Spatiotemporal Visualization of the Tongue Surface using Ultrasound and Kriging (SURFACES)

Vijay Parthasarathy[†], Maureen Stone[‡], and Jerry L. Prince[†].

[†] *Dept. of Electrical and Computer Engineering,*

Johns Hopkins University.

[‡] *Dept. of Biomedical Sciences, Dept. of Orthodontics,*

University of Maryland Dental School.

(Received April 26, 2004; accepted January 11, 2005)

Contact Address

J. L. Prince

105 Barton Hall, The Johns Hopkins University

3400 N. Charles St., Baltimore, MD 21218

Tel:410-516-5192, Fax:410-516-5566,

E-mail: prince@jhu.edu

Abstract

Analyzing the motion of the human tongue surface provides valuable information about speech and swallowing. One method to analyse this motion is to acquire two-dimensional ultrasound images and extract the tongue surface contours from them. Quantitative and statistical analysis of these extracted contours is made difficult because of the absence of physical fleshpoint markers on them. In this research, this problem is overcome by pre-processing the contours using Kriging. Pre-processing includes extrapolating and resampling the contours on a regular spatial grid. The preprocessed contours can then be visualised as spatiotemporal surfaces. A dedicated user interface called SURFACES is designed to aid in the generation, visualisation and quantitative comparisons of these spatiotemporal surfaces.

Keywords: Tongue, Ultrasound, Spatiotemporal surface, Extrapolation, Kriging, Visualization, SURFACES

1 INTRODUCTION

Motion of the surface of the human tongue is of interest because the tongue is critical in speaking, swallowing, and breathing. Being a deformable and volume-preserving object, the tongue can produce a variety of surface shapes through complex activation of its muscles [1;2]. Imaging techniques are often used to depict the shapes of the tongue and the vocal tract. These techniques include both fleshpoint measurements (X-ray microbeam and electromagnetic midsagittal articulator), and imaging techniques (ultrasound [3], X-ray [4], and magnetic resonance imaging (MRI) [4]). Compared to the imaging techniques, the fleshpoint measurements interfere with natural speech and also introduce the methodological problem of extrapolating the tongue surface between and beyond the fleshpoints [5]. The imaging techniques provide a more complete representation of the tongue surface, though they have their limitations. Among the imaging modalities, ultrasound is very attractive for producing an image sequence of tongue motion because of real-time capture rates (30 frames per second), convenience of experimentation, and cost. Ultrasound has been extensively used to analyse speech production [4;6] and to understand the act of swallowing [7;8].

In this paper, we have used a sequence of two-dimensional ultrasound images to understand the motion of the human tongue during speech and swallowing. The sequence of images is acquired at video frame rates and represents the mid-sagittal (lengthwise) section of the tongue (figure 1(a)). To account for the intra-subject variability in speech and swallowing, the image sequences are acquired for multiple repetitions of the same utterance or the same kind of swallow from a single subject. In order to increase the data analysis speed, automatic extraction and tracking of tongue surface contours have been implemented [9] (figure 1(b)). Each set of these extracted tongue contours constitutes a very high dimensional data set – a dense set of points on the tongue (typically around 100 sample points on the tongue) moving over time with data collected at a rate of 30 frames per second (figure 1(c)).

Insert figure 1 about here

While such high dimensional data can be visualised in a spatiotemporal fashion (see waterfall display in figure 1(d)), quantitative comparisons like averaging and comparison, are impossible

because of the absence of physical fleshpoint markers. The absence of physical markers implies that there is no simple point-to-point correspondence between contours, which is necessary for averaging and comparing two contours. Therefore, it is necessary that the contours be sampled on an identical spatial grid and that they be of the same length. If the contours are of equal length and if they are sampled on identical grids, then a spatial correspondence can be established between two spatial points on two contours that share the same x coordinate. But the following three factors lead to the apparent length differences and irregular sampling of contours.

1. Data loss at tongue tip and tongue root – The tongue tip and tongue root are difficult to image using ultrasound. The tongue tip is obscured by the air beneath it and the tongue root is obscured by the shadow of the hyoid bone. This might lead to a change in the apparent length of the extracted contours.
2. Change in tongue contour length – The tongue contours may be different lengths for different repetitions of the same speech-sound due to speaker imprecision. Moreover, the tongue length can change even during one utterance due to the volume-preserving nature of the tongue. For example, vertical expansion or compression must be balanced by an anterior-posterior expansion or compression, respectively, which changes the tongue length.
3. Contour sampling effects – An increase in the gradient of a portion of the extracted contour, increases the density of sampling in that portion. This behaviour of the contour extraction algorithm results in differences both in the spatial sampling locations and local sampling density.

To address these difficulties, our strategy is to pre-process the contours by equalising their lengths, and then resample them on the same grid. Pre-processing methods, such as registering, smoothing, extrapolating, and interpolating data, are necessary steps in many statistical applications [10]. A variety of pre-processing methods have been suggested by Stud et al in Ref. [5] and Stone et al in Ref. [11], for the application of principal component analysis (PCA) on coronal tongue contour data. Methods to equalise the lengths of the contours include combinations of the following three approaches: 1) truncation of the longer contours beyond a defined region; 2) extrapolation of shorter contours to the size of longer ones through linear or spline extension;

and 3) padding shorter curves with constant values. The truncation approach, although good for certain kinds of contours, discards interesting and valid data from the longer curves. Slud et. al [5] discarded the attempt to extrapolate using splines, because of unacceptable swings in the extrapolated contours; instead they used the 'padding' approach, where the shorter curves are padded with endpoint averages. They argue that, even though padding will introduce artificial discontinuities, it did not affect the PCA methods. But, these artificial discontinuities are visually unappealing and can be a problem in other statistical analyses. In this research our approach is to extrapolate the shorter contours using Kriging which produces smooth contours without discontinuities and oscillations.

