

# The use of ultrasound for magnetic phonetic fieldwork

**Bryan Gick**

Department of Linguistics  
Interdisciplinary Speech Research Laboratory (ISRL)  
University of British Columbia, Vancouver  
& Haskins Laboratories, New Haven, CT  
[gick@interchange.ubc.ca](mailto:gick@interchange.ubc.ca)

An increasingly wide variety of imaging and tracking technologies have been applied to measuring speech articulation in recent years. However, virtually all of these techniques have been restricted to laboratory or clinical settings. Portable ultrasound technology offers a way to conduct easy, non-invasive dynamic imaging of the whole tongue for less commonly studied languages, both in the field and in the laboratory.

## 1 Introduction

Ultrasound imaging technology has been used in studies of speech production since it came into regular clinical use in the 1960s and the 1970s. Ultrasound is able to capture dynamic tongue shape, enabling the study of such elusive lingual measures as the tongue root, sagittal groove, and interactions between vowels and lingual consonants. Early ultrasound studies of speech (Kelsey et al. 1969, Skolnick et al. 1975, Zagzebski 1975, MacKay 1977 and others) used relatively large and cumbersome hospital equipment to produce 1-dimensional (A-mode) measurements (recording movement along a single line), usually in the pharyngeal region. Even in this much more limited 1-D form, the advantages of ultrasound were clear, and ultrasound still remains the only available option for safe, non-invasive imaging of real movements of the whole tongue.

Since the 1970s, of course, advances in technology have made ultrasound an increasingly useful tool for speech research. Not only have 2-dimensional (B-mode), and even 3-dimensional techniques been developed (though the temporal resolution of the latter is as yet too low for measuring most natural speech), but also advancements in image processing have dramatically improved the quality and accuracy of the images. Because of these improvements, a number of researchers have used ultrasound to measure tongue shapes and movements with increasing effectiveness since the 1980s (Sonies et al. 1981; Keller & Ostry 1983; Munhall & Ostry 1985; Shawker et al. 1985; Stone, Morish et al. 1987; Stone, Shawker et al. 1988; Stone 1990; Stone, Faber et al. 1992; Stone & Lundberg 1996; Lundberg & Stone 1999). These studies, however, have typically been laboratory-based, still using cumbersome hospital ultrasound machines, and often employing involved means of maintaining or tracking head and transducer positions. This is of course necessary if one is to use ultrasound to make



**Figure 1** An example of a portable ultrasound recording package, including an Aloka SD-900 ultrasound machine, intercostal probe with laser pointer attachment, ultrasound gel, digital video camera, and microphone.

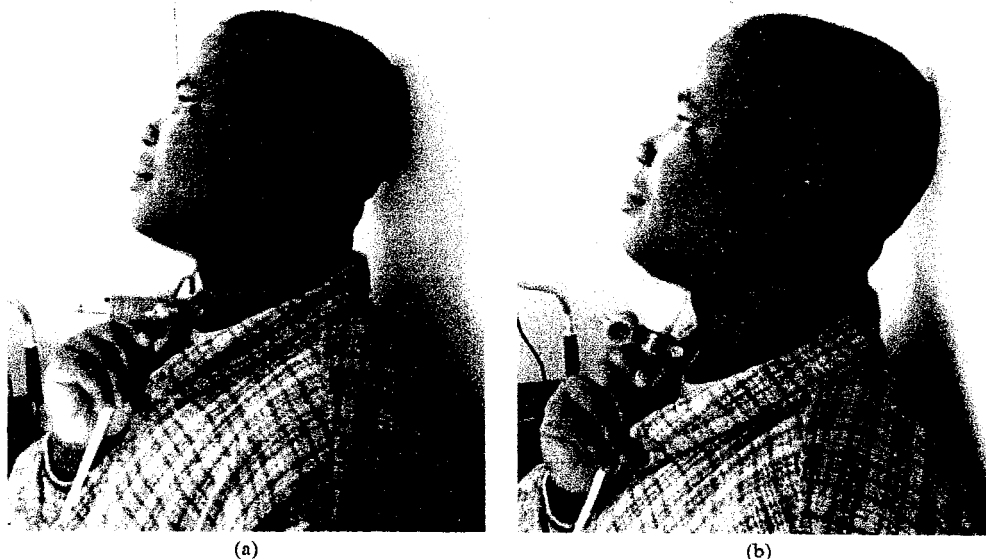
quantitatively verifiable measurements of absolute tongue location. But it is in cases where such a level of accuracy can be compromised to a degree, or where qualitative elements such as tongue shape are being measured – and both of these are certainly true of most linguistic phonetic field description – that the advantages of ultrasound make themselves most evident.

