Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance[a)]

Lisa Davidson[b)]
Department of Linguistics, New York University, 719 Broadway, 4th Floor, New York, New York 10003

(Received 31 October 2005; revised 24 April 2006; accepted 25 April 2006)

Ultrasound imaging of the tongue is increasingly common in speech production research. However, there has been little standardization regarding the quantification and statistical analysis of ultrasound data. In linguistic studies, researchers may want to determine whether the tongue shape for an articulation under two different conditions (e.g., consonants in word-final versus word-medial position) is the same or different. This paper demonstrates how the smoothing spline ANOVA (SS ANOVA) can be applied to the comparison of tongue curves [Gu, *Smoothing Spline ANOVA Models* (Springer, New York, 2002)]. The SS ANOVA is a technique for determining whether or not there are significant differences between the smoothing splines that are the best fits for two data sets being compared. If the interaction term of the SS ANOVA model is statistically significant, then the groups have different shapes. Since the interaction may be significant even if only a small section of the curves are different (i.e., the tongue root is the same, but the tip of one group is raised), Bayesian confidence intervals are used to determine which sections of the curves are statistically different. SS ANOVAs are illustrated with some data comparing obstruents produced in word-final and word-medial coda position. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2205133]

PACS number(s): 43.70.Jt [AL] Pages: 407–415

I. INTRODUCTION

Ultrasound imaging is becoming an increasingly popular technique for examining articulation in speech research. Previous research has shown that ultrasound imaging is a practical, low-cost, and noninvasive tool for acquiring articulatory data to examine tongue shapes corresponding to various sounds, answering phonological questions, conducting phonetic fieldwork, and use in speech rehabilitation (e.g., Bernhardt *et al.*, 2003; Bressmann *et al.*, 2005; Davidson, 2005; Gick, 2002; Stone, 2005; Stone *et al.*, 1992; Stone and Lundberg, 1996).

Ultrasound is an attractive technique for imaging articulation during speech because it provides an image of the length of the tongue. Other techniques for imaging the midsagittal contour of the length of tongue such as MRI and cinefluorography are also available. However, these methodologies are often prohibitively expensive or difficult to access. In most speech-related applications of ultrasound, researchers have focused on collecting data from the midsagittal contour of the tongue, although coronal slices have also been analyzed (Slud *et al.*, 2002). A sample image of a midsagittal tongue curve during the production of the fricative /z/ is shown in Fig. 1. In this and following ultrasound images, the tongue tip is on the right and the tongue root is on the left. The ability to image the entire contour of the tongue is a significant advantage of ultrasound over techniques like electromagnetic midsagittal articulography (EMMA) (Perkell *et al.*, 1992) or x-ray microbeam (Westbury, 1994), which only allow for the tracking of the flesh points to which the receivers are attached. Though a tongue surface can be partially reconstructed from fleshpoint data, there are two main shortcomings for fleshpoint tracking as compared to imaging techniques like ultrasound: (1) since the placement of receivers is limited by the gag reflex, it is difficult or impossible to acquire information about the shape or motion of the tongue root, and (2) there is always the possibility that an important shape of the tongue occurs between two receivers and cannot be accurately reconstructed.

While ultrasound has become important as a tool for both linguistic and clinical investigation, there has not been consensus regarding the quantification and statistical analysis of the data that are collected. Some methods that have been used so far include the overlay of a concentric grid with equally spaced radial lines on the tongue shape, which allows for measurements from a fixed point to the tongue surface on any of the lines in the grid (Bressmann *et al.*, 2005); a mean distance measure that averages the Euclidean distances between corresponding points on two curves being compared (Davidson, 2005); and principal components analysis (Slud *et al.*, 2002). Of these methods, the most common measurement technique for midsagittal tongue curves has been the concentric grid, which is implemented in several software packages for ultrasound imaging processing [e.g., University of Arizona's GLoSsatron (http://dingo.sbs.arizona.edu/∼apilab/), Queen Margaret University College's Articulate Assistant (http://www.articulateinstruments.com/), the University of British Columbia's Ultrax (http://www.linguistics.ubc.ca/isrl/index.html), University of Toronto's Ultra-CATs (http://www.slp.utoronto.ca/English/Ultra-CATS.html); all websites last viewed on April 21, 2006]. The image in Fig. 2 demon-

[a)]Portions of this work were presented at the Ultrafest III workshop at the University of Arizona, April 2005 and the 50th Acoustical Society of America meeting in Minneapolis, MN, October 2005.

[b)]Electronic mail: lisa.davidson@nyu.edu

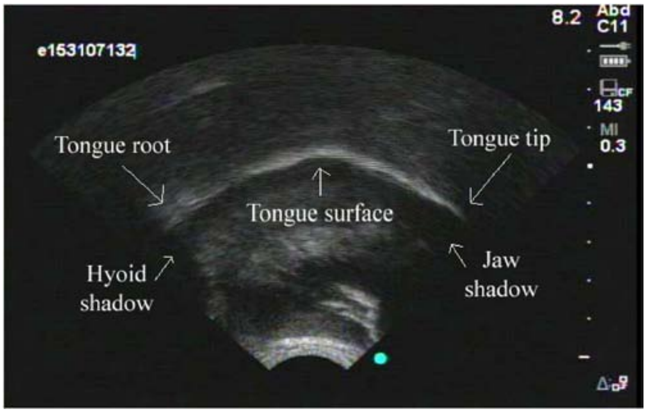


FIG. 1. (Color online) Midsagittal image of the frame corresponding to the midpoint of frication of the /z/ in the acoustic signal for "jazz dancer." The tongue tip is on the right and the tongue root is on the left.

strates a tongue curve representing the maximum constriction for the articulation of a /g/. Overlaid on top of the tongue curve is a grid with seven equally spaced radii. The origin of the radii is approximately at the center of the transducer. The ellipses indicate the points at which the radii intersect the tongue curve.

