultrasound machine, a piezoelectric crystal in the ultrasound transducer generates a sound-burst. After sending the burst, the same piezoelectric crystal is then put into receiving mode and listens 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 standard NTSC format that is used in television sets and video-recorders in North America. The NTSC video-frame rate is sufficiently fast to capture even the quicker aspects of tongue movement in speech such as those involved in the production of plosives.

Diagnostic ultrasound has long been used in phonetic research to examine the tongue shape in different speech sounds (Morrish, Stone, Sonies, Kurtz, & Shawker,1984; Shawker & Sonies, 1984; Stone, Morrish, Sonies, & Shawker, 1987, 1988; Wein, Bockler, Huber, Klajman, & Wilmes, 1990) as well as to assess temporal aspects of speech motor control (Munhall, 1985; Parush & Ostry, 1993). Much of the clinical application of ultrasound imaging for the study of tongue function to date has focused on the study of the oral phase of swallowing (Casas,Kenny, & Macmillan, 2003; Chi-Fishman, Stone, & McCall, 1998; Neuschäfer-Rube,Wein, Angerstein, Klajman, & Fisher-Wein, 1997; Peng, Jost-Brinkmann, Miethke, & Lin, 2000; Soder & Miller, 2002; Sonies, Baum, & Shawker, 1984; Shawker, Sonies, Stone, & Baum, 1983; Stone & Shawker, 1986). The method has proven useful even for babies (Bosma, Hepburn, Josell, & Baker, 1990). Swallowing and speech have been studied in different pathological populations including patients with cerebral palsies (Casas, Kenny, & McPherson, 1994; Casas, McPherson, & Kenny,1995; Kenny, Casas, & McPherson, 1989; Sonies & Dalakas, 1991), strokes (Wein, Alzen, Tolxdorff, Bockler, Klajman, & Huber, 1988a; Wein, Angerstein, & Klajman, 1993), geriatric patients (Sonies et al., 1984), glossectomy (Schliephake, Schmelzeisen, Schönweiler, Schneller, & Altenbernd, 1998), and malocclusions (Cheng, Peng, Chiou, & Tsai, 2002; Kikyo, Saito, & Ishikawa, 1999).

The fact that the tongue shape can be displayed in real time on the screen makes ultrasound potentially very attractive to speech-language pathologists as a tool for biofeedback for oral deaf speakers and also for patients with dysarthrias or compensatory articulation errors associated with cleft palate. However, only a few studies so far have used ultrasound as a tool for biofeedback in speech therapy. Shawker and Sonies (1985) described the use of ultrasound imaging for the speech therapy of an individual with an articulation disorder. The authors found that the subject was able to improve her articulatory distortions of the /r/ sound over a course of 3 months of ultrasound biofeedback therapy. A recent study by Bernhardt et al. (Bernhardt, Bacsfalvi, Gick, Radanov, & Williams, this issue; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003) compared ultrasonographic and electropalatographic feedback in four speakers with mild to severe hearing loss. The authors found that all subjects were able to improve their articulation and that both feedback methods were equally effective.

For the research in the Voice and Resonance Laboratory at the University of Toronto, we use a low-end General Electric Logiq Alpha 100 MP ultrasound machine with a 6.5 MHz micro convex curved array scanner with a 114° view (Model E72, General Electric Medical Systems, P.O. Box 414, Milwaukee, Wisconsin 53201). The machine and the transducer are displayed in Figure 1. During the ultrasound examination, the video-output of the ultrasound machine is captured with a generic digital video camera (Canon ZR 45 MC, Canon Canada Inc., 6390 Dixie Road, Mississauga, Ontario L5T 1P7). Parallel sound recordings are made onto the same digital videotape using an AKG C420 headset condenser microphone (AKG Acoustics, 914 Airpark Center Drive, Nashville, Tennessee 37217) with a Behringer Ultragain Pro 2200 line-driver (Behringer Ltd., 18912 North Creek Pkwy, Suite 200, Bothell, Washington 98011). After the recording, the ultrasound films are downloaded from the digital video camera onto a computer and saved as a digital file. Figure 2 shows typical midsagittal tongue contours during the sustained production of the cardinal vowels /a/, /i/, and /u/.

Problems in 2D ultrasound imaging: Head fixation

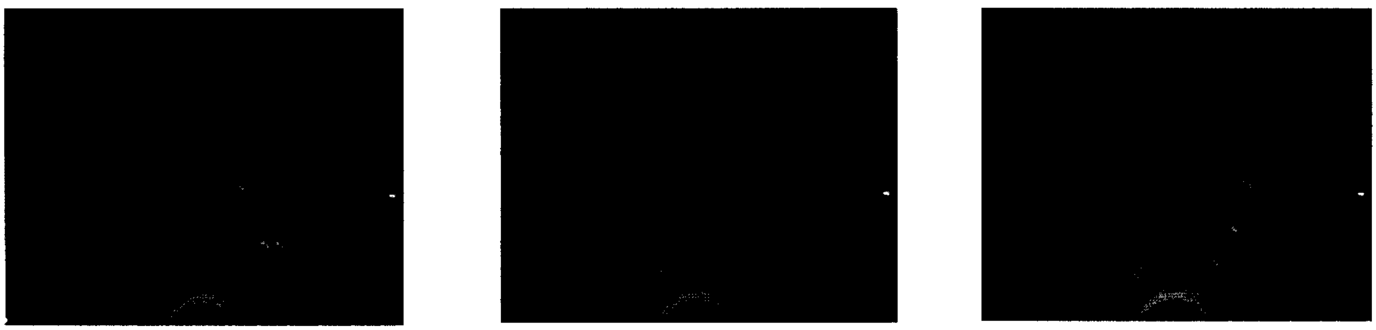
The ultrasound image is acquired by holding the transducer against the neck of the research participant. If the transducer is held manually against the neck of a subject who sits in a standard office chair, there are a number of moving elements that might lead to measurement error:

- The examiner's hand with the ultrasound transducer may move or change the angle of the transducer;
- The subject's mandible moves up and down which may change the position of the transducer;
- The subject's head and shoulders may move which can affect the coupling or the angle of the transducer.

Figure 1

Figure 1. Ultrasound machine with 114° endocavity transducer and ultrasound gel. The image also shows the PC Bird electromagnetic movement tracking system, which is used to reconstruct three-dimensional ultrasound volumes.

Figure 2



(a) (b) (c)

Figure 2. Midsagittal ultrasound contours of the tongue during the production of sustained vowels. The anterior tongue is towards the right side of the images. To facilitate viewing, the tongue contours are marked with a grey line. (a) sustained /a/; (b) sustained /i/; (c) sustained /u/.