The problem of spatial data interpolation and extrapolation is common to many scientific areas, for example, image processing, economic forecasting and geostatistics. Various methods have been used for interpolation [12–15] e.g. inverse distance weighting, Kriging, polynomial splines, Hardy's multi-quadratic method, and tension finite difference method. Some of these interpolation algorithms have been used directly for extrapolation, but the results differ in their accuracy. Among these methods, the inverse distance weighting method is considered to be robust in terms of estimation error. This robustness is due to the weighted averaging of data values, resulting in estimates not too far from the actual data. This method, however, introduces abrupt changes in contours, which make the contour non-differentiable, unsmooth and visually unappealing (see figure 2(a)). This makes the contours difficult to be used for further analysis. On the other hand, polynomial splines, especially the higher order splines, sometimes lead to undesirable oscillations in the extrapolated values depending on the gradients of the values near the end of the contours (see figure 2(a)) [16;17].

Insert figure 2 about here

To illustrate the problem, figure 2 shows a typical extracted tongue surface contour. The data corresponding to the extracted contour is represented in the form of a stem plot descending from the top of the plot. Note that the sampling density is higher in locations where the slope is larger, such as at the back of the tongue (on the left). This occurs because the contour itself is sampled

uniformly along its length. Thus, as the slope of the contour increases, the density of sampling with respect to the x-axis also increases. Figure 2(a) shows two extrapolated methods for the same data set: one using inverse square distance weighting (a special case of inverse distance weighting where the weighting exponent is two) and the other using cubic splines. In both cases the quality of interpolation within the tongue surface is good. The problem starts to appear in the extrapolated part. In the case of inverse square distance weighting, the value of the extrapolated values are constrained to stay within the values of the data. Hence there is an abrupt change in shape, which is uncharacteristic of a tongue shape. In the case of cubic splines, clearly there is a non-intuitive and extreme fluctuation in the extrapolation.

In order to avoid the above problems, we use Kriging [18] to extrapolate the tongue shape. Kriging is a statistical estimation technique that uses the statistics of the sampled function to estimate a continuous function that interpolates between the sampled points and also extrapolates beyond the endpoints of the contours. The output of Kriging is a smooth, visually appealing fit of the data, making Kriging suitable for pre-processing the contours. Both the oscillation and the abruptness are absent in figure 2(b), where the extrapolation is done using Kriging. The key to Kriging's improved performance in extrapolation is its spatial asymptotic properties. Also, given the sample data points and their statistics, Kriging estimates a continuous function that best fits the data points. Therefore, the resulting continuous function can be resampled at any given spatial grid. After each contour has been extrapolated and resampled, the contours can be visualised as a spatiotemporal surface and can be analysed using a dedicated software tool called SURFACES, which we also present in this paper.

2 METHODS

2.1 Data acquisition

We acquire a sequence of ultrasound images of the mid-sagittal section of the tongue (figure 1(a)). The sequence of ultrasound images is acquired as the subject either speaks a given utterance or swallows a particular bolus. One of the images in an ultrasound sequence is shown in figure 1(b),

with the extracted contour overlayed as white dots. The ultrasound scan rate is set to 30 images per second. Each subject is asked to repeat the utterance multiple times (usually 7 times, with the first and last omitted from further processing) in order to account for the intra-subject variability in speech production. The audio data is also recorded, but it is not directly useful in the context of this paper. The sequence of images is acquired both in analog and in digital format. The images are then input into the contour extraction program, which is described in the next section.

2.2 Automatic contour extraction and tracking

Each image in the sequence is processed using the algorithm proposed by Li et al [9]. The algorithm uses a discrete form of deformable contours and imposes speech, tongue, and ultrasound imaging constraints. The initial contour of the tongue shape is user-defined; it is then used as the initialisation for the deformable model. Using the initial contour and the model constraints, the algorithm tracks the tongue surface over the series of images. The algorithm also imposes regularising constraints on the deformable contours, so that the resulting contour is smooth. Each contour is represented as a set of $\underline{y}$ values, which represents the height of the tongue (calculated from the top of the image) measured at sampling locations determined by the $\underline{x}$ values (figure 1(c)). A dedicated software tool incorporating the algorithm is used to extract and track the contours from the ultrasound image sequences (see Li et al in this volume for more details). These contours are the input for the pre-processing using Kriging.

2.3 Introduction to Kriging

Kriging (pronounced with a long i-vowel) is named after the South African mining engineer D. Krige who developed it for estimating mineral deposits from scattered ore samples [12;19]. Since then it has been used to interpolate spatially dependent data in a wide variety of disciplines. Kriging is a modified linear regression technique that estimates a value at a point by assuming that the value is spatially related to the known values in the neighborhood of the point. Kriging computes the value for the unknown data point using a weighted linear sum of known data values. The weights

are chosen to minimise the estimation error variance while keeping the average estimation error zero. Hence, Kriging is called the best linear unbiased estimator because it theoretically tries to minimise the variance of estimation error, while being an unbiased estimation procedure [19].

Direct minimisation of error variance is not possible because the true values are unknown. Hence, Kriging uses a random function model, where the data points are assumed to be realisations of random variables and the point to be estimated is also a random variable. These random variables are assumed to have specific covariance structure; selection of which is crucial in the estimation procedure. So given the model, the error variance can be modelled and then minimised under the unbiasedness constraint to get the Kriging solution.

2.4 Derivation of Kriging solution

Given observations at spatial points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p$, we want to estimate the value of the function at any spatial point $\mathbf{x}$. Kriging estimates a continuous function $s(\mathbf{x})$, so that the average estimation error is zero and the error variance is minimum. Since Kriging estimates a continuous function, we can get the value of the function at any point $\mathbf{x}$.