More recently, the increased availability and lower price of portable and PC-based ultrasound units, digital video recording equipment, and image analysis software have brought ultrasound within financial reach of many linguistic phonetics labs. Used ultrasound units may also be available. Recent work in the Interdisciplinary Speech Research Laboratory (ISRL) at the University of British Columbia has focused on developing the potential of using 2-D ultrasound in these field situations. Because of its portability and its ability to image the harder-to-access parts of the vocal tract non-invasively, particular focus has been placed on using ultrasound to address long-standing descriptive questions in less studied languages. Especially useful have been those results pertaining to tongue root activity, timing of articulatory events, overall tongue shape, and non-sagittal lingual measures such as lateral release and tongue grooving.

## 2 Description of methods

There already exist many references describing ultrasound technology in general (see e.g. Stone 1997). The present paper, following a brief discussion of the general technique, will focus on applying ultrasound directly to phonetic fieldwork.

Our studies at ISRL to date have used an Aloka SSD-900 portable ultrasound machine weighing 14.5 kg, with a small 3.5 MHz electronic convex intercostal probe (Aloka UST-9102; see figure 1). Any small probe with a sharp (60-90-degree) field of view should work reasonably well. There is an increasing variety of options available for even more compact portable ultrasound, including both PC-based and packaged systems, some of which are

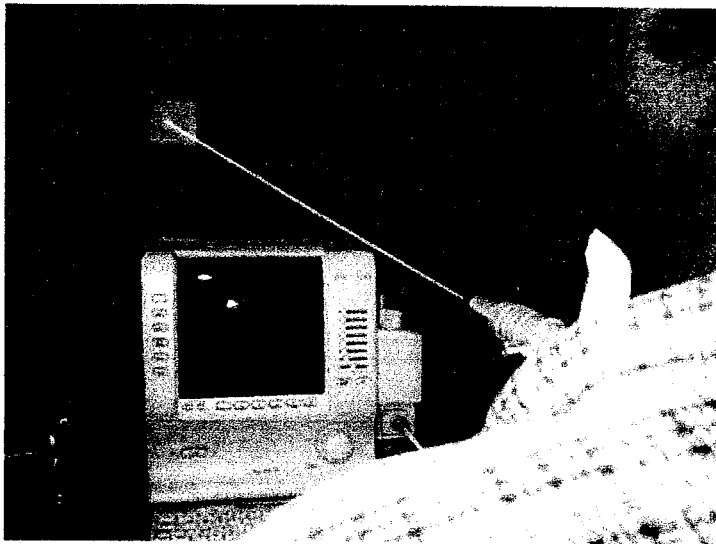


**Figure 2** Placement of transducer for (a) midsagittal and (b) coronal sections.

battery powered (such as SonoSite, [www.sonosite.com](http://www.sonosite.com)), which is about the size and weight of a laptop computer). A number of similar products may be located with a web search.

Ultrasound images are reconstructed from the echo patterns of ultra-high-frequency sound both emitted and received by piezoelectric crystals contained in a small hand-held transducer, or 'probe'. Ultrasound allows for imaging of soft tissue, but not of bone or air. Thus, when held against the soft tissue below the chin and aimed up and back toward the tongue surface (figure 2a), 2-dimensional cross-sections of the superior surface of the tongue can be imaged from root to blade (except for 'shadows' cast by the hyoid bone, the mandible, and the sublingual cavity). Likewise, by turning the transducer 90 degrees, any coronal or transverse section of the tongue may be imaged (figure 2b). Experimenting with transducer placement will clarify the best position for different measures and subjects. Examples of images thus collected are shown in figures 4 and 5 below.