If one wishes to generate quantitative tongue movement data from an ultrasound film, it is important to reduce movement of the transducer and the subject's head. On the other hand, the fixation mechanism should ensure good coupling during regular mandibular movement in speech. Different groups of researchers have used head stabilization devices such as headrests for the back of the head (Davidson , 2004; Stone et al., 1988), headrests for the forehead (Peng et al., 2000), complete head fixation (Stone & Davies, 1995), helmets with transducer attachments (Hewlett, Vasquez, Zharkova, & Zharkova, 2004) or position control with laser pointers (Gick, 2002). In an alternative approach, Whalen et al. (2004) suggested minimizing the head movement with a headrest and tracking the residual movement with a three-dimensional optical tracking system.

In our laboratory at the University of Toronto, we developed the Comfortable Head Anchor for Sonographic Examinations (CHASE; Carmichael, 2004) that was roughly modelled on the device described by Peng et al. (2000). The CHASE, which is depicted in Figure 3, consists of a headrest for the subject's forehead and a transducer cradle with a suspension-spring mechanism. Since a large number of our research participants are head and neck cancer patients, it was our goal for the development of the CHASE to make the device as unintimidating as possible. For this reason, the CHASE only anchors and stabilizes the participant's head while avoiding forced head fixation. Our experience to date shows that this effectively reduces all head and transducer movement to an acceptable minimum (less than 1.5 mm lateral wandering after ten minutes of speech recordings).

Figure 3

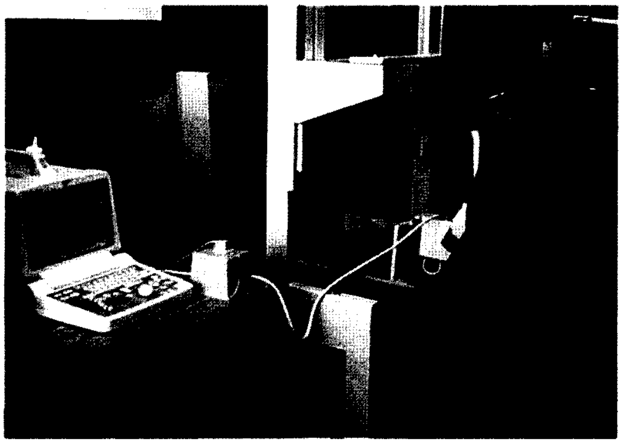


Figure 3. A research participant in the CHASE head anchor.

Problems in 2D ultrasound imaging: Image analysis

The set-up for ultrasound imaging and the data acquisition require minimal advance preparation. However, the data analysis can be labour-intensive. Every second of film generates 30 separate image frames, and we are interested in the positions of different parts of the tongue. It is extremely time consuming to do these analyses by hand. It is therefore desirable to automate the data analysis as much as possible.

At first, the automatic extraction of tongue contours from an ultrasound film may seem a trivial task because the tongue movement can be easily visualized and appears clearly on the screen of the ultrasound machine. However, ultrasound is an acoustic imaging method and, as a consequence, the image is often noisy and may contain artefacts. On close inspection, the tongue contours that look so clear to the experimenter's eye are in fact blurry patches of diffuse shades of grey. The human observer has the advantage of Gestalt perception, which means that a pattern of elements is perceived as a unified whole. While the moving tongue surface is easily discernible for the human observer, the automatic extraction of tongue contours from an ultrasound film is a challenging problem for computer vision programming.

Over the years, the researchers at the Vocal Tract Visualization Laboratory at the University of Maryland have developed a number of different successful motion trackers (Akgul, Kambhamettu, & Stone, 1999; Unser & Stone, 1992). The current version of the EdgeTrak software (Li, Kambhamettu, & Stone., 2003) can be downloaded at http://speech.umaryland.edu. Other programs are currently being developed at Queen Margaret University College in Edinburgh (Wrench , 2004) and at the University of British Columbia in Vancouver (Gick & Rahemtulla, 2004). However, even the best automatic motion trackers will often lose a tongue contour that they are tracking. The deformation of the tongue shape can be very rapid. A good example for a situation that often leads to the failure of a motion tracker is when the speaker changes from a low to a high vowel. The tracking of ultrasonographic tongue contours is probably more an artificial intelligence rather than an image-processing problem, and the current generation of motion trackers cannot be expected to perform flawlessly. It is therefore important that the user reviews and corrects the automatic tracking results.

We recently completed the development of our own software, named the Ultrasonographic Contour Analyzer for Tongue Surfaces (Ultra-CATS; Gu, Bressmann, Cannons, & Wong, 2004). The Ultra-CATS software can be downloaded from www.slp.utoronto.ca/People/Labs/TimLab/ultracats.htm. The Ultra-CATS was designed with a focus on the semi-automated analysis of the ultrasound data. The program also incorporates an automatic tracking option. The main goal of the software was to facilitate the manual frame-by-frame analysis of an image sequence. In the semi-automated analysis, the user traces the tongue surface with a drawing tool on each image frame. The software then measures points on the tongue surface with a polar grid. In our experience, the manual tracing of a single ultrasound frame will take a trained experimenter about 7 seconds on average. At this pace, every ten seconds of continuous ultrasound film will take approximately 35 minutes to analyze. An additional feature of the Ultra-CATS software is an automatic image-processing algorithm. The automatic tracker can be set, run, stopped, corrected and set back on track at any point during the analysis. The Ultra-CATS program saves all measurements to a text file so that they can be edited and analyzed using a spreadsheet editor or a statistical analysis program. An image of the program interface of the Ultra-CATS can be found in Figure 4.

Since we have completed the development of the Ultra-CATS software, we have used it to analyze the speech and swallowing of normal speakers as well as patients with tongue cancer and lingual paralyses. Figure 5 shows a waterfall display for a water swallow of a normal male participant. The waterfall display shows the elevation of the back of the tongue that prevents predeglutitive aspiration. As the oral transport phase of the swallow begins, the back of the tongue lowers in order to allow passage of the bolus. We can then appreciate the progressive elevation of the tongue from the front to the back as the bolus is cleared from the oral cavity.

Figure 6 shows a waterfall display of the phrase 'ninety-three years old'. This segment was taken from a reading of the second sentence of the Grandfather passage ('Well, he is nearly ninety-three-years old'; van Riper, 1963), spoken by a normal female speaker. The displayed data represent 1.3 seconds of data. Note the large number of posture adjustments that the tongue makes to produce all the required phonemes. Also note the immediate anticipatory elevation of the anterior dorsum of the tongue after the final /d/ plosion, which leads into the first high front vowel of the next sentence ('He dresses himself in an ancient black frock coat ...').