In our case, the observations are the 'y_i' values that measure the height of tongue contours from the top of the ultrasound image at sampling points 'x_i'. Since the x_i's are one-dimensional, we let $\mathbf{x} = x$, a one-dimensional variable. Kriging models the estimated function $s(x)$ as consisting of two components

$$s(x) = u(x) + \mathbf{f}^{\mathrm{T}}(x)\mathbf{d}\,. \tag{1}$$

The first term $u(x)$ is a zero-mean random function with known covariance function $k(x_a, x_b)$ [20]. The covariance function models the spatial correlation in the data. The second term $\mathbf{f}^{\mathrm{T}}(x)\mathbf{d}$ is the mean of the function $s(x)$. The term $\mathbf{f}(x)$ is $r \times 1$ vector of known 'drift functions' and $\mathbf{d}$ is the $r \times 1$ vector of unknown 'drift coefficients'. The mean of the function $s(x)$ is deterministic, but unknown. Usually the drift functions are taken to be monomials of degree less than or equal to a chosen value q. In our 1-D case, $r = q + 1$. Given the shape of the tongue contours, we have

selected $q = 1$, which leads to linear drift functions ($r = 2$),

$$\mathbf{f}(x) = [1 \;\; x]^{\mathrm{T}}.$$

Intuitively, Kriging can be seen as estimating two components of the contour $s(x)$. The mean, $\mathbf{f}^{\mathrm{T}}(x)\mathbf{d}$, captures the global shape of the contour, while the zero-mean random function, $u(x)$, captures the variation of the contour around its mean. The mean is a linear combination of drift functions. In this research we use linear drift functions which means that the global shape of the contours is captured with a linear function with a given slope and an intercept. The variation around the mean is captured by the zero mean random function $u(x)$. The behaviour of these two components is critical in determining how the extrapolated curve will look. A brief discussion on the extrapolation properties is discussed after the Kriging solution has been developed.

Given this statistical model for the data, Kriging produces the Best Linear Unbiased Estimate (BLUE), which consists of a linear combination of the observations.

$$\hat{s}(x) = \mathbf{a}^{\mathrm{T}}\mathbf{y}\,,$$

where $\mathbf{y}$ is a vector of the observations ('y values) at $x_1, x_2, ..., x_p$ and $\mathbf{a}(x)$ is a $p \times 1$ vector of coefficients, which we want to estimate. The constraint of unbiasedness of the estimate leads to the constraint on the coefficients,

$$F\mathbf{a}(x) = \mathbf{f}(x)\,,$$

where

$$F = [\mathbf{f}(x_1) \ldots \mathbf{f}(x_p)]\,, \tag{2}$$

which is an $r \times p$ matrix. The Kriging estimate is then obtained by finding $\hat{s}(x)$ which minimises the estimation error variance

$$E[(s(x) - \hat{s}(x))^2]$$

subject to the unbiasedness constraint. The constrained minimisation problem can be solved using the method of Lagrange multipliers and the solution depends only on $\mathbf{f}(x)$, F, data covariance

matrix

$$K = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_p) \\ \vdots & \ddots & \vdots \\ k(x_p, x_1) & \dots & k(x_p, x_p) \end{bmatrix}, \tag{3}$$

and the covariance vector

$$\mathbf{k}(x) = [k(x, x_1) \dots k(x, x_p)]^{\mathrm{T}}. \tag{4}$$

The solution is

$$\hat{s}(x) = \mathbf{k}^{\mathrm{T}}\mathbf{w} + \mathbf{f}^{\mathrm{T}}(x)\hat{\mathbf{d}}, \tag{5}$$

where

$$\begin{aligned} \mathbf{w} &= K^{-1}[I - F^{\mathrm{T}}(FK^{-1}F^{\mathrm{T}})^{-1}FK^{-1}]\mathbf{y}, \\ \hat{\mathbf{d}} &= (FK^{-1}F^{\mathrm{T}})FK^{-1}\mathbf{y}. \end{aligned}$$

For more details on the derivation of Kriging, please see reference 21. Details of the algorithm implemented in this paper are given in Appendix A.

Thus, the solution of Kriging is a continuous function $\hat{s}(x)$, which can be resampled on an arbitrary spatial grid, thus overcoming the irregular sampling problem. Notably, the spatial grid can include extrapolated points that are beyond the original range of x_i's over which the data was collected. This compensates for apparent length changes of tongue contours because of data loss and speaker imprecision.

The selection of the covariance structure of the data is important in Kriging estimation. In our algorithm, we use the generalised covariance function, $k(x_a, x_b) = \|x_a - x_b\|^2 \ln\|x_a - x_b\|^2$. The use of this covariance function makes our Kriging solution the same as the thin-plate spline solution [22]. Thin-plate spline is an interpolation method that estimates a smooth curve that passes through all given data points so that the final curve is minimally bent. The name 'thin plate spline' refers to a physical analogy involving the bending of a thin sheet of metal, when the tongue heights are set as deflections of the metal plate in the z-direction. In our case, we deal with a 1-D analog of

this bent metal sheet. It has been shown that the thin-plate spline is a special form of Kriging [23] and under certain conditions they are the same.

Thin plate splines are smooth and asymptotically parallel to the mean of the estimated function. In the extrapolated region, while the zero mean random function tends to flatten out, the mean function continues on its trend, thus dominating the behaviour of the curve. So, during extrapolation the contour typically follows the global trend of the contour, which in our case is linear because of the use of linear drift terms. Since the thin-plate spline solution has a smoothing term built in, the extrapolation will be smooth. Unacceptably huge drifts can occur while extrapolating with thin-plate splines; but it happens only at points that are much further away from the data, when compared to the spread in data locations. In typical cases, we extrapolate less than 6 mm on either side of the tongue, where the extrapolation performs reasonably well. A detailed validation of the quality of the extrapolation and the estimation errors are presented later in this paper in section 4.

A typical ultrasound data set contains 13-40 contours depending on the length of the speech utterance or swallow and the video frame rate of the ultrasound scanner. Each contour is extrapolated and resampled using the above method. Then the contours are stacked as a spatiotemporal surface [see figure 3(b)]. Similar processing can be done on different repetitions of the same speech utterance or swallow, and resulting surfaces can be averaged to yield an average spatiotemporal surface.