Using a radial grid, researchers can measure from the origin of the radii to the point at which each radius intersects the curve. For example, if a researcher were examining the difference in constriction degree for a velar stop like /g/ versus a velar fricative like /x/, measurements along one or more radii could be compared for multiple repetitions of the /g/ to those for the /x/, and then statistically analyzed with a *t* test. However, unless information from the entire tongue is recorded, it would be easy to miss taking a measurement at the most important location. For example, in the case of the /g/ in Fig. 2, the apex of the curve, marked with an X, is taken to be the point of maximum constriction. Since this point does not fall on a radius, the most relevant measurement is missed. Alternatively, the number of radii on the grid could be increased in order to make as many measurements as possible, but such a decision is an incomplete attempt to characterize the entire tongue surface, which is what the smoothing spline ANOVA described in this paper was explicitly developed to do. The other measurement techniques mentioned earlier are similarly dissatisfactory: either they are not suitable for individual comparisons (principal components), or there is no principled way of comparing individual sections of the tongue curve to determine where a difference lies (mean distance measures).

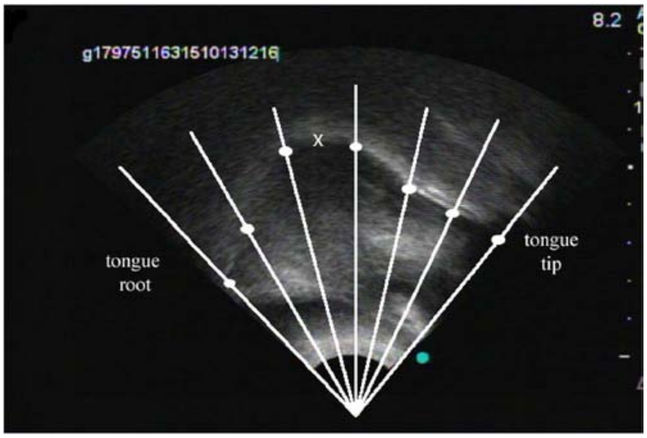


FIG. 2. (Color online) Midsagittal ultrasound image of the maximally raised position of the tongue dorsum for /g/ in "Baghdad" with a radial grid overlay. The white ellipses indicate where the radii intersect the tongue curve. The "X" indicates the location of maximum constriction along the tongue curve.

To address these issues, this paper introduces the smoothing spline ANOVA [SS ANOVA (Gu, 2002)] as a method for comparing tongue curve shapes. The SS ANOVA is a statistical method that allows for the holistic comparison of the entire tongue curve, whether it is obtained from ultrasound, MRI, or cinefluorography. This procedure has been used in other fields where similarities and differences of curve shapes must be assessed, such as plots of circadian rhythms in normal adults, patients with Cushing's syndrome, and patients with depression (Wang *et al.*, 2003). Because the mathematical details of both smoothing splines and the SS ANOVA have been well covered in both statistical and applied literature, this paper is intended primarily as a descriptive introduction of the technique for linguists or speech scientists who use ultrasound [or similar techniques, such as x-ray (e.g., Iskarous, 2005)] for speech research. References are provided for those desiring a more technical explanation of the procedures described in this paper.

To demonstrate the smoothing spline ANOVA for tongue curve comparison, it is illustrated with respect to the degree and location of maximum constriction of consonants in different word positions (e.g., *ba**g** dazzled* versus *Ba**g**hdad*). The data used in this paper to present the SS ANOVA come from an unpublished experiment, but this paper is not a report of the results of that study.

II. ULTRASOUND DATA COLLECTION

A. Data collection procedure

The stimuli consisted of three pairs of words and phrases containing the same consonant in different positions: *blac**k** top* versus *blac**k**top*, *ba**g** dazzled* versus *Ba**g**hdad*, and *ja**zz** dancer* versus *NA**S**DAQ*. These consonants were chosen because they are all lingual articulations that are easily imaged by the ultrasound. These words were produced by five monolingual native speakers of American English.

Midsagittal images of the tongue were recorded from a Sonosite Titan portable ultrasound machine using a 5–8 MHz Sonosite C-11 transducer with a 90° field of view and a depth of 8.2 cm. The incoming video signal from the ultrasound machine and an audio signal from an Audio Technica AT-813 microphone were synchronized and captured directly to a Dell computer using a Canopus ADVC-1394 capture card and Adobe Premiere 6.0. The Canopus card is designed to assure audio-video synchrony throughout the duration of the recording. The video frame rate is 29.97 Hz.

In order to compare images from different utterances, it is important to ensure that neither the speaker's head nor the transducer move during the experiment (Stone, 2005; Stone and Davis, 1995). Participants were seated in a sound-proof booth and their heads were stabilized using a moldable head

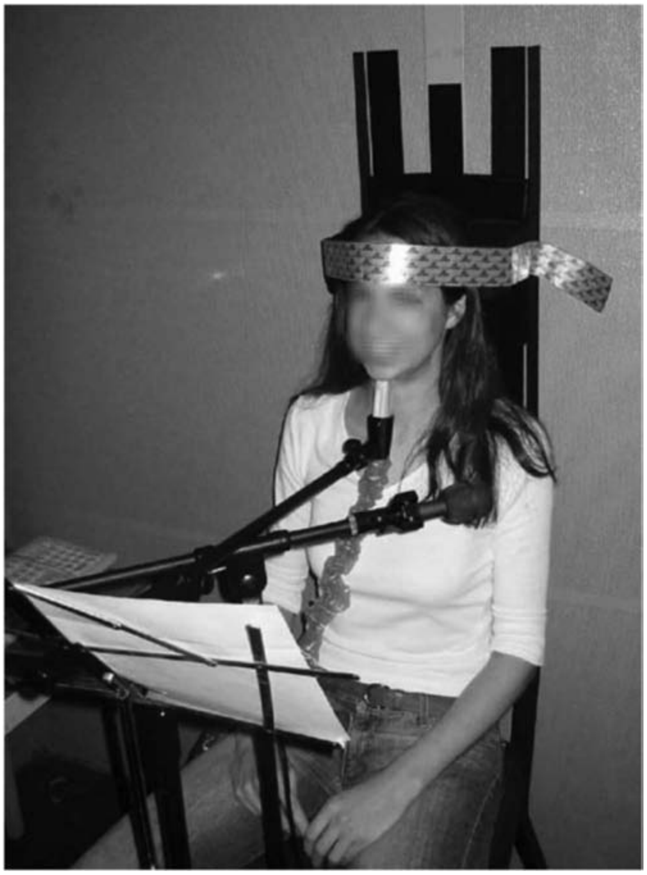


FIG. 3. Head and transducer stabilization setup. The speaker's head is encompassed by the moldable head stabilizer, which can be moved up and down on the Velcro strips against the wall of the soundproof booth. Another Velcro strap is pulled against the speaker's head for further stabilization. The transducer is stabilized with a microphone stand.

stabilizer (Comfort Company). The moldable head stabilizer is a rigid U-shaped foam brace designed to assist elderly people with low neck tone who have difficulty keeping their heads upright. The stabilizer is affixed to a wall in the soundproof booth with Velcro and is placed at the height of the participant's temples. Another piece of Velcro is then used to strap the speaker into the head stabilizer so that the head is entirely enclosed. Once the speaker is placed in the head stabilizer, one microphone stand to hold the transducer and another stand to hold the microphone are set up. The stand with the transducer is placed underneath the chin and the placement is adjusted until a satisfactory midsagittal tongue image is obtained. A picture of this setup is shown in Fig. 3.