Figure 7 shows two waterfall displays of repeated syllables. In Figure 7a, we see five repeated utterances of

Figure 4

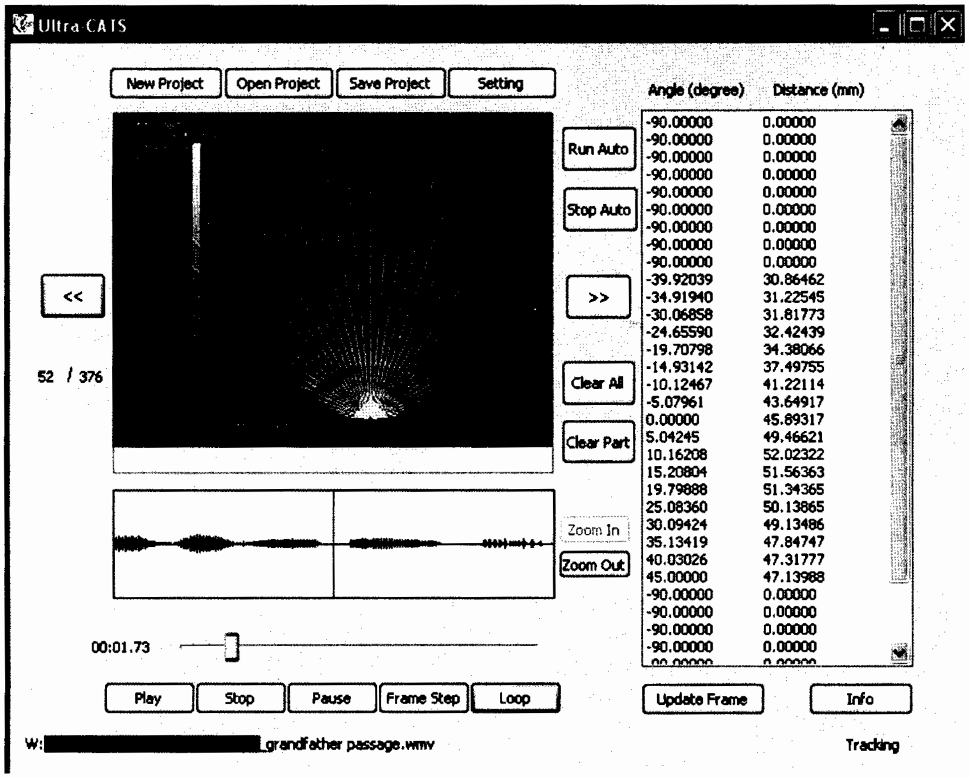


Figure 4. Screenshot of the Ultra-CATS software

Figure 5

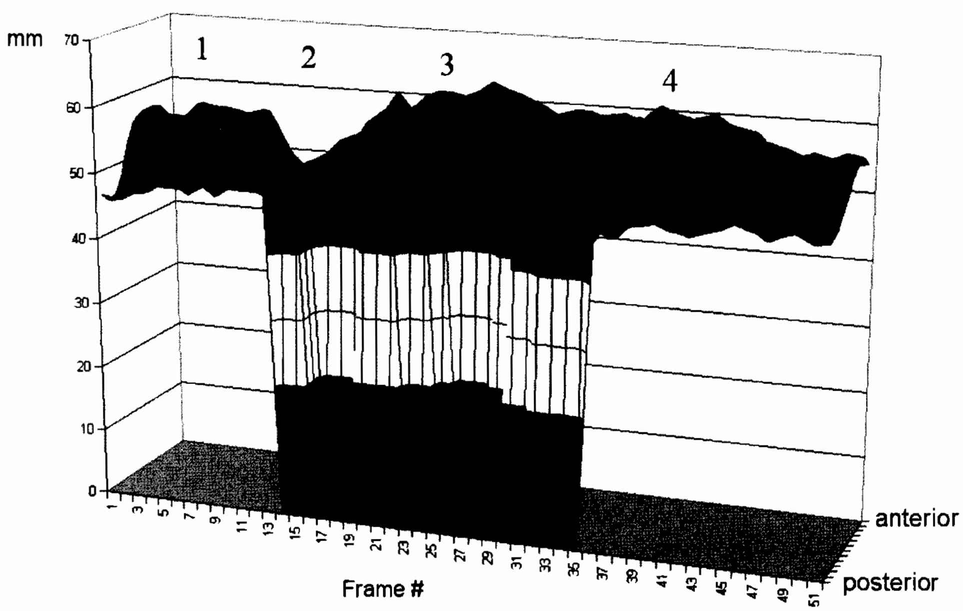


Figure 5. Waterfall display of a water swallow by a normal female participant. The numbers indicate different phases of the swallow. (1) The dorsum of the tongue is elevated to prevent predeglutitive aspiration. (2) The swallow is initiated. The tongue lowers so that the water bolus can pass into the oropharynx. (3) First the tip and then the dorsum of the tongue elevate to clear the bolus from the oral cavity. During this part of the swallow, the hyoid bone moves forward to open the upper oesophageal sphincter. The forward movement of the hyoid is indicated by the missing data/ zeros which are visible in the posterior tongue during this phase of the swallow. (4) After the swallow, the tongue returns to a neutral rest position and then lowers in preparation for the next swallow.

Figure 6

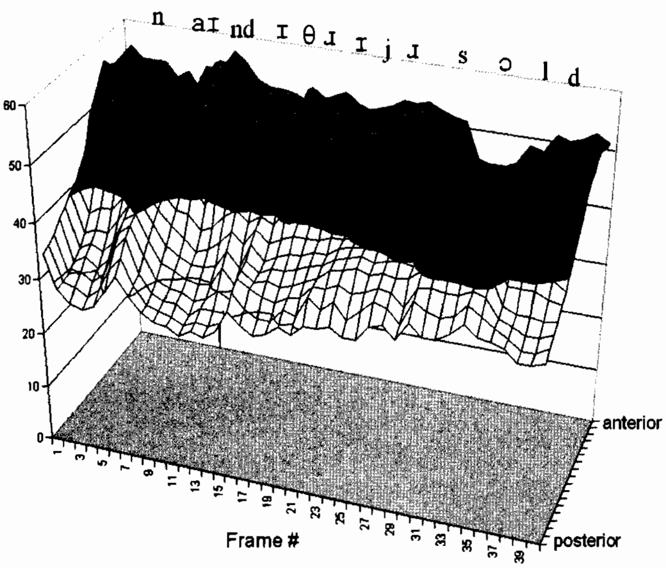


Figure 6. Waterfall display of the midsagittal tongue contours during the phrase 'ninety-three years old' from the second sentence of the Grandfather passage ('Well, he is nearly ninety-three years old.'), spoken by a normal adult female participant.