Insert figure 3 about here

2.5 SURFACES software

Figure 3(a) shows a snapshot of the graphical user interface (GUI) for SURFACES (available for download at www.speech.umaryland.edu/software). The GUI and the algorithm were implemented in MATLAB Version 6 (Mathworks, Natick MA, USA) and ported to a stand-alone version. The GUI has five main panels. The functions in the first and second panels pre-processes individual contour for further analysis. The program reads in the initial contour sequences and allows the user to select maximum and minimum values of $\underline{x}$, within which each contour will be cut or extended,

smoothed (estimated) using Kriging and then resampled. The 'Krige and Show Surface' button kriges all the contours resulting in a spatiotemporal surface, as shown in figure 3(b). This surface looks similar to the waterfall display in figure 1(d). Unlike the waterfall display, this surface can be directly used for further processing (e.g. averages, differences etc), because all the contours have an equal number of samples on the same grid.

The spatiotemporal surfaces that are derived from the kriged contours can be used to qualitatively analyse a speech utterance. For example, figure 3(b) shows the spatiotemporal surface for the word 'golly'. Noting that the front of the tongue is on the right, the nearest contour shows the 'g', which is arched in the middle [see 1 in figure 3(b)]. As time advances, the tongue flattens and the tip rises for the 'l' [see 2 in figure 3(b)]. Finally the tongue arches again for 'y' [see 3 in figure 3(b)].

Panel 3 of the software is for averaging different repetitions of the same utterance that have been kriged and resampled in Part 1. Since the samples are on a regular spatial grid, the averaging is done for different y values at each x coordinate. By averaging different y values at each x coordinate, we are implicitly making a point-to-point correspondence of different y values which share the same x coordinate. The result is an averaged spatiotemporal surface, and a variance surface. Panels 4 and 5 of the software are used for comparison of two spatiotemporal surfaces like overlaying surfaces and calculating local or global differences. These spatiotemporal surfaces can be either individual repetitions or average surfaces (see figure 4(a) for an example of an overlay of two such surfaces). The current version of SURFACES implements two algorithms for calculating the difference between spatiotemporal surfaces. These include a simple difference of y at each x and a nearest-neighbor algorithm [24] to find the shortest distance between two surfaces. These distances are further used for calculating L2 difference norms and root mean squared differences. More details about the algorithms used can be found in the user manual for SURFACES (www.speech.umaryland.edu/software).

3 Results and applications

Applications of the SURFACES software is demonstrated on two kinds of data: 1. speech data collected to find the effect of gravity on tongue and 2. swallowing data collected to find the effects of anterior open bite on swallowing stability.

Insert figure 4 about here

3.1 Application to speech data

This application demonstrates the use of comparative analysis between two spatiotemporal surfaces corresponding to two different speech utterances. The goal of this study was to understand the effects of gravity on the tongue during speech [25]. The subjects were asked to repeat the same utterances in a supine position first and then in an upright position. Tongue contours were extracted from the ultrasound data, kriged, averaged and visualised using SURFACES. The overlaid surfaces in figure 4(a) shows a typical result during the utterance of the word 'golly'. We see that the supine surface (filled surface) is rotated backward from the upright surface (white mesh) during the entire word. A secondary effect that can also be observed is that tongue tip is elevated in the supine position during the 'l' (see arrow). The two surfaces can also be visualised as a difference image (figure 4(b)) with the colors denoting the amount of difference between each pixel of the two surfaces.

3.2 Application to swallowing data

This application demonstrates the use of visualising the spatiotemporal surfaces and drawing qualitative physiological inferences from it. This experiment studied the effects of anterior open bite on swallowing stability [26]. Figure 4(c) shows the spatiotemporal surface of a 20 cc water swallow. We observe that the water is initially contained anteriorly, with the tongue tip depressed and the back elevated to protect the airway. Subsequently, the tongue deforms around the bolus as it is propelled backwards. Finally, the tongue elevates from front to back to make contact with the palate after the water's passage. This spatiotemporal surface can also be rotated to various views.

Figure 4(d) shows the spatiotemporal surface of figure 4(c) as a 2-D image where the color denotes the tongue height. The black lines in figures 4(c) and (d) separate the regions that which contain true data from the regions that contain extrapolated data.

4 Validation of contour extension

Recall that, given the data values at specified spatial points, a Kriging solution estimates the value at any spatial point. The Kriging estimate is the 'best' in the sense that it theoretically minimises the error variance while maintaining the mean error zero. Hence it is possible to get an estimate of the minimum error variance even before the estimation is done. But this estimate of the minimum error variance is the predicted error variance of the model and cannot be completely trusted without testing the model on real data [12]. The validation test on real data is more crucial in the case of extrapolation, since the range of errors produced in extrapolation tends to be larger than in interpolation.

4.1 Validation materials and methods

A total of 1612 tongue contours were used for the validation test. The data was collected for the upright-supine study [25] and was in compliance with an approved human subject experiment protocol. The validation data set contained contours from 4 different words (golly, oslo, he sought, he taught), 5 different speakers and 2 different positions (upright and supine).

Portions of the tongue contour of length approximately 1 mm, 3 mm, 5 mm and 10 mm were artificially cut (see figure 5). All the lengths are distances along the surface of the tongue contour and not along the spatial axis (x-axis). The center region combined with the lighter gray (lower) regions was the initial full contour. The artificial cuts were made from both the back and the front of the tongue contour (gray) regions in figure 5). These cuts simulated the loss of data and the apparent change in length as discussed in the introduction. Kriging was then used to restore these cut portions (black regions in figure 5) and the error was measured as estimated curve minus the true curve. This procedure was done for all the 1612 contours. The

errors were also averaged separately for the front and the back of the tongue. Standard deviation of the errors was also calculated.