While many methods have been used by previous researchers to control transducer position, in our fieldwork, the transducer is held by hand, usually by the speaker him- or herself. Probe position is maintained using a small laser pointer attached to the probe. A speaker is seated in a solid chair, preferably with the head leaning against a headrest or wall, facing a vertical surface at a distance of approximately 2 meters. The laser pointer projects an image of crosshairs into a 10 cm square target area attached to the surface. Speakers are instructed to keep the crosshairs upright and within this target area during data collection. Depending on the level of accuracy needed, an additional laser pointer may be attached to the speaker's head to control for head movement; however, this is normally too much for a typical field situation. After a few minutes of practice, most speakers are easily able to do this while speaking and/or reading from a list of stimuli. Accuracy of the laser pointer is monitored visually by the investigators during the experiment, and tokens



**Figure 3** Projection of laser pointer crosshairs onto target (the white line has been added to show the trajectory of the beam).

where the crosshairs leave the target area are re-collected. This arrangement is shown in figure 3.

Images are recorded to a standard digital video camera and thus have the frame rate of standard video (about 30 frames/sec). This frame rate has shown itself to be adequate for capturing most articulatory movements. The acoustic signal is simultaneously recorded to VHS or digital video using an external microphone to ensure synchronization with the video signal. After collection, videos may be digitized or transferred to a computer for analysis using video editing software such as Adobe Premiere, Final Cut Pro, or other, cheaper alternatives. Images may then be edited and analyzed using graphics software such as the public domain NIH Image program (developed at the US National Institutes of Health, available at <http://rsb.info.nih.gov/nih-image>), or its cross-platform alternative ImageJ. (These are also available from [www.scioncorp.com](http://www.scioncorp.com).) Specialized ultrasound analysis software has been developed at the ISRL, Haskins Laboratories, and other labs where ultrasound has been used (see e.g. Stone 1990, Stone et al. 1992, Stone & Lundberg 1996, Iskarous et al. 2001).

All of the equipment described above will run off of a single standard grounded outlet. The only component that requires an outlet is the ultrasound machine itself, and if necessary this can be powered by a generator that supplies standard voltage. Alternatively, if no external power source is available, a smaller system running on batteries may be more appropriate.

The obvious limitations of this method are the difficulty in capturing the tongue tip and epiglottis (because of the 'shadows' created by the sublingual and epiglottic cavities, respectively), difficulty in controlling for the position of the transducer relative to the head, and the fact that no stable reference points or opposing surfaces (e.g. the palate and the rear/lateral pharyngeal walls) are measured. However, studies have shown these problems to be less dramatic than might be expected. One study (Stone 1990) found that for vowels, /s/ and /l/, only about 1 cm of the tip is lost (comparable to X-ray Microbeam and EMMA tracking systems, which typically position the first pellet 1 cm from the tip). Another study (Magen & Kang 2001) shows the stability of the rear pharyngeal wall during the full range of

English vowels, suggesting that tongue root measurements alone are sufficient to approximate pharyngeal constrictions for English vowels. To further mitigate these limitations, additional transducer pressure can improve tongue tip imaging, and, where more precise measurements are required, tracking devices can be used to correct for position. In addition, it is possible to image the palate by filling the mouth with water, but reliable synchronizing of palate and tongue images requires head tracking, which is assumed to be beyond the scope of most field phonetics applications. In any case, the limitations of ultrasound are dramatically outweighed by its being the only safe, uninvasive, and portable technique available for tongue imaging. An additional advantage of ultrasound that is not to be underestimated for field research is that, unlike many less tangible techniques, the tongue images (that can be viewed in real time by the speaker) are generally of both immediate interest and obvious relevance to most speakers. This has resulted in a high level of enthusiasm and support from almost every speaker to date.