the syllable /aka/ by a normal female speaker. The lowering of the tongue from the rest position towards the position for the /a/ can be appreciated. The back of the tongue then elevates to achieve velar closure for the /k/, and the procedure is repeated over. Figure 7b demonstrates the same repeated syllables spoken by an older female patient with a flaccid paralysis of the tongue resulting from post-Polio syndrome. It can be observed that the lingual paralysis leads to an undifferentiated elevation of the tongue, rather than the lowering of the tongue for /a/, which perceptually resulted in a centralized vowel. The patient cannot elevate the posterior tongue for /k/. Perceptually, this articulatory undershoot results in significantly reduced speech intelligibility.

Static Three-Dimensional Ultrasound

The 2D display of tongue contours is interesting and affords us fascinating insights into the movement of the tongue. However, the tongue is a complex, non-rigid three-dimensional structure (Hiiemae & Palmer, 2003; Stone 1990). It is especially important to recover the three-dimensional information about the shape of the tongue in glossectomy patients because the lingual resection and reconstruction rarely ever lead to a symmetrical outcome. This means that a midsagittal image will often be an incomplete representation of the surgically altered tongue shape. So far, three-dimensional ultrasound has mostly been used in feasibility studies in normal speakers (Lundberg and Stone, 1999; Stone and Lundberg, 1996; Watkin and Rubin, 1989; Wein, Klajman, Huber, & Doring, 1988b). While it was demonstrated that three-dimensional ultrasound has the capabilities to deliver exact representations of the lingual surface in different speech sounds, the above studies remained descriptive and did not attempt to quantify lingual movement ranges in the reconstructed three-dimensional volume. However, a quantitative approach to the three-dimensional deformation shape of the tongue during the production of speech sounds would be particularly desirable for the analysis of patients with glossectomy in order to compare pre- and postoperative movement ranges and to evaluate the effect of different reconstructive techniques on the deformation and symmetry of the tongue tissue.

Figure 7

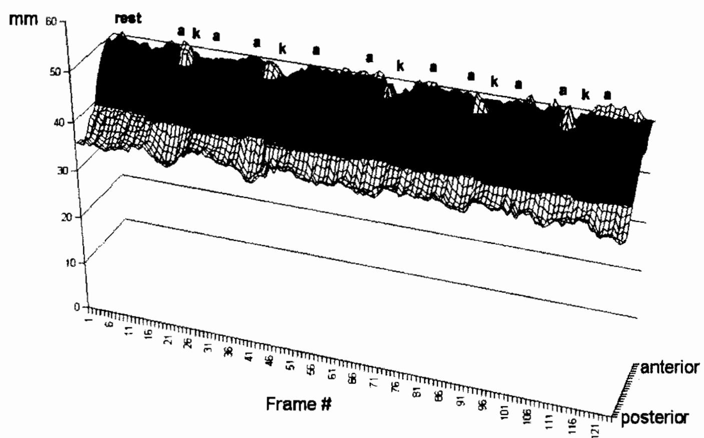


Figure 7b

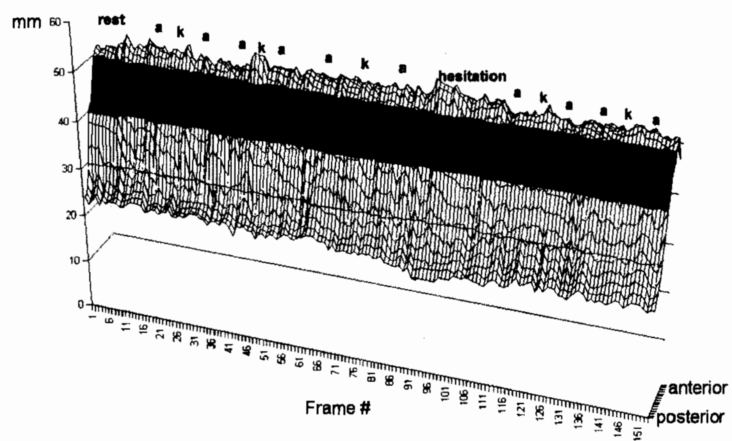


Figure 7. Waterfall display of the midsagittal tongue contour during five repetitions of /aka/. (a) Normal adult female participant. Note the elevated tongue rest position before the beginning of the utterance. (b) Female patient with lingual paralysis resulting from post-Polio syndrome.

Figure 8

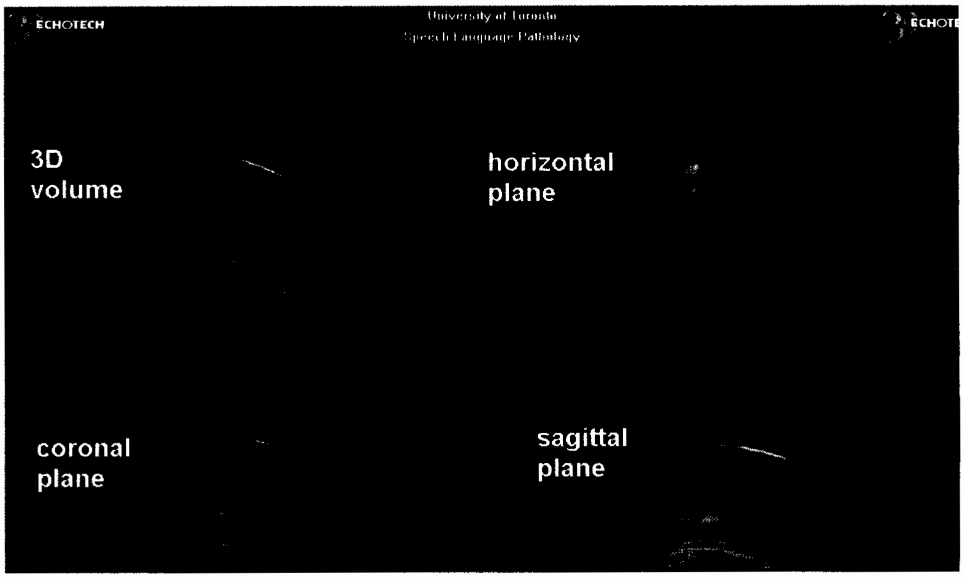


Figure 8. Orthogonal planes of a three-dimensional ultrasound volume of the sustained vowel /a/, spoken by a normal male adult participant.