Insert figure 5 about here

Insert figure 6 about here

Insert figure 7 about here

4.2 Validation results

Figure 6 shows the average errors (at each point) for the four different length cuts. The black curve shows the error at the back of the tongue, whereas the gray curve shows the error at the front of the tongue. The x-axis in these graphs represents the distance from the estimated point to the cut, (i.e. the last data points on the edge of the contour represented as black circles in figure 5). Error bars represent standard deviation. For a given length of extrapolation, the error measures in Figure 6 give an estimate of the amount of confidence that can be placed on the calculated values, depending on whether the extrapolation is done in the front or in the back of the tongue. For example if the contour is extrapolated to a length of 5 mm (figure 6(c)), then at a point 4 mm away from the actual data, the expected error is around -3.2 mm in the back and around -4 mm in the front of the tongue. Also, we note that in all cases the error is negative, which implies Kriging always underestimates the values. This underestimation of the true curve is because, the extrapolated contour tries to follow the global shape of the tongue contours, which in many cases, a line with a positive slope. Therefore, in the front of the tongue, the extrapolated contour curves up, whereas in the back of the tongue the extrapolated contour curves down. In the actual data the back of the tongue slopes downward, whereas the front of the tongue stays flat, in most cases (see for example figure 1(c)). Hence, we see that error at the front of the tongue is slightly, but consistently, higher than the back of the tongue. This behaviour, however, depends entirely on the global scope of the particular tongue surface that is being analysed.

Figure 7 represents the worst case analysis of Kriging extrapolation. Maximum expected error is plotted as a function of the length of tongue cut for both the front and the back. The maximum

error occurs at the point which is farthest away from the data. The maximum error and the error variance also increase with increasing length of data loss.

4.3 Discussion

We notice that the errors are large when the amount of extrapolation is large. It is natural to expect this trend because we are moving away from where actual data exists. The inverse distance methods may provide a lower error measure because the values are always constrained to be within the data values. But this lower error measure does not have a physical meaning and is only of statistical interest. This is because the curve estimated using inverse distance methods has large discontinuities, thus ignoring the physical reality of the tongue. Moreover the contours that are estimated using inverse distance methods can neither be used for averaging repetitions nor for doing comparative studies.

It is also important to note that the errors mentioned in this section are extrapolation errors. Using Kriging with the generalized covariance function for interpolation is extremely robust and the estimates have a very low value of errors [12] (see figure. 2(b)). So, the values estimated in the interior of the tongue have low errors (see the central overlapping region of the gray and black curves in figure 5).

Even though the Kriging solution is useful for visualisation, averaging and comparative analysis, physiological inferences derived using these extrapolated regions should be used with caution because the extrapolation errors. On the other hand, interpolation error is very small and hence the quantitative measures in the non-extrapolated regions of the tongue can be used with high degree of confidence. With regard to the issue of knowing which regions of the contours are extrapolated, the 'SURFACES' software has two important features: 1) When visualising a spatiotemporal surface a mask is generated which tell the user which data points are real and which have been artificially kriged (figure 4(c) and (d)); 2) when averaging different repetitions, rules have been implemented so that an averaged value will be generated only at those $\underline{x}$ points where a certain number of real (non-extrapolated) $\underline{y}$ values are available.

One of the limitations of Kriging is that its solution, like all spline-based interpolation methods,

becomes unstable if there are two points with the same x value, but different y values. This situation can occur when the tongue surface curls or when the tongue surface becomes exactly vertical. In these cases, Kriging might fail to give reasonable contours. Therefore, for such cases, the 'SURFACES' software implements a local contour adjustment routine. The tongue contour is locally tweaked by changing the x-coordinate of one of the two points. Different amounts of tweaking were tried on locally vertical contours from data sets of the upright-supine study. The minimum of these local tweaks that provided reasonable results for all contours was ± 0.3 mm. Therefore a bias of ± 0.3 mm was chosen as final amount of tweaking. This bias is within the typical ultrasound measurement error of ± 0.5 mm [27]. The adjusted contour is then subsequently kriged and visualised.

5 Conclusion

We described a method of visualising, quantifying and comparing tongue surface features from contour sequences. Kriging was used to extrapolate the tongue surface contours that are extracted from ultrasound image sequences of the tongue. The resulting kriged contours are then stacked and visualised as a spatiotemporal surface. A dedicated software tool, SURFACES, which incorporates the Kriging algorithm is presented. The tool is used for averaging and comparative analysis of different tongue shapes. The calculation and visualisation of spatiotemporal mid-sagittal tongue surfaces helps in understanding tongue deformations during speech and swallowing. It is hoped that this methodology will further help in quantification and statistical comparison of complex tongue motion.

The main problem that was overcome by this research is the lack of point-to-point correspondence between the extracted tongue contours. This problem was solved by equalising the length of the contours and resampling them on an identical grid, thus establishing a correspondence of two points which share the same $\underline{x}$ coordinate. Ongoing research in this field (Li et. al in this issue) is to design algorithms for estimating true point-to-point correspondences based on curvature and other shape properties of the tongue. These correspondences can also be used for registering the

data in time and space. In the future, these algorithms can be combined with Kriging to further improve the quantitative measures of tongue shapes.

6 Acknowledgments

This research was supported in part by a grant (R01 01758) from National Institutes of Health of United States of America.

References

[1] A. J. Lundberg and M. Stone. Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data. *Journal of the Acoustical Society of America*, 105(5):2858–2867, November 1999.

[2] W. M. Kier and K. K. Smith. Tongue, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83:307–324, 1985.

[3] M. Stone and A. Lundberg. Three-dimensional tongue surface shapes of english consonants and vowels. *Journal of the Acoustical Society of America*, 99(6):3728–3737, June 1996.

[4] M. Stone. Imaging the tongue and vocal tract. *British Journal of Disorders of Communication*, 26:11–23, 1991.

[5] E. Slud, M. Stone, P. Smith, and M. Goldstein Jr. Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*, 59:108–133, August 2002.

[6] E. Keller and D. J. Ostry. Computerized measurement of tongue dorsum movements with pulsed-echo ultrasound. *Journal of the Acoustical Society of America*, 73:1309–1315, 1983.

[7] G. Chi-Fishman, M. Stone, and G. N. McCall. Lingual action in normal sequential swallowing. *Journal of Speech, Language, and Hearing Research*, 41:771–785, 1998.

[8] B. Wein, S. Klajman, W. Huber, and W. H. Doring. Ultrasound examination of coordination disorders of the tongue during swallowing. *Nervenzart*, 59:154–158, 1988.