A music stand was placed directly in front of the speaker at eye level. Eight pieces of paper each containing a randomization of the stimuli and fillers were placed on the music stand. The participant read the list, and then the experimenter turned the page. This resulted in eight repetitions of each phrase.

B. Edge extraction

After data collection, sections of the video files collected with Adobe Premiere containing the target phrases were transformed into JPEG stills. For the stops /k/ and /g/, the ultrasound frame with the most raised tongue body within the period of stop closure was chosen for comparison of tongue shapes for word-final versus word-medial codas. For the fricative /z/, the ultrasound frame roughly corresponding to the midpoint of the duration of frication on the acoustic record was chosen as the comparison frame. A sample image for the most raised tongue body for the /g/ in "Baghdad" is shown in Fig. 2 and the midpoint of the fricative for /z/ in "jazz dancer" is illustrated in Fig. 1. The decision to compare single frames as opposed to a sequence of frames was carried out both for theoretical reasons and for simplicity of presentation. First, one question that speech scientists may ask is whether the point of maximum constriction of a consonant differs with respect to some variable, such as word position, speech rate, or phonological environment (e.g., Browman and Goldstein, 1995; Kochetov, 2006). Second, in order to illustrate the SS ANOVA, the point of maximum constriction is used as a simple test case. However, the SS ANOVA has also been used to investigate comparisons along spatial and temporal dimensions, as illustrated by statistical methods developed to examine changes in the electroencephalograms (EEG) of epileptic patients (Guo *et al.*, 2003) or spatiotemporal changes in surface air temperature (Luo *et al.*, 1998).

For each repetition of the target phrases, the JPEG stills were loaded into EdgeTrak (version 1.0.0.4) for measurement (Li *et al.*, 2005). EdgeTrak is a computer program that automates the tracking of tongue contours by extracting (x,y) coordinates from the lower edge of the white curve in the ultrasound image. First, a few points on the tongue image are manually chosen, and then EdgeTrak uses an active contour model to determine the location of the tongue edge in the image. If the automatic tracking of the tongue edge does not produce satisfactory results, points can be manually added or subtracted to obtain the best fit. Sixty-four points were extracted for each tongue curve, which were then used for statistical analysis. A screenshot of the tongue curve extraction in EdgeTrak for the frame of the /g/ of "bag dazzled" is shown in Fig. 4.

III. SMOOTHING SPLINE ANOVA FOR COMPARING TONGUE SHAPES

A. Smoothing splines

The 64 points for each of the eight repetitions of /g/ for "bag dazzled" and "Baghdad" extracted with EdgeTrak are shown in Fig. 5. These repetitions are plotted in the statistical package S-Plus 2000 (the commercial version of the open-source R language for statistical computing). The first step is to fit the data using smoothing splines (Eubank, 1988; Green and Silverman, 1994; Wahba, 1990). Smoothing splines have also previously been employed in speech production research. For example, Ramsay *et al.* (1996) provides a technical introduction to the use of smoothing splines in a study of lip motion using OPTOTRAK, an optoelectronic tracking system that transduces the 3-D position of reflective markers. In what follows, a more intuitive introduction to smoothing splines is presented, focusing on how it applies to ultrasound data.

Smoothing splines are a type of natural cubic spline, which is a piecewise polynomial function that connects dis-

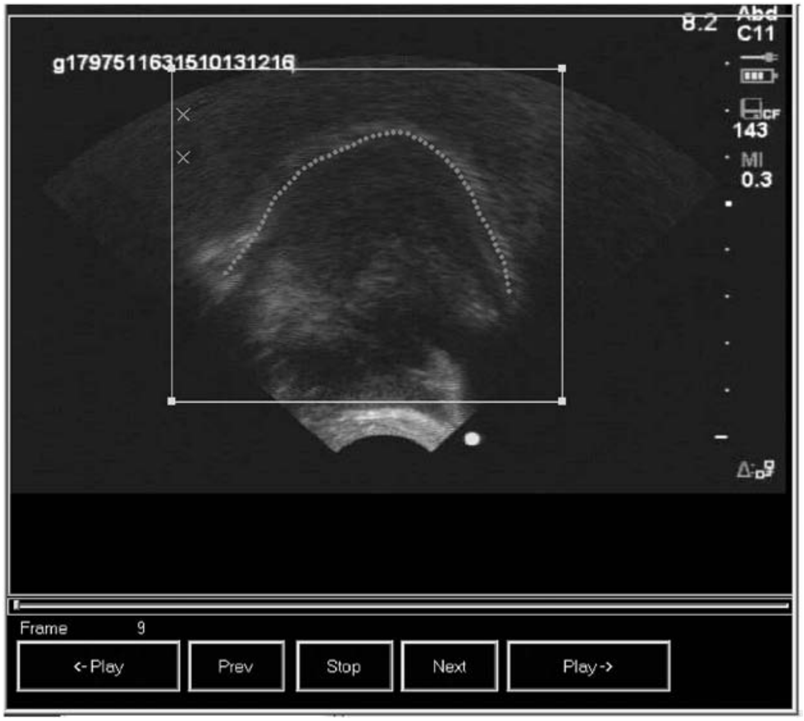


FIG. 4. Screenshot of the EdgeTrak extraction for the frame for /g/ shown in Fig. 1. The program is asked to provide 64 points to characterize the curve shape.

crete data points called knots. Smoothing splines include a smoothing parameter to find the best fit when the data tend to be noisy. More specifically, the function defining the smoothing spline contains two terms: one that attempts to fit the data and one that penalizes a fit which does not have the appropriate amount of smoothness. Although the penalty term does not allow the function to fit the data precisely, it ensures that the resulting spline has a suitable amount of smoothness. Natural cubic splines have the advantage that the shape of the data does not have to be known *a priori*.