In a series of ongoing studies at the Voice and Resonance Laboratory at the University of Toronto, we use our ultrasound machine in combination with a three-dimensional motion sensor (PC Bird, Ascension Technology Corporation, P.O. Box 527, Burlington, Vermont 05402). The FreeScan V7.04 computer program (Echotech 3D Imaging Systems, 85399 Halbergmoos, Germany) is used for the data acquisition and the reconstruction of the three-dimensional volumes. This set-up allows us to make three-dimensional scans of static structures and, consequently, all volume scans have to be obtained from sustained speech sounds. During the 3D data acquisition procedure, the subject is seated upright on a chair and instructed to slightly overextend his or her head. The transducer is held in a coronal scanning position and swept from the chin to the upper border of the thyroid cartilage. In a typical examination, a research participant sustains the following English phonemes: /a/, /i/, /u/, /s/, /ʃ/, /ɹ/, /l/, /n/ and /ŋ/. Each speech sound is repeated three times. The sound is sustained for approximately 5 seconds while the ultrasound scan is made. A 3D ultrasound scan of the tongue usually takes 2-3 seconds.

Using the FreeScan software, we can then browse through the three-dimensional ultrasound volume in any direction. Figure 8 illustrates how multiple planes of a three-dimensional ultrasound volume can be visualized. The sound shown is sustained /a/ spoken by a normal male speaker. Figure 9 shows the reconstructed three-dimensional tongue volume of the same sustained /a/. In order to make measurements of the 3D tongue surface, it is important to define an anchor point in the tongue upon which all measurements can be based. We first align all 3D scans so that the lingual septum is as exactly vertical as possible and that muscle fibres between the chin and the hyoid bone that are formed by the geniohyoid and the inferior genioglossus are as exactly horizontal as possible. We then define the 'midpoint' of the tongue as the halfway point between the mandible and the hyoid on the superior border of the geniohyoid muscle in the sagittal view, and as the point of intersection of the lingual septum and the geniohyoid muscle in the coronal view. Using this midsagittal midpoint, we then identify a left and a right parasagittal plane that is exactly parallel to the midsagittal slice. Based on the anchor point, we then superimpose a concentric grid with measurement lines spaced out in 11.25° intervals on the slices and measure the sagittal tongue form in three parallel sagittal planes. Figure 10 demonstrates how we take measurements in the midsagittal

Figure 9

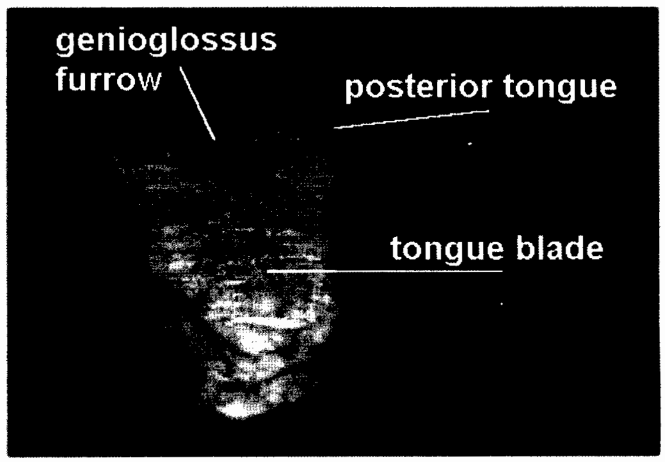


Figure 9. Reconstructed three-dimensional volume of the sustained /a/ from Figure 8.

Figure 10

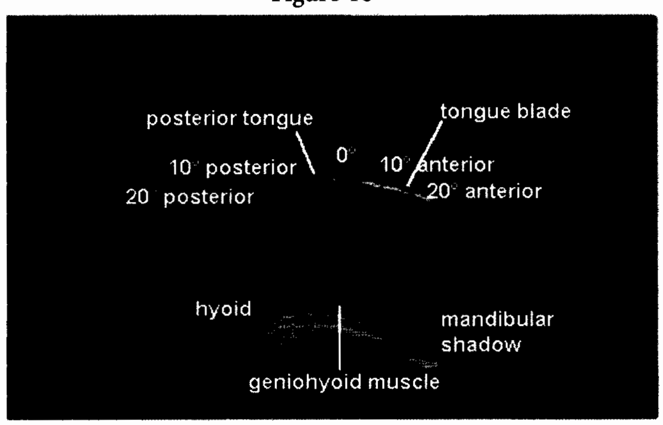


Figure 10. Midsagittal plane of a sustained /a/ with overlay of a measurement grid.

plane of the same ultrasound volume of the sustained /a/ shown in Figures 8 and 9.

This procedure generates a data matrix for the tongue surface for a speech sound. The data can be used to reconstruct a rough visual representation of tongue surface shapes as demonstrated in Figure 11. Figure 11a shows a composite of tongue surface data during the production of /ʃ/ for 12 normal speakers. Note the midsagittal groove for this sound that is necessary to produce a clear /ʃ/. Figure 11b shows the tongue of a patient with a lateral carcinoma of the tongue during the production of the same sound before the ablative cancer surgery to the right side of her tongue. The patient was able to produce an adequate midsagittal groove and her /ʃ/ was perceptually acceptable. Figure 11c shows the same patient after a lateral resection of the right side of her tongue and defect reconstruction with a radial forearm flap. The patient was now unable to form a consistent midsagittal groove, which led perceptually to a distortion and lateralization of the /ʃ/sound.

In two studies (Bressmann, Uy, & Irish, 2005; Bressmann, Thind, Uy, Bollig, Gilbert, & Irish, 2005), we used the quantitative tongue surface data that we generated from three-dimensional tongue volumes to extract underlying components of tongue movement by using mathematical procedures such as principal component analysis or multi-dimensional scaling. We also developed a number of quantitative descriptors for the degree of protrusion of the tongue in the oral cavity (anteriority index), the degree of three-dimensional midsagittal grooving along the length of the tongue (concavity index) and the symmetry of the elevation of left and right lateral tongue (asymmetry index). We established orienting values for a group of normal speakers and demonstrated the usefulness of these measures for the analysis of a patient with glossectomy in whom we compared pre- and postoperative movement ranges and evaluated the effect of the defect reconstruction on the deformation and symmetry of the tongue.