[9] M. Li, C. Kambhamettu, and M. Stone. Snake for band edge extraction and its applications. In *Intl. Conf. on Computers, Graphics and Imaging*, August 2003.

[10] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, USA, 1997.

[11] M. Stone, M. H. Goldstein, and Y. Zhang. Principal component anlysis of cross sections of tongue shapes in vowel production. *Speech Communication*, 22:173–184, 1997.

[12] E. H. Isaaks and R. M. Srivastava. *An introduction to Applied Geostatistics*. Oxford University Press, New York, NY, 1989.

[13] C. Caruso and F. Quarta. Interpolation methods comparison. *Computers Math. Applic.*, 35(12):109–126, 1998.

[14] R. Franke. Scattered data interpolation:test of some methods. *Mathematics of Computation*, 38(157):181–200, 1982.

[15] H. Theil. *Applied Economic Forecasting*. North-Holland Publishing Company, Amsterdam, 1966.

[16] P. Lancaster and K. Salkauskas. *Curve and Surface Fitting - An Introduction*. Academic Press, London, UK, 1986.

[17] P. K. Kitanidis. *Introduction to Geostatistics*. Cambrige University Press, Cambrige, UK, 1997.

[18] R. Christensen. *Linear Models for Multirate, Time series, and Spatial Data*. Springer-Verlag, New York, USA, 1991.

[19] R. Parrot, M. R. Stytz, P. Amburn, and D. Robinson. Towards statistically optimal interpolation for 3-d medical imaging. *IEEE Eng. in Medicine and Biology*, pages 49–59, September 1993.

[20] W. S. Kerwin. *Space-Time Estimation of left Ventricular Motion from Tagged Magnetic Resonance Images*. PhD thesis, Johns Hopkins University, 1999.

[21] G. S. Waston. Smoothing and interpolation by kriging and with splines. *Math. Geo.*, 16(6):601–615, 1984.

[22] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[23] J. T. Kent and K. V. Mardia. The link between kriging and thin-plate splines. In *Probability, Statistics and Optimization – A Tribute to Peter Whittle*, pages 325–338. John Wiley and Sons, 1994.

[24] M. Stone, E. P. Davis, A. Douglas, M. NessAiver, R. Gullapalli, W. Levine, and A. Lundberg. Modeling tongue surface contours from cine-mri images. *Journal of Speech, Language, and Hearing Research*, 44(5):1026–1040, October 2001.

[25] M. Stone, U. Crouse, and M. Sutton. Exploring the effects of gravity on tongue motion using ultrasound image sequences. In *J. Acoust. Soc. Am.*, volume 111, page 2476, May 2002.

[26] M. Noorani. Effect of simulated bite opening on swallowing patterns in normal adults. Master's thesis, Univ. of Maryland Dental School, Baltimore, MD, February 2003.

[27] M. Stone, T. Shawker, T. Talbot, and A. Rich. Cross-sectional tongue shape during the production of vowels. *J. Acoust. Soc. Amer*, 83(4):1586–1596, 1988.

A Appendix – Kriging algorithm

Given a contour in terms of x_i (spatial sampling locations) and y_i (height of the point from the top of the image), the problem is to estimate the value of a continuous function $s(\cdot)$ at arbitrary spatial position $x \in \mathbb{R}$. Kriging algorithm is applied on the contour to estimate $\hat{s}(x)$. The detailed algorithm is given below.

Algorithm 1 *1. Form the data vector,* $\mathbf{y} = [y_1 \dots y_p]$.

2. *Select the drift function* $\mathbf{f}(x)$ *and calculate* F *as defined in Eq. (2).*
   We used the linear drift function, $\mathbf{f}(x) = [1 \;\; x]^{\mathrm{T}}$.

3. *Select the covariance function for the data,* $k(x_a, x_b)$ *and calculate the matrix* K *and vector* $\mathbf{k}(x)$ *as defined in Eqs. (3) and (4) respectively. We used the generalised covariance function,* $k(x_a, x_b) = \|x_a - x_b\|^2 \ln \|x_a - x_b\|^2$.

4. *Select the noise covariance matrix* Σ*, a* $p \times p$ *matrix that characterises the statistics of the noise in the data.*
   In this work we assumed zero noise variance in the contour data. The contour extraction algorithm incorporates smoothing routines and hence the output contour is already smooth and noise free. But noise can be easily incorporated into the algorithm, but assuming white zero mean noise with variance equal to $\sigma \;\; \mathrm{mm}^2$*. Hence,* $\Sigma = \sigma I$*,where* I *is the* $p \times p$ *identity matrix. The use of non-zero noise variance make Kriging a smoother rather than interpolator [21].*

5. *Calculate the matrices*

$$\begin{aligned} L &= (K + \Sigma)^{-1} \\ M &= (FLF^{\mathrm{T}})^{-1} FL \\ G &= KL(I - F^{\mathrm{T}} M) \end{aligned}$$

6. *Calculate the coefficient vectors*

$$\begin{aligned} \hat{\mathbf{d}}_s &= M\mathbf{y} \\ \mathbf{w}_s &= K^{-1}G\mathbf{y} \end{aligned}$$

7. *Calculate the desired estimate using*

$$\hat{s}(x) = \mathbf{k}^{\mathrm{T}}(x)\mathbf{w}_s + \mathbf{f}^{\mathrm{T}}(x)\hat{\mathbf{d}}_s.$$