The smoothing spline is estimated by minimizing the function in (1):

$$G(x) = \frac{1}{n}\sum_{\text{all } i} (y_i - f(x_i))^2 + \lambda \int_a^b (f''(u))^2\, du, \tag{1}$$

where n is the number of data points, and a and b are the x coordinates of the endpoint of the spline. The smoothing parameter λ is critical to the performance of the spline estimate. If λ is large, the curve will be smoother, whereas a small λ produces a wavier curve that attempts to fit each of the individual data points. The smoothing parameter is determined automatically using the generalized cross validation (GCV) method (technical details on GCV are discussed in Craven and Wahba, 1979; Ramsay *et al.*, 1996). The same function is used to estimate a spline whether the data contain the 64 points of one repetition or the 512 points of eight repetitions.

An example of the smoothing splines corresponding to each data set from the eight repetitions of /g/ in “bag dazzled” and “Baghdad” for subject TO is shown in Fig. 6(a). In this figure, the axes are in pixels, where 1 mm =2.63 pixels. The vertical lines in the figure are a rough division of the tongue into three parts corresponding to the tongue anterior, the body/dorsum, and the root. This type of tentative division allows for the determination of statistical significance in the part of the tongue most relevant to the research question. For the purpose of the data discussed in this paper, the main region of interest for the coronal fricative /z/ is the rightmost third of the tongue corresponding to the anterior parts of the tongue, including the tip and blade. For the velar stops /k/ and /g/, the focus is on the middle third corresponding to the tongue body/dorsum. For now, the tongue is divided into three equal parts for lack of a better assumption about the most linguistically relevant way to determine such divisions. This issue will be discussed again in the general discussion.

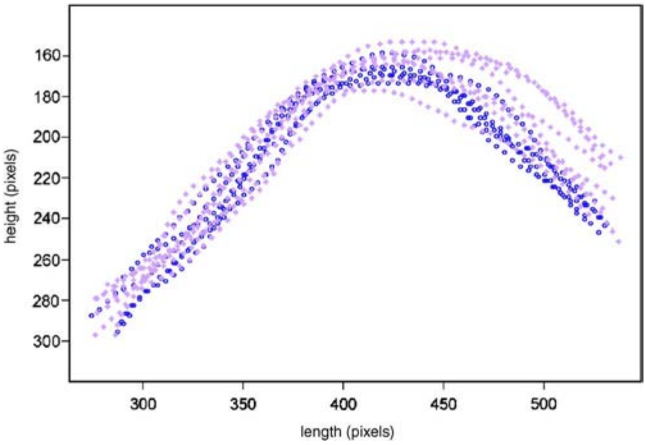


FIG. 5. (Color online) Raw data points from eight repetitions for comparison of the shapes for /g/ in “bag dazzled” and “Baghdad” for speaker TO. “bag dazzled” is represented by the dark blue “o” data points, and “Baghdad” by the pink ‘+’ data points. The x axis is the length of the tongue, and the y axis is the height of the tongue. The scales correspond to the pixels of the original JPEGs, where 1 mm=2.63 pixels and the origin is in the top left corner (accounting for why the values on the y axis increase). Like the ultrasound images, the tongue tip is on the right and the tongue root is on the left.

Some differences between the consonantal articulations

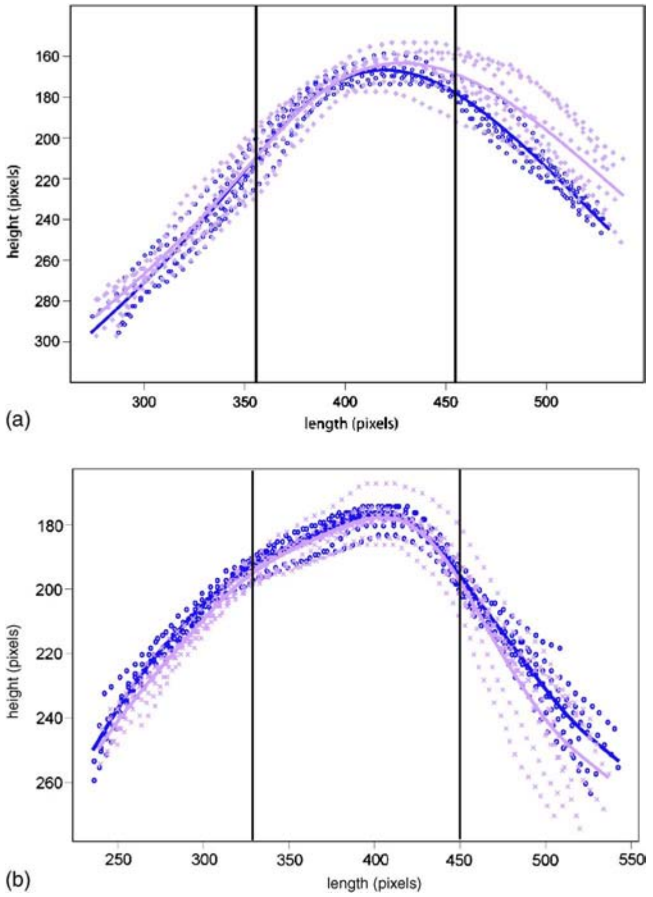


FIG. 6. (Color online) (a) Data points from eight repetitions and smoothing spline estimate (solid lines) for the /g/ in "bag dazzled" and "Baghdad" for speaker TO. "bag dazzled" is represented by the dark blue line and the "o" data points, and "Baghdad" by the pink line and '+' data points. (b) Data points and smoothing spline estimate for the /z/ in "jazz dancer" (dark blue) and "NASDAQ" (pink) for speaker RE.

can be seen impressionistically in Fig. 6. In Fig. 6(a), for example, the tongue blade and body for the /g/ of TO's "Baghdad" is somewhat higher and fronted. Figure 6(b) displays a comparison of the /z/ in "jazz dancer" and "NASDAQ" for speaker RE, with slight raising of the tongue anterior for the /z/ in "jazz dancer." Although no palate shape data were collected for this study, these differences likely correspond to differences in the degree and/or location of constriction for the consonant being produced. In the case of the /g/ of "Baghdad," the constriction location may be more fronted, whereas the /z/ in "jazz dancer" appears to have an increased constriction degree.