Towards Dynamic Three-Dimensional Ultrasound

With the current generation of ultrasound machines, we are faced with the dilemma that we can either visualize a two-dimensional dynamic motion or a three-dimensional static volume. Obviously, our ultimate goal would be to visualize the motion of the tongue in 3D. This would be especially important for our research on glossectomy patients because the partial tongue resections and reconstructions rarely ever lead to a symmetrical outcome. In recent years, so-called 4D ultrasound machines have become commercially available. These machines are currently able to visualize up to 30 volumes per second. However, while this technology is used with great success in obstetrics and gynaecology, it is less suitable for the imaging of the tongue in speech. The reason for this is that the air in the oral cavity causes echo artefacts in the ultrasound scan that obscure the true surface of the tongue in the three-dimensional volume. Consequently, 4D scans acquired this way would necessitate extensive post-processing. Yang & Stone (2002) at the Vocal Tract

Figure 11

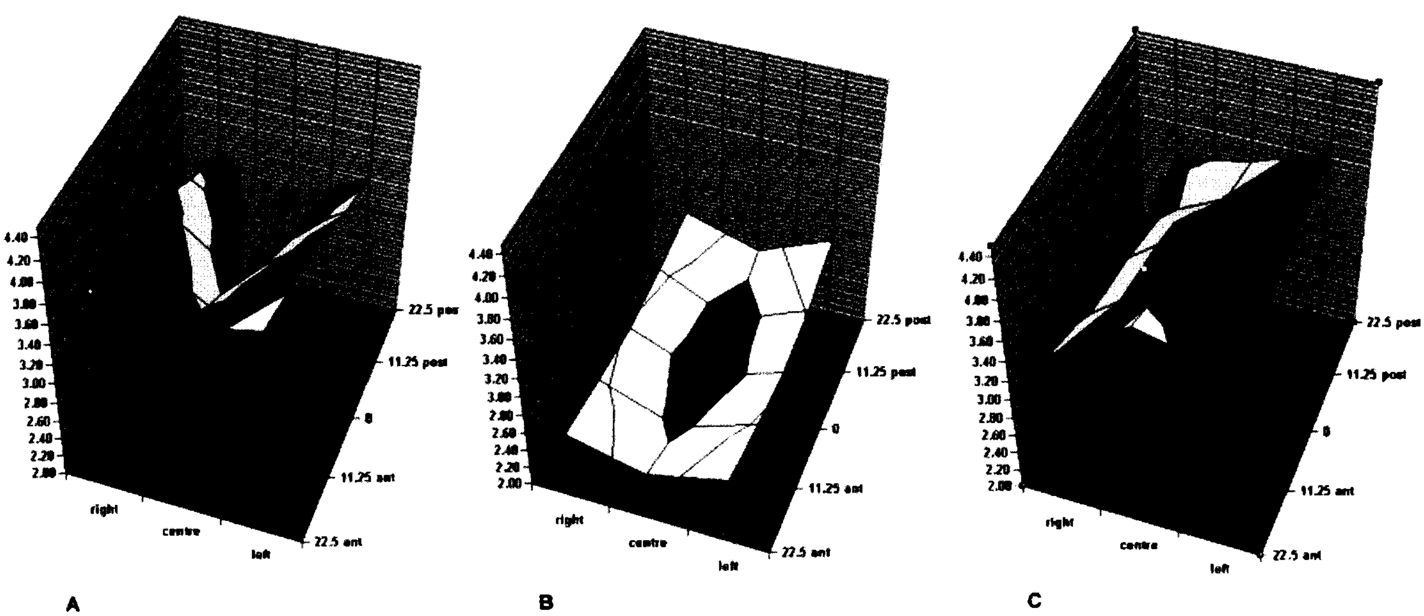


Figure 11. Surface plots for the postalveolar fricative /ʃ/. (a) Composite surface plot for 12 normal speakers; (b) A patient with a carcinoma of the right lateral tongue preoperatively; (c) The same patient after the tumour resection and reconstruction with a radial forearm flap. Note the decreased lingual grooving and the postoperative asymmetry of the tongue.

Visualization Laboratory at the University of Maryland recently suggested an interesting new approach to the reconstruction of three-dimensional tongue motion from multiple two-dimensional image sequences. The researchers recorded repeated utterances of the same sentence in multiple parallel sagittal and coronal planes and then reassembled the data in a reconstructed three-dimensional surface using the Dynamic Programming method. Using multiple two-dimensional scans to reconstruct a three-dimensional moving tongue surface is an elegant way to circumvent the current technological limitations of the available ultrasound machines. However, the high number of slices and repetitions required for the method described by Yang & Stone (2002) relies on the participation of a highly compliant research volunteer. It is probably less practical for clinical patient examinations and research in pathological speaker groups.

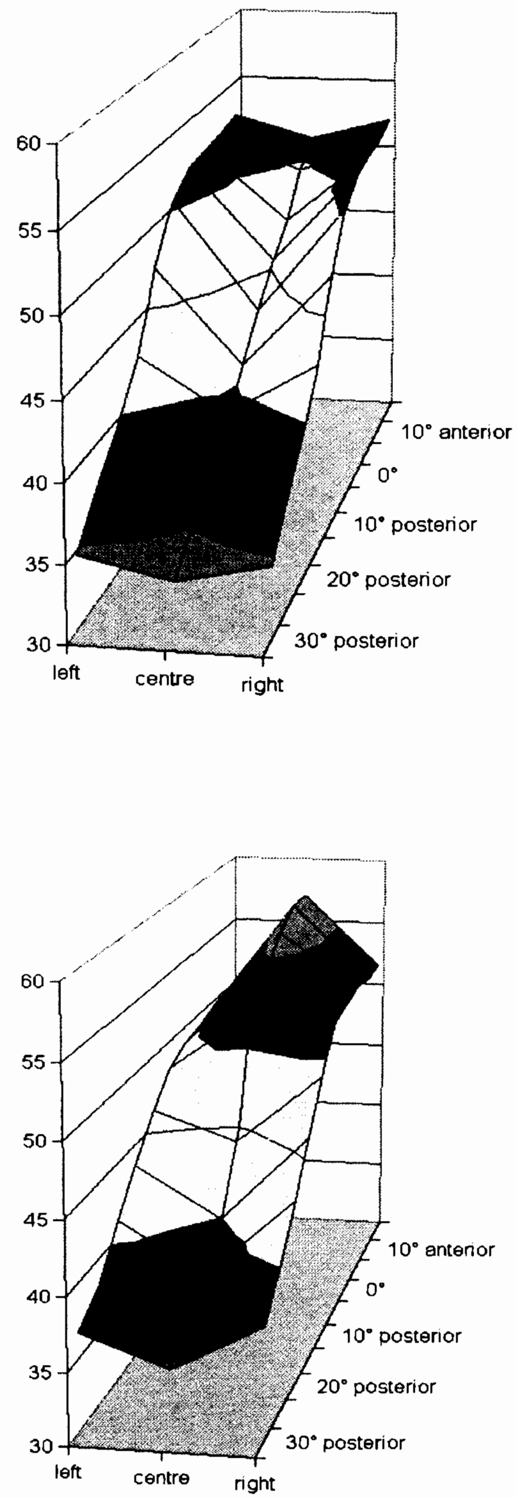
In a recent series of experiments at the Voice and Resonance Laboratory, we have started to acquire parallel sagittal scans to make sparse three-dimensional surface plots of the moving tongue. For this research, we use the CHASE device to stabilize the head of the research participant for repeated scans in three parallel sagittal planes. Instead of aligning the images retrospectively with Dynamic Programming, we pace the speech of the subject using a digital metronome. In order to facilitate the task of keeping a steady rhythm at 60 beats per minute, the metronome is set to 120 beats per minute and the subject is instructed to speak at half-tempo. So far, we have used only repeated VCV syllables for this examination technique. The stress is on the CV segment and coincides with the metronome beat (i.e., /a'ta/). The examination is repeated in three sagittal planes.