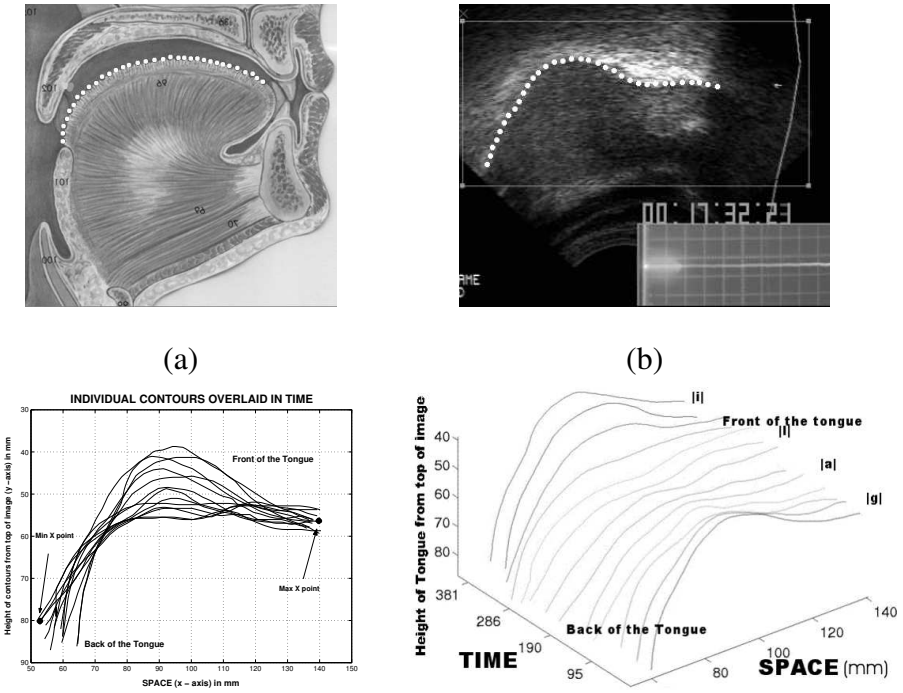
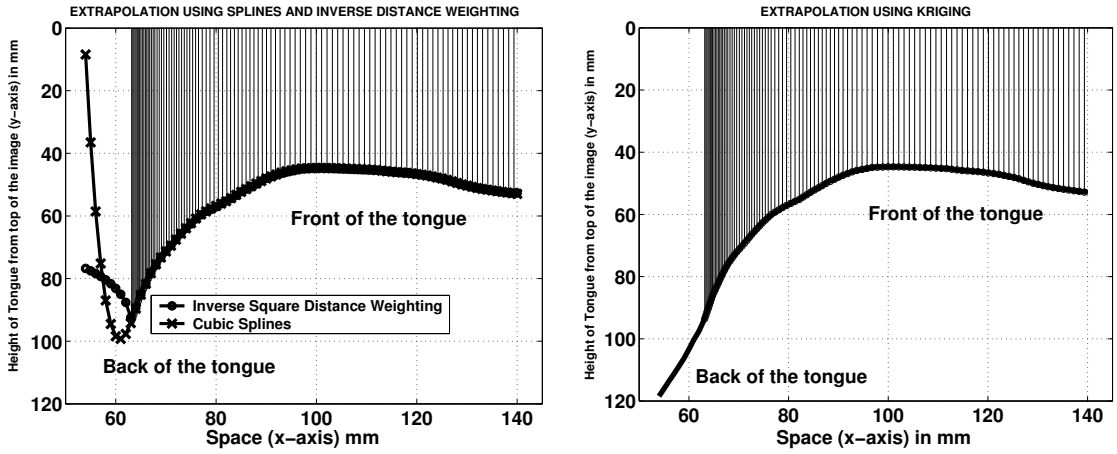
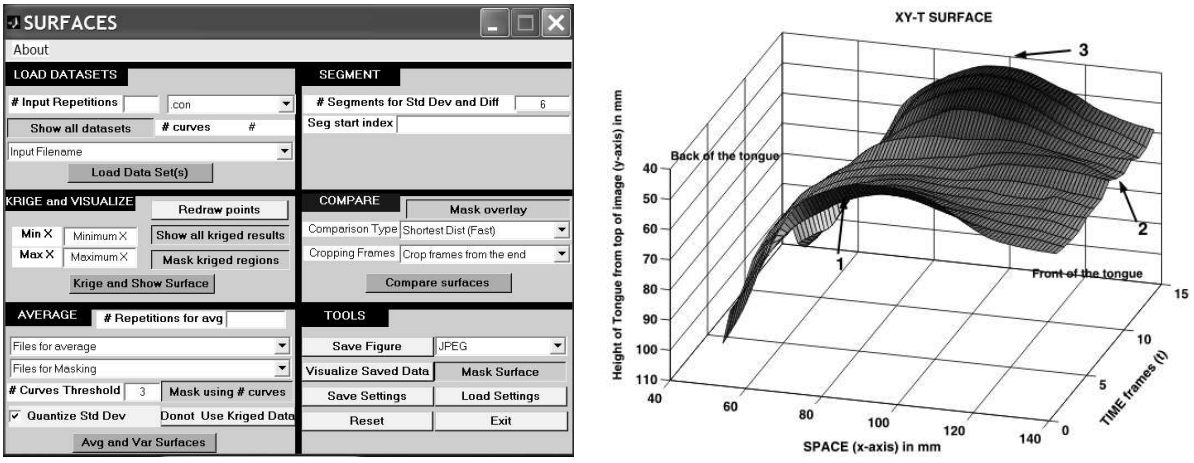


Figure 1: (a) Midsagittal tongue schematic with superimposed surface contours points. (tip of tongue on the right). (b) Midsagittal ultrasound image with tracked surface contour points (c) Sequence of tongue contours in time overlaid on each other. (d) Waterfall display of contours for the word ‘golly’.



(a) (b)

Figure 2: Comparison of Kriging with inverse square distance weighting and cubic splines: (a) Extrapolation using cubic splines and inverse square distance weighting; note the swing in the case of cubic spline(cross) and the unsmooth contour produced by inverse square distance weighting methods. (b) Extrapolation using Kriging; note the improved performance of Kriging at the back of the tongue.



(a) (b)

Figure 3: (a) A snapshot of the SURFACES software and (b) a spatiotemporal surface of the word ‘golly’.