B. Smoothing spline ANOVA

The SS ANOVA has been used in applications that require a statistical technique to determine whether the shapes of multiple curves are significantly different from one another. In addition to the study of circadian rhythm mentioned in the Introduction (Wang *et al.*, 2003), SS ANOVAs have also been applied to studies in environmental science and epidemiology (Gu and Wahba, 1993a, b; Wahba *et al.*, 1995).

The SS ANOVA was implemented in S-Plus 2000 using the ASSIST library for fitting spline-based models (Wang and Ke, 2002). The SS ANOVA model is of the form in Eq. (2). Each component of f is estimated with a smoothing spline:

$$f = \mu + \beta x + \text{main group effect} + \text{smooth}(x) + \text{smooth}(x;\text{group}). \tag{2}$$

Unlike a standard ANOVA, the SS ANOVA does not return an F value. Instead, the smoothing parameters of the components smooth(x) and smooth$(x;\text{group})$ are compared to determine their relative contributions to the equation. In the ANOVA model, the main group effects correspond to the smoothing splines for each data set [for example, the dark blue data for "bag dazzled" versus pink data for "Baghdad" in Fig. 6(a), color online], smooth(x) is the single smoothing spline that would be the best fit for all of the data put together (not represented in these diagrams), and the interaction term smooth$(x;\text{group})$ is the smoothing spline representing the difference between a main effect spline and the smooth(x) spline.

The interaction term smooth$(x;\text{group})$ is examined to determine whether the curves representing each group are significantly different. If the two curves being compared have different shapes, then smooth$(x;\text{group})$ will be a significant component of f. Significance is determined by comparing the smoothing parameter value for the interaction term smooth$(x;\text{group})$ with the smoothing parameter value for smooth(x). If smooth(x) and smooth$(x;\text{group})$ are of the same order of magnitude, then it is likely that at least some regions along the two curves are significantly different. In this case, the order of magnitude refers to the nearest power of 10; thus, smoothing parameters with values of 8 and 30 would be considered to be within the same order of magnitude, since both are numerically close to 10^1. However, smoothing parameter values of 8 and 110 would be within different orders of magnitude, since 110 is nearest to 10^2. Furthermore, it should be emphasized that the order of magnitude criteria for the smoothing parameters is only a rough metric that does not guarantee that differences are not significant. In cases of extreme difference, such as values of 0.1 versus 10 000, it may be assumed that there are no significant differences among the curves. However, differences of 0.1 versus 10 may still contain a significant difference at some point along the curve. For the comparison of the /g/ of "bag dazzled" and "Baghdad," the smoothing parameters for smooth(x) and smooth$(x;\text{group})$ are 6.04 and 28.05, respectively. These values are within the same order of magnitude. Visual inspection suggests that the front third of the tongue shapes in Fig. 6(a) are significantly different from one another, but in order to confirm this, 95% Bayesian confidence intervals can be constructed to determine whether the curves are significantly different at any point in the comparison (Gu and Wahba, 1993b; Wahba, 1983).

C. Bayesian confidence intervals

The first step is to construct 95% Bayesian confidence intervals around the smoothing splines for the main effects curves themselves. This is illustrated in Fig. 7. When the

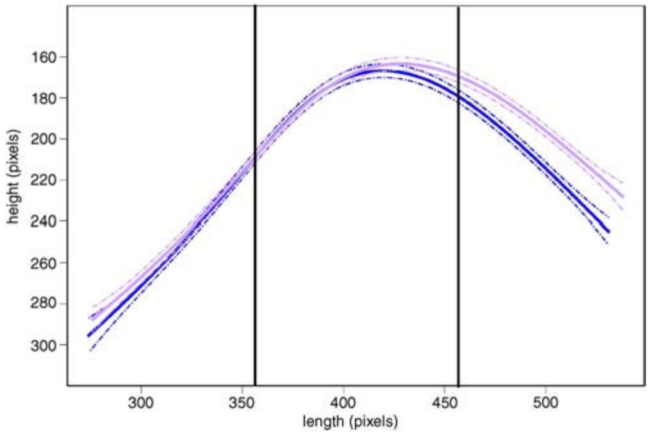


FIG. 7. (Color online) Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the mean curves for /g/ in “bag dazzled” and “Baghdad” for subject TO. “bag dazzled” is represented by the dark blue line, and “Baghdad” by the pink line. The axes and scales are the same as in Fig. 6.

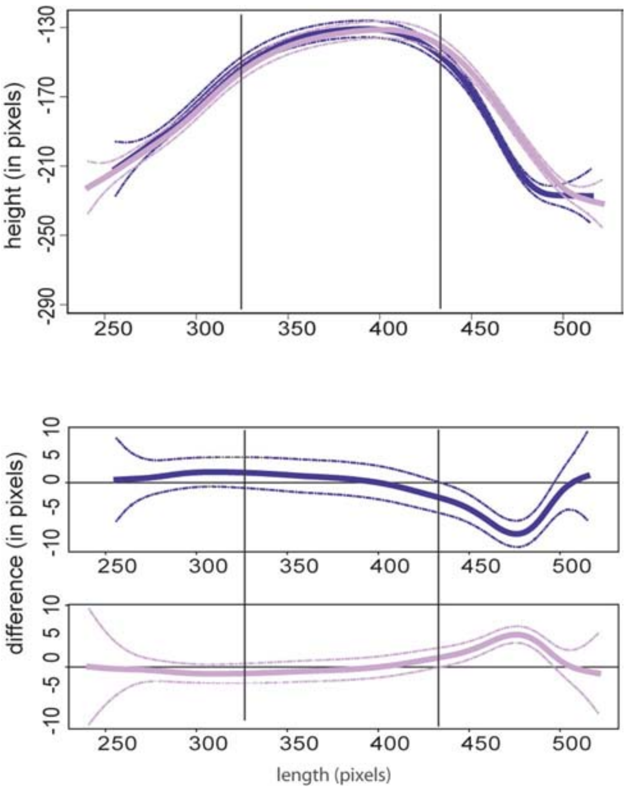
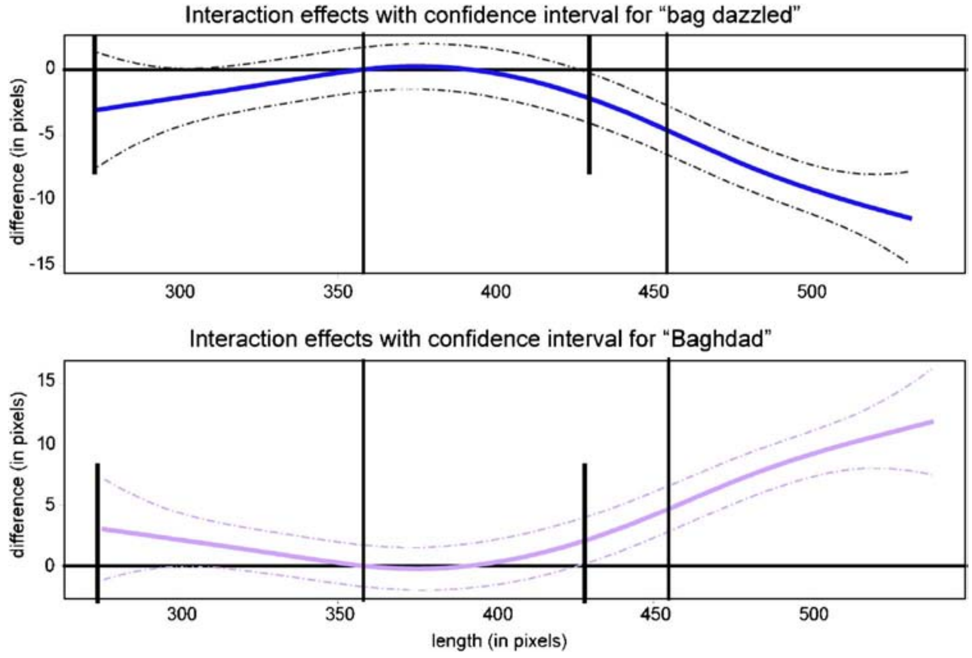