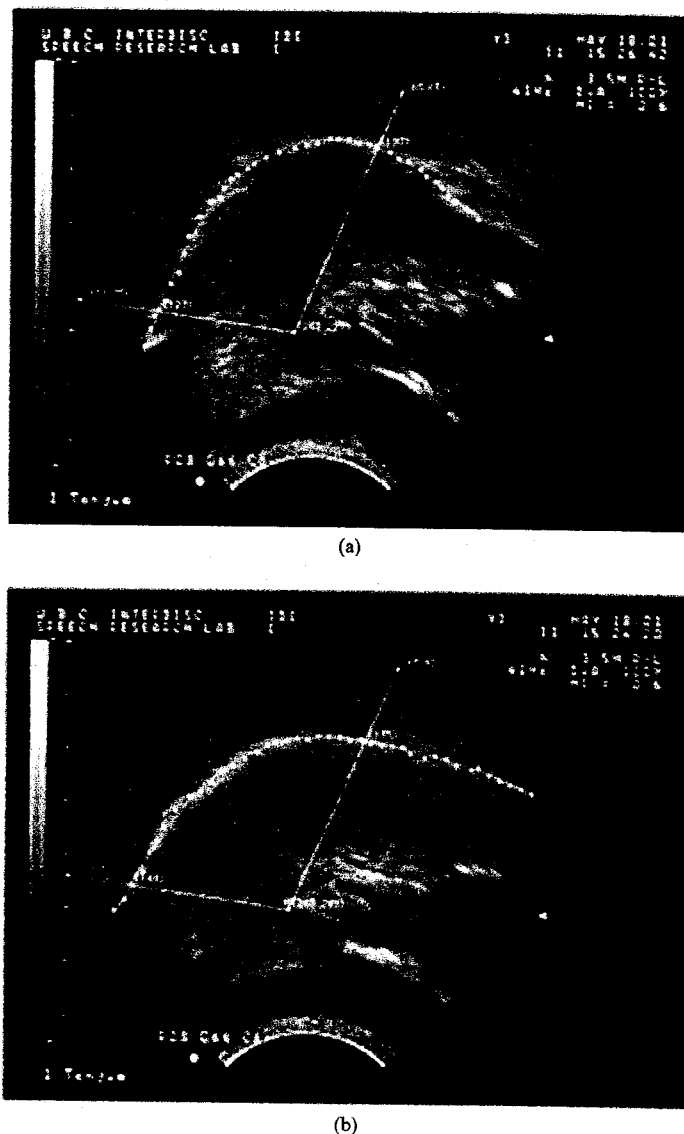
### 3 Sample results

In a number of ongoing studies, ultrasound imaging is being applied to clarify issues of interest to linguistic phonetics and phonology. Some of the areas being studied include the timing of articulatory events (Gick & Wilson 2001, Gick et al. 2001), overall tongue shape (Oh 2002), and non-sagittal measures (e.g. lateral release and tongue grooving). Studies currently in progress focus particularly on various African, Asian, and Indo-European languages, and languages of the northwest coast of North America.

In the remainder of this section, illustrations will be given from two ongoing studies focusing on the tongue root. The first of these concerns tongue root harmony in Kinande, and the second pharyngealization and vowel lowering in Nuuchah-nulth.

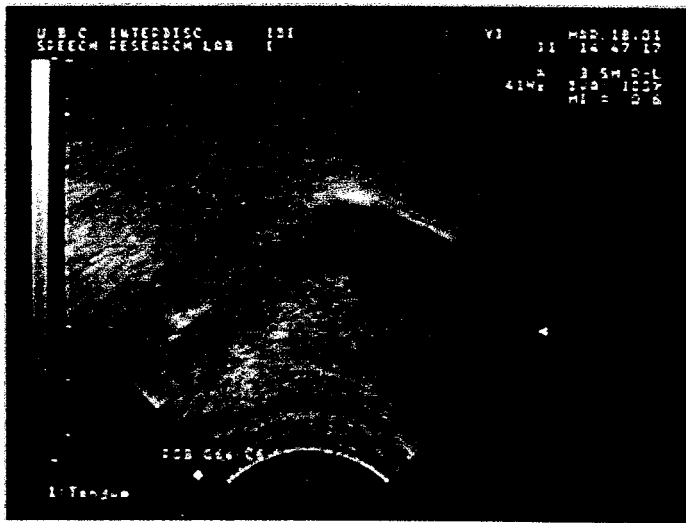
Kinande, a Bantu language of Congo, has been described as having [ATR] harmony (see Archangeli & Pulleyblank 1994 for references). However, a number of questions have persisted around this and other supposed ATR systems, including tongue root interactions with low vowels, interactions with laryngeal configurations, and even whether the relevant characteristic distinguishing these vowel systems is, in fact, advancement vs. retraction of the tongue root. Ongoing ultrasound studies have investigated these issues in ATR languages such as Kinande and Yoruba. Whalen & Gick (2001), for example, showed that both tongue root advancement/retraction and tongue body height interact with intrinsic fundamental frequency effects. Further study of additional languages will help to clarify whether this is a passive or active interaction, and whether this is a universal effect. Results of these studies will contribute to our understanding of the phonological status of the ATR feature, its phonetic realization, and its relation to other phonological and phonetic processes. Figure 4 shows examples of the advanced and retracted variants of the vowel /e/. These images show an interaction between tongue root position (more horizontal line) and tongue body height (more vertical line).