We then use a video-editing program with a parallel oscillogram display (ScreenBlast, Sony Corp., 550 Madison Avenue, New York, New York 10022) to identify the bursts of the metronome in the acoustic signal. The metronome bursts are used to identify key-frames that help us synchronize the image sequences for the same utterance recorded in three sagittal planes. The image sequences are analyzed using the Ultra-CATS software and the results are plotted as surfaces. Figure 12 shows a number of frames of the pseudo-3D surface plots of tongue movement during the production of the syllable /aka/. In our laboratory, this technique is still largely experimental at this time. While we have only used it on normal speakers to date, we are hoping to incorporate a similar procedure into future patient examinations.

Conclusion

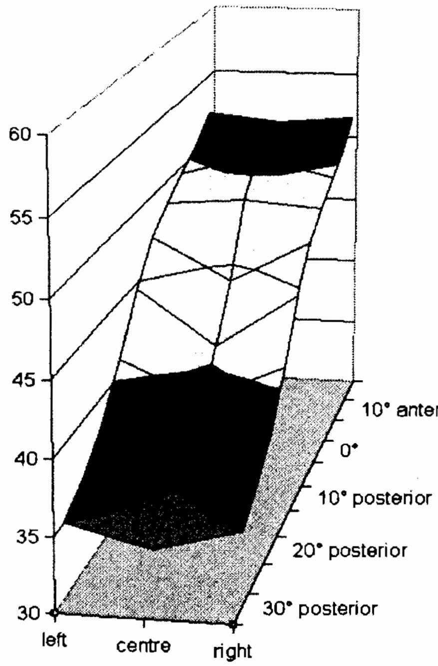
Ultrasound offers exciting new possibilities for researchers and therapists in speech-language pathology. The main advantages of ultrasound imaging are its non-invasiveness, bio-safety, cost-effectiveness and, last but not least, the ease of the image acquisition. Ultrasound allows us to acquire extensive amounts of speech data, and

Figure 12



a

b



c

Figure 12. Posterior view of the reconstructed three-dimensional surface plots of the tongue during the utterance /a'ka/. (a) First /a/: Note the prominent genioglossus furrow during the production of /a/; (b) Maximal dorsal elevation for the production of /k/; (c) Second /a/: Note the reduced genioglossus furrow following the velar closure.

the examination sessions can be as long as necessary. We also find that the direct visualization of the tongue shape on the ultrasound screen is a good motivator for research participants. Obviously, there are still a number of methodological issues associated with the use of ultrasound for research or clinical applications. These will have to be addressed in further research. Nevertheless, the benefits outweigh the disadvantages of this highly promising imaging method.

The costs for basic ultrasound machines have dropped significantly over the last few years, and it is likely that this trend will continue in the future. In addition, many manufacturers now make small portable ultrasound machines that are operated with batteries. In the future, many more speech-language pathologists may be able to use ultrasound for speech and swallowing assessments as well as a biofeedback device for therapy.

Acknowledgements

The authors gratefully acknowledge the invaluable contributions of the following people (in alphabetical order) to the research described here: Yarixa Barillas, Carmen Bollig, Arpita Bose, Amanda Braude, Kevin Cannons, Brent Carmichael, Michelle Ciccia, Heather Flowers, Jiayun (Jenny) Gu, Gajanan (Kiran) Kulkarni, Marilia Sampaio, Parveen Thind, Catherine Uy and Willy Wong. Funding for this research was provided by the Canadian Institutes for Health Research (grant fund # MOP 62960).

References

Akgul, Y.S., Kambhamettu, C., & Stone, M. (1999). Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging, 18,* 1035-1045.

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

Bosma, J.F., Hepburn, L.G., Josell, S.D., & Baker, K. (1990). Ultrasound demonstration of tongue motions during suckle feeding. *Developmental Medicine and Child Neurology, 32,* 223-229.

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

Bressmann, T., Thind, P., Uy, C. Bollig, C., Gilbert, R.W., & Irish, J.C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from twelve normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics, 19,* 573-588.

Bressmann, T., Gu J., Cannons, K., Wong, W., Heng, C. L., & Carmichael, B. (in preparation). Quantitative analysis of tongue motion using semi-automatic edge detection software in B-mode ultrasound films: Design, methods and procedure validation.

Carmichael, B. (2004). *The Comfortable Head Anchor for Sonographic Examinations (CHASE).* Toronto: University of Toronto.

Casas, M. J., Kenny, D. J., & Macmillan, R. E. (2003). Buccal and lingual activity during mastication and swallowing in typical adults. *Journal of Oral Rehabilitation, 30,* 9-16.

Casas, M.J., Kenny, D.J., & Mc Pherson, K.A. (1994). Swallowing/ventilation interactions during oral swallow in normal children and children with cerebral palsy. *Dysphagia, 9,* 40-46.

Casas, M.J., McPherson, K.A., & Kenny, D.J. (1995). Durational aspects of oral swallow in neurologically normal children and children with cerebral palsy: an ultrasound investigation. *Dysphagia, 10,* 155-159.

Cheng, C. F., Peng, C. L., Chiou, H. Y., & Tsai, C. Y. (2002). Dentofacial morphology and tongue function during swallowing. *American Journal of Orthodontics and Dentofacial Orthopedics, 122,* 491-499.