FIG. 9. (Color online) Smoothing splines for data sets (top) and interaction effects with Bayesian confidence intervals (bottom) for the shapes for /k/ in “black top” (dark blue) and “blacktop” (pink) for speaker RE.

confidence intervals of the main effects curves overlap, the differences between two curves are not significant.

To better examine where significant differences are, Bayesian confidence intervals can also be constructed for the interaction curves. The interaction curves for each of the data sets being compared are a plot of the difference of the smoothing spline for each data set from the smoothing spline that is the best fit to all of the data [i.e., smooth(x)]. The interaction effects for the main effects curves shown in Fig. 7 are illustrated in Fig. 8. Though the mean interaction curves for each data set are mirror images, the confidence intervals for each one may be different, which is why both interaction curves are provided in the figures. If the confidence interval encompasses the zero on the y axis at any point along the interaction curve, there is no difference between the two curves being compared; the interaction at that point is not statistically significant. In Fig. 8, the Bayesian confidence intervals encompass zero for about two-thirds of the entire length of the tongue, starting at the tongue root. Thus, the front part of the tongue curves for “bag dazzled” and “Baghdad” are significantly different than one another.

Figure 9 contains the smoothing splines for the production of /k/ in “black top” versus “blacktop” for speaker RE. For this comparison, the smoothing parameter values for smooth(x) and smooth(x; group) were 0.88 and 0.44, respec-


Interaction effects with confidence interval for “bag dazzled”

Interaction effects with confidence interval for “Baghdad”

FIG. 8. (Color online) Interaction effects with Bayesian confidence intervals for the shapes for /g/ in “bag dazzled” and “Baghdad” for speaker TO. The splines representing the interaction effect are mirror images because they represent the difference of main effect spline (as shown in Fig. 7) from the spline that best fits all data for “bag dazzled” and “Baghdad.” However, both images are shown because the confidence intervals can be different. The x axis is length, and the y axis is the difference between each data set and the spline that fits all data for “bag dazzled” and “Baghdad.” When the confidence interval encompasses 0, the curves are not significantly different. The short, thick lines in each image demarcate the part of the interaction curve that is not significantly different.

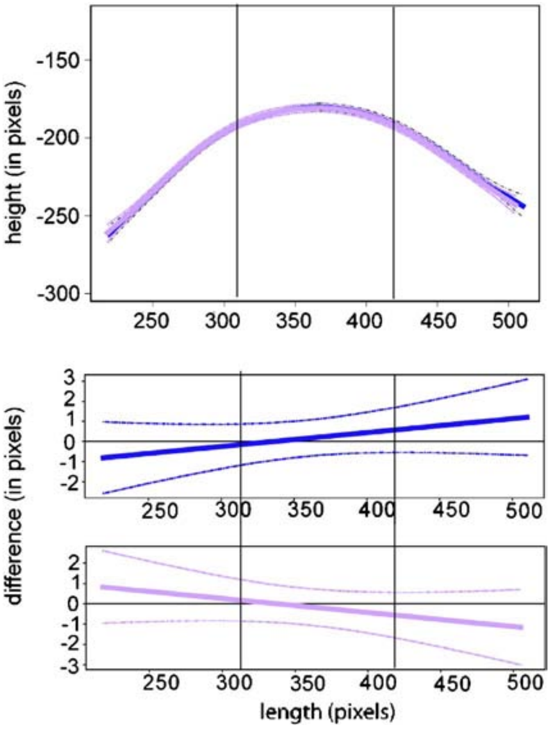


FIG. 10. (Color online) Smoothing splines for data sets (top) and interaction effects with Bayesian confidence intervals (bottom) for the shapes for /z/ in "jazz dancer" (dark blue) and "NASDAQ" (pink) for speaker SH.

tively. In this example, the confidence intervals for the interaction effects are different for "black top" (dark blue line) and "blacktop" (light/pink line) (color online). This is most evident at the ends of the curves, where ultrasound data are often less consistent since the imaging quality at the ends of the tongue curve may be slightly degraded, and therefore harder to accurately track. Such variability will be reflected in the Bayesian confidence intervals of the smoothing splines for both the main effects and the interaction. However, it is also clear in this figure that the confidence interval surrounding the interaction effect for "black top" (dark blue line) is wider than the interval for "blacktop" (light/pink line). This indicates greater variability for "blacktop."