Another ongoing study concerns post-velar consonants in Nuuchah-nulth, a Wakashan language spoken on the west coast of Vancouver Island. In this language, as with many languages of the nearby Salish family, upper and lower pharyngeal consonant constrictions have been described as affecting vowel height, with the strongest apparent effect on vowels following the consonant in question, but with some effect on preceding vowels. One of the many questions to be asked about the post-velar consonants in this language is whether this relationship between tongue retraction and vowel height is categorical/phonological or gradient/phonetic. If the lowering is phonetic, then the degree of tongue body lowering should be simultaneous with and directly relatable to the degree of tongue root retraction. Ultrasound data was collected from two female speakers, both approximately 60 years old, of the Ahousaht dialect of Nuuchah-nulth. Each session resulted in approximately 30 minutes

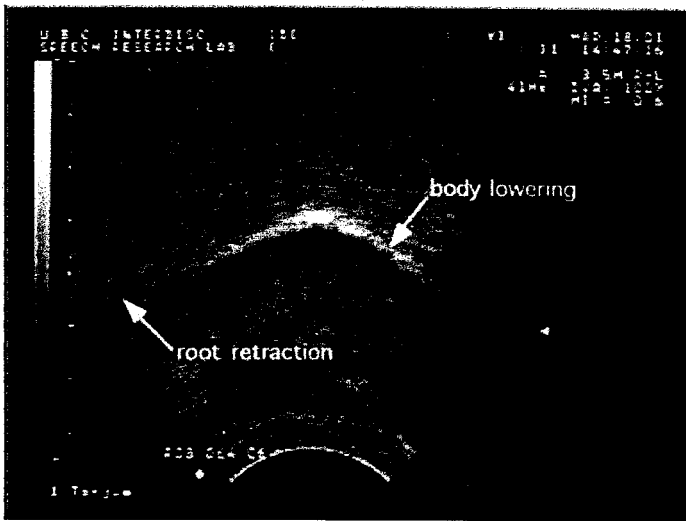


**Figure 4** Midsagittal ultrasound images of Kinande, showing (a) ATR and (b) RTR varieties of the vowel /e/. The tongue tip is at the right of the picture, and the root is at the left (note that the images are tilted at the same angle at which the probe was held during data collection; see transducer angle in figure 2a). The tongue surface can be seen as the lower edge of the white curved region.

of recorded data, with occasional breaks between sets. Although the data collection process was straightforward, both speakers commented that it was tiring holding the ultrasound probe up to their necks for extended periods. This kind of problem can be solved by attaching the probe to a microphone stand or the like. An example of the data is shown in figure 5.



(a)



(b)

**Figure 5** Midsagittal ultrasound images of Nuu-chah-nulth, showing the vowel /i/ (a) in a non-pharyngeal context, and (b) following a pharyngeal glide. The tongue tip is at the right of the picture, and the root is at the left (note that the images are tilted at the same angle at which the probe was held during data collection; see transducer angle in figure 2a). Observe the tongue root retraction and lower tongue body position following the pharyngeal (note the one-centimeter hash marks along the left and lower edges of the image).

## 4 Conclusion

Portable ultrasound is proving to be an extremely useful tool for field linguistic studies, and at the current rate of technological development, it promises only to become more desirable and affordable as time passes. In addition, with the ongoing reliance of medical fields on ultrasound imaging, this technology will always be available, along with a level of technical

support unfamiliar to linguistic phoneticians using traditional methods. In combination with acoustic measurements and analysis, ultrasound can provide a type and quality of articulatory imaging data previously unavailable to the field linguist.