Chi-Fishman, G., & Sonies, B. C. (2002). Kinematic strategies for hyoid movement in rapid sequential swallowing. *Journal of Speech, Language, and Hearing Research, 45,* 457-468.

Chi-Fishman, G., Stone, M., & McCall, G. N. (1998). Lingual action in normal sequential swallowing. *Journal of Speech, Language, and Hearing Research, 41,* 771-785.

Davidson, L. (2004 April). *Assessing tongue shape similarity: comparing L2 norms and area measures.* Paper presented at the meeting of the Second Ultrasound Roundtable, Vancouver, BC.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association, 32,* 113-122.

Gick, B., & Rahemtulla, S. (2004 April). *Recent developments in quantitative analysis of ultrasound tongue data.* Paper presented at the meeting of the Second Ultrasound Roundtable, Vancouver, BC.

Gu, J., Bressmann, T., Cannons, K., & Wong, W. (2004). *The Ultrasonographic Contour Analyzer for Tongue Surfaces (Ultra-CATS).* Toronto: University of Toronto.

Hewlett, N., Vazquez, Y., Zharkova, A., & Zharkova, N. (2004 April). *Ultrasound study of coarticulation and the "Trough Effect" in symmetrical VCV syllables: A report of work in progress.* Paper presented at the meeting of the Second Ultrasound Roundtable, Vancouver, BC.

Hiiemae, K.M., & Palmer, J.B. (2003). Tongue movements in feeding and speech. *Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists, 14,* 413-29.

Kenny, D.J., Casas, M.J., & McPherson, K.A. (1989). Correlation of ultrasound imaging of oral swallow with ventilatory alterations in cerebral palsied and normal children: preliminary observations. *Dysphagia, 4,* 112-117.

Kikyo, T., Saito, M., & Ishikawa, M. (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.

Li, M., Kambhamettu, C, & Stone, M. (2003 August). EdgeTrak, a program for band edge extraction and its applications. Paper presented at the Sixth IASTED International Conference on Computers, Graphics and Imaging, Honolulu, HI.

Lundberg, A., & Stone, M. (1999). Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data. *Journal of the Acoustical Society of America, 106,* 2858-2867.

Morrish, K.A., Stone, M., Sonies, B.C., Kurtz, D., & Shawker, T. (1984). Characterization of tongue shape. *Ultrasonic Imaging, 6,* 37-47.

Munhall, K.G. (1985). An examination of intra-articulator relative timing. *Journal of the Acoustical Society of America, 78,* 1548-1553.

Neuschäfer-Rube, C., Wein, B.B., Angerstein, W., Klajman, S. Jr., & Fischer-Wein, G. (1997). Sektorbezogene Grauwertanalyse videosonographisch aufgezeichneter Zungenbewegungen beim Schlucken. *HNO, 45,* 556-562.

Parush, A., & Ostry, D.J. (1993). Lower pharyngeal wall coarticulation in VCV syllables. *Journal of the Acoustical Society of America, 94,* 715-722.

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.

Schliephake, H., Schmelzeisen, R., Schönweiler, R., Schneller, T., & Altenbernd. C. (1998). Speech, deglutition and life quality after intraoral tumour resection. A prospective study. *International Journal of Oral and Maxillofacial Surgery, 27*, 99-105.

Shawker, T.H., & Sonies, B.C. (1984). Tongue movement during speech: a real-time ultrasound evaluation. *Journal of Clinical Ultrasound, 12*, 125-133.

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

Shawker, T.H., Sonies, B., Stone, M., & Baum, B.J. (1983). Real-time ultrasound visualization of tongue movement during swallowing. *Journal of Clinical Ultrasound, 11*, 485-490.

Soder, N., & Miller, N. (2002). Using ultrasound to investigate intrapersonal variability in durational aspects of tongue movement during swallowing. *Dysphagia, 17*, 288-297.

Sonies, B.C., Baum, B.J., & Shawker, T.H. (1984). Tongue motion in elderly adults: initial in situ observations. *Journal of Gerontology, 39*, 279-283.

Sonies, B.C., & Dalakas, M.C. (1991). Dysphagia in patients with the post-polio syndrome. *New England Journal of Medicine, 324*, 1162-1167.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America, 87*, 2207-2217.

Stone, M., & Davis, E.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, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America, 99*, 3728-3737.

Stone, M., Morrish, K., Sonies, B.C., & Shawker, T.H. (1987). Tongue curvature: A model of shape during vowel production. *Folia Phoniatrica, 39*, 302-315.

Stone, M., & Shawker, T.H. (1986). An ultrasound examination of tongue movement during swallowing. *Dysphagia, 1*, 78-83.

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

Unser, M., & Stone, M. (1992). Automated detection of the tongue surface in sequences of ultrasound images. *Journal of the Acoustical Society of Canada, 91*, 3001-3007.

Van Riper, C. (1963). *Speech correction: principles and methods, 4th edition*. Englewood Cliffs, NJ: Prentice-Hall.

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*, 496-499.

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

Wein, B., Alzen, G., Tolxdorff, T., Bockler, R., Klajman, S., & Huber, W. (1988). Computersonographische Darstellung der Zungenmotilität mittels Pseudo-3D-Rekonstruktion. *Ultraschall in der Medizin, 9*, 95-97.

Wein, B., Bockler, R., Huber, W., Klajman, S., & Willmes, K. (1990). Computersonographische Darstellung von Zungenformen bei der Bildung der langen Vokale des Deutschen. *Ultraschall in der Medizin, 11*, 100-103.

Wein, B., Klajman, S., Huber, W., & Doring, W. H. (1988). Ultraschalluntersuchung von Koordinationsstörungen der Zungenbewegung beim Schlucken. *Nervenarzt, 59*, 154-158.

Whalen, D.H., Iskarous, K., Tiede, M.K., Ostry, D.J., Lehnert-LeHoullier, H., Vatikiotis-Bateson, E., & Hailey, D.S. (2004 April). *HOCUS: the Haskins Optically Corrected Ultrasound System*. Paper presented at the meeting of the Second Ultrasound Roundtable, Vancouver, BC.

Wrench, A. (2004 April). *QMUC Matching, merging and means: Spline productivity tools for ultrasound analysis*. Paper presented at the meeting of the Second Ultrasound Roundtable, Vancouver, BC.

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

Author Note

Please address correspondence to: Tim Bressmann, Ph.D., Assistant Professor, Graduate Department of Speech-Language Pathology, University of Toronto, 500 University Avenue, Toronto, ON M5G 1V7 Canada, tim.bressmann@utoronto.ca

Received: November 15, 2004
Accepted: March 16, 2005