In the example in Fig. 9, the section of the curve relevant to determining whether there is a difference in constriction for the two different types of /k/ is again the middle third. The interaction curves indicate that there is no significant difference anywhere in that region. There is a significant difference along most of the section corresponding to the anterior part of the tongue (the rightmost third), although at the very end of the curve the curves are again very close to one another. This is due to the increased variability at the tongue tip, as indicated by the widening confidence intervals.

Figure 10 demonstrates the production of /z/ in "jazz dancer" (black) and "NASDAQ" (gray) by speaker SH, in which there are no significant differences at all along any point in the curve. The smoothing parameter values for smooth(x) and smooth(x;group) were 7.92 and 2 608 893, respectively.

IV. GENERAL DISCUSSION

The smoothing spline ANOVA is a useful technique for providing a statistical analysis of differences among tongue shapes acquired by ultrasound imaging. When multiple repetitions of an utterance are collected, smoothing splines in conjunction with Bayesian confidence intervals are an appropriate method to account for the shapes that best fit the data and the variance in production. In the examples given above, the articulation corresponding to the most constricted position of a word-final consonant (e.g., *blac**k** top*) was compared to that of a word-medial consonant (e.g., *blac**k**top*). By looking at either the whole tongue curve or a particular region, depending on the researcher's interest, it can be determined whether the tongue shapes for a given articulation are the same or different when some context is varied. In the case of the /g/ in "bag dazzled" versus "Baghdad" for speaker TO (Fig. 7), a significant difference in the rightmost section of the tongue extending into the middle third of the tongue suggests a difference between the constrictions of /g/. In the example of /k/ for "black top" and "blacktop" for speaker RE (Fig. 9) and /z/ for "jazz dancer" and "NASDAQ" for speaker SH, however, there were no significant differences in the relevant regions.

One advantage of the SS ANOVA technique is that any changes in shape, rotation, or translation are taken into account in the statistical analysis. When the head and transducer are stabilized, it can be assumed that any changes not just in the tongue shape, but also in translation (shift on the x or y axis) or rotation of the tongue curve, are of interest to the question being researched. Translation changes, for example, may indicate a change in the backness dimension for a vowel, or may reflect the effects of coarticulation on the production of a consonant.

A few comments about the interpretability of the SS ANOVA should be mentioned. First, unlike methods such as electromagnetic midsagittal articulography (EMMA) (Perkell *et al.*, 1992) or cine-MRI (Stone *et al.*, 2001), ultrasound is not a point tracking technique. Although the smoothing splines representing the tongue shapes for articulation being compared may touch or cross in some spots, it is not the case that the location of contact occurs at the exact same point of the tongue. Thus, the fact that there will be no statistically significant difference between the curves at the point where tongue curves cross should be interpreted with care. As noted in the introduction, a point-tracking technique like EMMA is limited in that it can only provide information about tongue shape and motion for as many pellets as are placed on the tongue (usually around four), whereas the whole midsagittal or coronal contour of the tongue can be imaged by ultrasound.

Second, it is not immediately obvious how the tongue should be divided into linguistically relevant regions. While factor analysis and principle components analysis have been applied to the characterization of tongue configurations in vowel production, these methods are best suited to classifying the tongue shapes of related classes of sounds, not for examining differences in particular regions of interest (Harshman *et al.*, 1977; Hoole, 1999; Nix *et al.*, 1996; Stone and Lundberg, 1996). For example, Harshman *et al.* (1977) developed the PARAFAC ("parallel factors") algorithm in an effort to reduce the number of factors necessary to describe tongue shape. The measurements submitted to the algorithm were based on tracings of midsagittal tongue curves from cinefluorograms which were divided into 18 sections individually determined for each speaker. The results of the PARAFAC analysis indicated that tongue shapes could generally be accounted for by two factors referred to as "front-raising" and "back-raising," which characterize the motion and shapes of the tongue blade and tongue dorsum, respectively. While this method is useful for classifying the overall tongue shape for particular articulations, it does not, for example, lend itself well to determining whether the constriction location and degree for an obstruent consonant in word-final position are statistically different from the same consonant in word-medial position.

In the examples presented in this paper, the tongue was partitioned into three equal sections that can be thought to roughly correspond to the tongue tip/blade, body/dorsum, and root. When examining the location of constriction for velar consonants like /k/ and /g/, the region of greatest interest was the middle third, or the dorsum of the tongue, since this is the section of the tongue that is most relevant to the formation of a velar constriction. However, it is possible—even likely—that the equal division of the tongue surface into three sections is neither the most anatomically nor linguistically accurate method for examining movements and constrictions of different parts of the tongue. One proposal by Iskarous *et al.* (2003) for segmenting the tongue uses conic arcs to model constriction location and constriction degree; perhaps this technique could be used in conjunction with the SS ANOVA to fully quantify tongue shape curves.

Third, related to the issue of linguistically relevant divisions is how to interpret a significant difference in a region of the tongue that is not obviously pertinent to the question being investigated. For example, if a researcher were studying a language that appeared to have a vowel distinction marked by advanced tongue root (ATR) (Ladefoged and Maddieson, 1996), it might be hypothesized that the only region of interest is the tongue root, which should be more advanced or retracted depending on the vowel being produced. However, since the SS ANOVA and the Bayesian confidence intervals for the interaction provide information about the entire tongue (that is, for example, a researcher cannot avoid the statistical comparison of the tongue blade even if it is not the region of interest), it is possible that significant differences will be revealed both in the tongue root and tongue blade region. Would the researcher want to assign any linguistic import to the distinction in the tongue blade? Or, if a difference were found only in the tongue blade region, would the researcher be forced to conclude that the vocalic distinction in question was not an ATR distinction? Such possibilities ought to be considered by researchers in advance so that they are prepared to interpret findings in which the SS ANOVA reveals an unexpected significant difference in some region of the tongue.

Finally, it is important to emphasize that the SS ANOVA is not appropriate for studies that would involve data collection over multiple sessions. It is extremely difficult to ensure that the transducer is placed in exactly the same place across more than one recording session, which results in a different slice of the tongue being imaged each time. This would rule out, for example, pretreatment/posttreatment studies that aim to use the SS ANOVA to quantify the effect of clinical intervention on an articulation of interest. However, the SS ANOVA could still be useful in clinical applications, such as the comparison of the tongue shapes collected within a single session corresponding to correctly produced velar stops with the disordered productions of alveolar stops as palatalized velar stops (Gibbon *et al.*, 1993).