## Acknowledgements

The author wishes to acknowledge Shaffiq Rahemtulla, Barbara Bernhardt, Fiona Campbell, Sunyoung Oh, and Ian Wilson for contributions to the development and testing of the UBC ISRL ultrasound system; and Oladiipo Ajiboye, Fiona Campbell, Ngessimo Mutaka, Douglas Pulleyblank, and Ian Wilson for contributions to the specific language sections of this paper. Additional helpful discussion and comments have come from John Esling, Khalil Iskarous, Maureen Stone, Doug Whalen, and several audiences, including in particular the December 2001 meeting of the Acoustical Society of America. Financial support for this project has been provided from a New Opportunities grant from the Canadian Foundation for Innovation, a research grant to the author from the Natural Sciences and Engineering Research Council of Canada, and NIH Grant DC-02717 to Haskins Laboratories.

## References

- ARCHANGELI, D. & PULLEYBLANK, D. (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.
- GICK, B., CAMPBELL, F. & OH, S. (2001). A cross-linguistic study of articulatory timing in liquids. *Journal of the Acoustical Society of America* **110**(5), Pt. 2, 2656.
- GICK, B. & WILSON, I. (2001). Pre-liquid excrescent schwa: What happens when vocalic targets conflict. In Dalsgaard, P., Lindberg, B. & Benner, H. (eds.), *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 273–276. Aalborg, Denmark: Center for Personkommunikation (CPK).
- ISKAROUS, K., WHALEN, D. H. & MATTINGLY, I. (2001). Modeling tongue shapes with conic arcs. *Journal of the Acoustical Society of America* **110**(5), Pt. 2, 2760.
- KELLER, E. & OSTRY, D. (1983). Computer measurement of tongue dorsum movement with pulsed echo ultrasound. *Journal of the Acoustical Society of America* **73**, 1309–1315.
- KELSEY, C. A., WOODHOUSE, R. J. & MINIFIE, F. D. (1969). Ultrasonic observations of coarticulation in the pharynx. *Journal of the Acoustical Society of America* **46**(4), 1016–1018.
- 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.
- MACKAY, I. R. A. (1977). Tenseness in vowels: An ultrasonic study. *Phonetica* **34**, 325–351.
- MAGEN, H. & KANG, A. M. (2001). Rear pharyngeal wall placement during vowel production. *Journal of the Acoustical Society of America* **110**(5), Pt. 2, 2761.
- MUNHALL, K. G. & OSTRY, D. J. (1985). Ultrasonic measurement of laryngeal kinematics. In Titze, I. R. & Scherer, R. C. (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, 145–162. Denver, CO: Denver Center for the Performing Arts.
- OH, S. (2002). Cross-language blending of /l/ gestures by bilingual Korean-English children. In Do, A., Fish, S. & Skarabela, B. (eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development*, 473–484. Somerville, MA: Cascadia Press.
- SHAWKER, T. H., STONE, M. & SONIES, B. C. (1985). Tongue pellet tracking by ultrasound: Development of a reverberation pellet. *Journal of Phonetics* **13**, 135–146.
- SKOLNICK, M. L., ZAGZABSKI, J. A. & WATKIN, K. L. (1975). Two-dimensional ultrasonic demonstration of lateral pharyngeal wall movement in real time – a preliminary report. *Cleft Palate Journal* **12**, 299–303.
- SONIES, B. C., SHAWKER, T. H., HALL, T. E., GERBER, L. H. & LEIGHTON, S. B. (1981). Ultrasonic visualization of tongue motion during speech. *Journal of the Acoustical Society of America* **70**, 683–686.
- 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. (1997). Laboratory techniques for investigating speech articulation. In Hardcastle, W. J. & Laver, J. (eds.), *Handbook of Phonetic Sciences*, 11–32. Blackwell Handbooks in Linguistics Series. Oxford: Blackwell.
- STONE, M., FABER, A., RAPHAEL, L. J. & SHAWKER, T. H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics* 20, 253–270.
- 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., MORISH, 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., TALBOT, T. & RICH, A. (1988). Cross-sectional tongue shape during vowel production. *Journal of the Acoustical Society of America* 83, 1586–1596.
- WHALEN, D. H. & GICK, B. (2001). Intrinsic F0 and pharyngeal width in ATR languages. *Journal of the Acoustical Society of America* 110(5), Pt. 2, 2761.
- ZAGZEBSKI, J. A. (1975). Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research* 18, 308–318.