In conclusion, the smoothing spline ANOVA is a promising method for speech researchers examining the tongue contour of an articulation at a moment in time, such as the most extreme articulation of a gesture of interest. In the future, development of methods that facilitate the analysis of changes over time, including an extension of the SS ANOVA to sequential frames of ultrasound data, will permit researchers to compare changes that span more than just the single frame representing the articulation of a sound being studied.

ACKNOWLEDGMENTS

The author thanks statistician Dr. Kyung Sin (Ph.D., UCSB) for her invaluable help in choosing and implementing the smoothing spline ANOVA for analyzing tongue curves. Thanks also to the members of the New York University Phonetics/Phonology Lab and the participants of Ultrafest III workshop at the University of Arizona, April 2005, for their comments on this work. Maureen Stone and Stefan Benus also provided invaluable input on the manuscript. This work was supported in part by the National Science Foundation CAREER Grant No. BCS-0449560.

Bernhardt, B., Gick, B., Bacsfalvi, P., and Ashdown, J. (**2003**). "Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners," Clin. Linguist. Phonetics **17**(3), 199–216.

Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R., and Irish, J. (**2005**). "Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving, and symmetry: Data from 12 normal speakers and a partial glossectomee," Clin. Linguist. Phonetics **19**(6/7), 573–588.

Browman, C., and Goldstein, L. (**1995**). "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (American Institute of Physics, New York).

Craven, P., and Wahba, G. (**1979**). "Smoothing noisy data with spline functions," Numer. Math. **31**, 377–403.

Davidson, L. (**2005**). "Addressing phonological questions with ultrasound," Clin. Linguist. Phonetics **19**(6/7), 619–633.

Eubank, R. (**1988**). *Spline Smoothing and Nonparametric Regression* (Dekker, New York).

Gibbon, F., Dent, H., and Hardcastle, W. (**1993**). "Diagnosis and therapy of abnormal alveolar stops in a speech-disordered child using electropalatography," Clin. Linguist. Phonetics **7**(4), 247–267.

Gick, B. (**2002**). "The use of ultrasound for linguistic phonetic fieldwork," J. Int. Phonetic Assoc. **32**(2), 113–122.

Green, P. J., and Silverman, B. W. (**1994**). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (Chapman and Hall, London).

Gu, C. (**2002**). *Smoothing Spline ANOVA Models* (Springer, New York).

Gu, C., and Wahba, G. (**1993a**). "Semiparametric analysis of variance with tensor product thin plate splines," J. R. Stat. Soc. Ser. B. Methodol. **55**, 353–368.

Gu, C., and Wahba, G. (**1993b**). "Smoothing spline ANOVA with component-wise Bayesian confidence intervals," J. Comput. Graph. Stat. **2**, 97–117.

Guo, W., Dai, M., Ombao, H., and von Sachs, R. (**2003**). "Smoothing spline ANOVA for time-dependent spectral analysis," J. Am. Stat. Assoc. **98**(463), 643–652.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.

Iskarous, K. (**2005**). "Patterns of tongue movement," J. Phonetics **33**, 363–381.

Iskarous, K., Goldstein, L., Whalen, D., Tiede, M., and Rubin, P. (**2003**). "CASY: The Haskins Configurable Articulatory Synthesizer," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Solé, D. Recasens, and J. Romero (Universitat Autónoma de Barcelona, Barcelona), pp. 185–188.

Kochetov, A. (**2006**). "Syllable position effects and gestural organization: Articulatory evidence from Russian," in *Papers in Laboratory Phonology VIII*, edited by L. Goldstein, D. Whalen, and C. Best (Mouton de Gruyter, Berlin).

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford).

Li, M., Kambhamettu, C., and Stone, M. (**2005**). "Automatic contour tracking in ultrasound images," Clin. Linguist. Phonetics **19**(6/7), 545–554. EdgeTrak available at http://speech.maryland.edu/software.html. (website last viewed on 21 April 2006).

Luo, Z., Wahba, G., and Johnson, D. R. (**1998**). "Spatial-temporal analysis of temperature using smoothing spline ANOVA," J. Clim. **11**, 18–28.

Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross-linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707–3717.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Ramsay, J. O., Munhall, K., Gracco, V., and Ostry, D. (**1996**). "Functional data analyses of lip motion," J. Acoust. Soc. Am. **99**, 3718–3727.

Slud, E., Stone, M., Smith, P., and Goldstein, M. (**2002**). "Principal components representation of the two-dimensional coronal tongue surface," Phonetica **59**, 108–133.

Stone, M. (**2005**). "A guide to analyzing tongue motion from ultrasound images," Clin. Linguist. Phonetics **19**(6/7), 455–502.

Stone, M., and Davis, E. P. (**1995**). "A head and transducer support system for making ultrasound images of tongue/jaw movement," J. Acoust. Soc. Am. **98**, 3107–3112.

Stone, M., and Lundberg, A. (**1996**). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. **99**, 3728–3737.

Stone, M., Faber, A., Rafael, L., and Shawker, T. (**1992**). "Cross-sectional tongue shape and linguopalatal contact patterns in [s], [esh], and [1]," J. Phonetics **20**(2), 253–270.

Stone, M., Davis, E. P., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W. *et al.* (**2001**). "Modeling tongue surface contours from Cine-MRI images," J. Speech Lang. Hear. Res. **44**, 1026–1040.

Wahba, G. (**1983**). "Bayesian confidence intervals for the cross validated smoothing spline," J. R. Stat. Soc. Ser. B. Methodol. **45**(1), 133–150.

Wahba, G. (**1990**). *Spline Models for Observational Data* (Society of Industrial and Applied Mathematics, Philadephia).

Wahba, G., Wang, Y., Gu, C., Klein, R., and Klein, B. (**1995**). "Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy," Ann. Stat. **23**(6), 1865–1895.

Wang, Y., and Ke, C. (**2002**). *ASSIST: A Suite of S-Plus Functions Implementing Spline Smoothing Techniques*. Available at http://www.pstat.ucsb.edu/faculty/yuedong/research (website last viewed on 21 April 2006).

Wang, Y., Ke, C., and Brown, M. (**2003**). "Shape-invariant modeling of circadian rhythms with random effects and smoothing spline ANOVA decompositions," Biometrics **59**, 804–812.

Westbury, J. (**1994**). *X-ray Microbeam Speech Production Database User's Handbook, Version 1* (Univ. of Wisconsin, Madison).