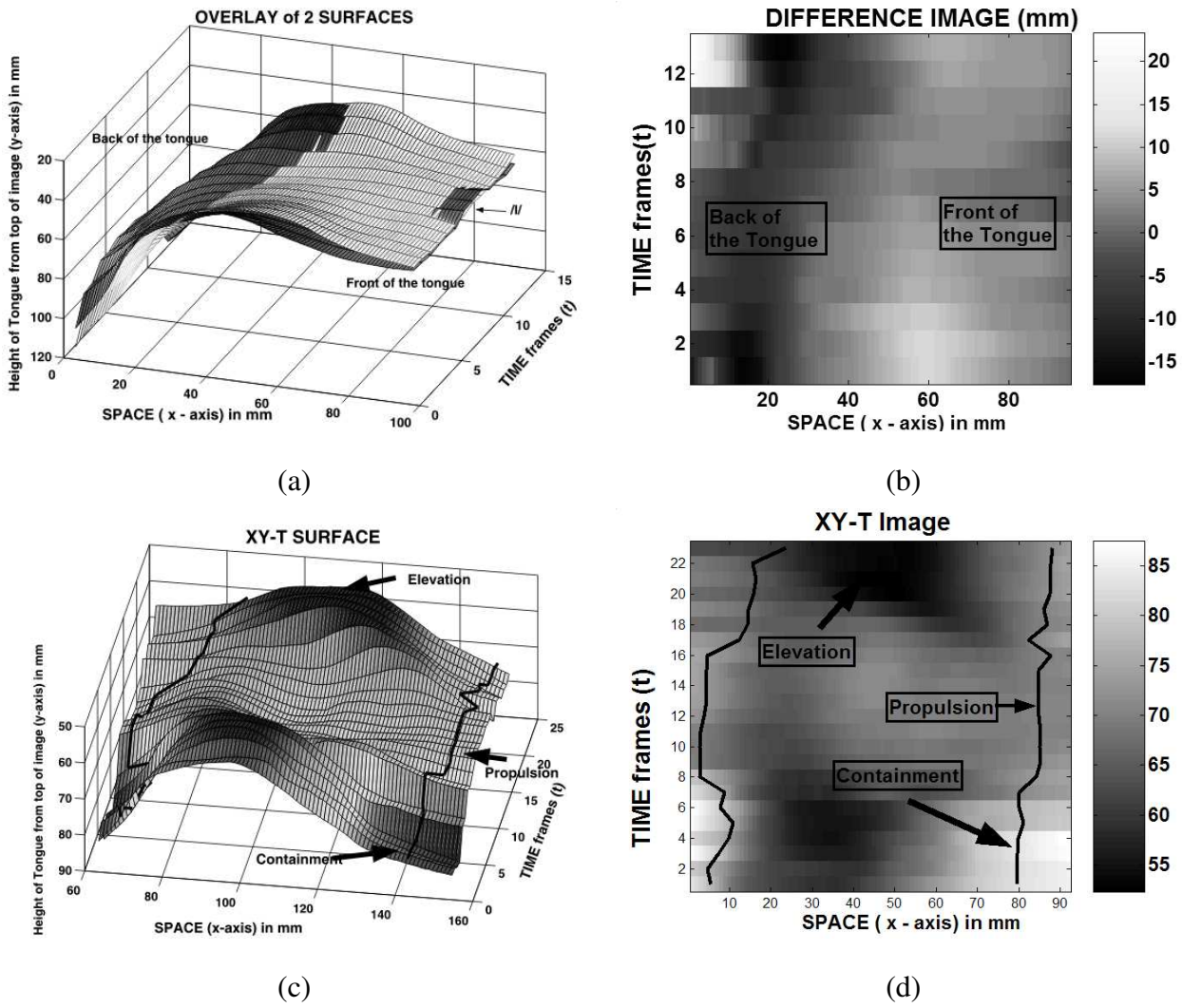


Figure 4: Applications of SURFACES: (a) Overlay of 2 spatiotemporal tongue surfaces of the word 'golly'. Mesh surface was spoken in 'upright' position and the filled surface in 'supine' surface. (b) Difference between the two spatiotemporal surfaces in form of an image. (c) Spatiotemporal surface of a 20cc swallow. The water is contained in front of tongue, then propelled backwards, followed by tongue surface elevation after the water's passage. (d) Visualization of the swallow as an image with gray scale denoting height. Note in (c) and (d) the black lines separate the regions that contain real data from the regions that contain extrapolated data.

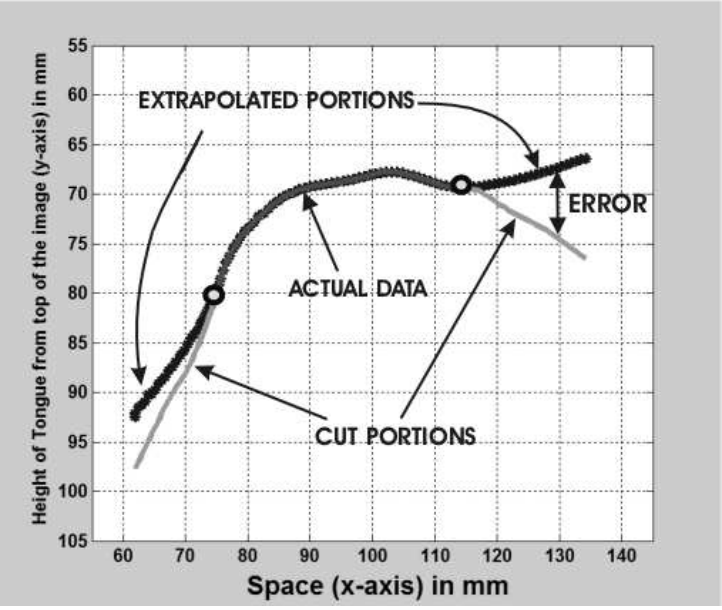


Figure 5: Validation experiment: The center line along with the lower lines (gray) on either side is the actual tongue contour. The edges are artificially cut in order to see how Kriging performs in extrapolation. The last data points on the edge of the contour are represented as black circles. The extrapolated lines (black) are the contours that Kriging estimated. The difference between the black and gray regions is measured as error.

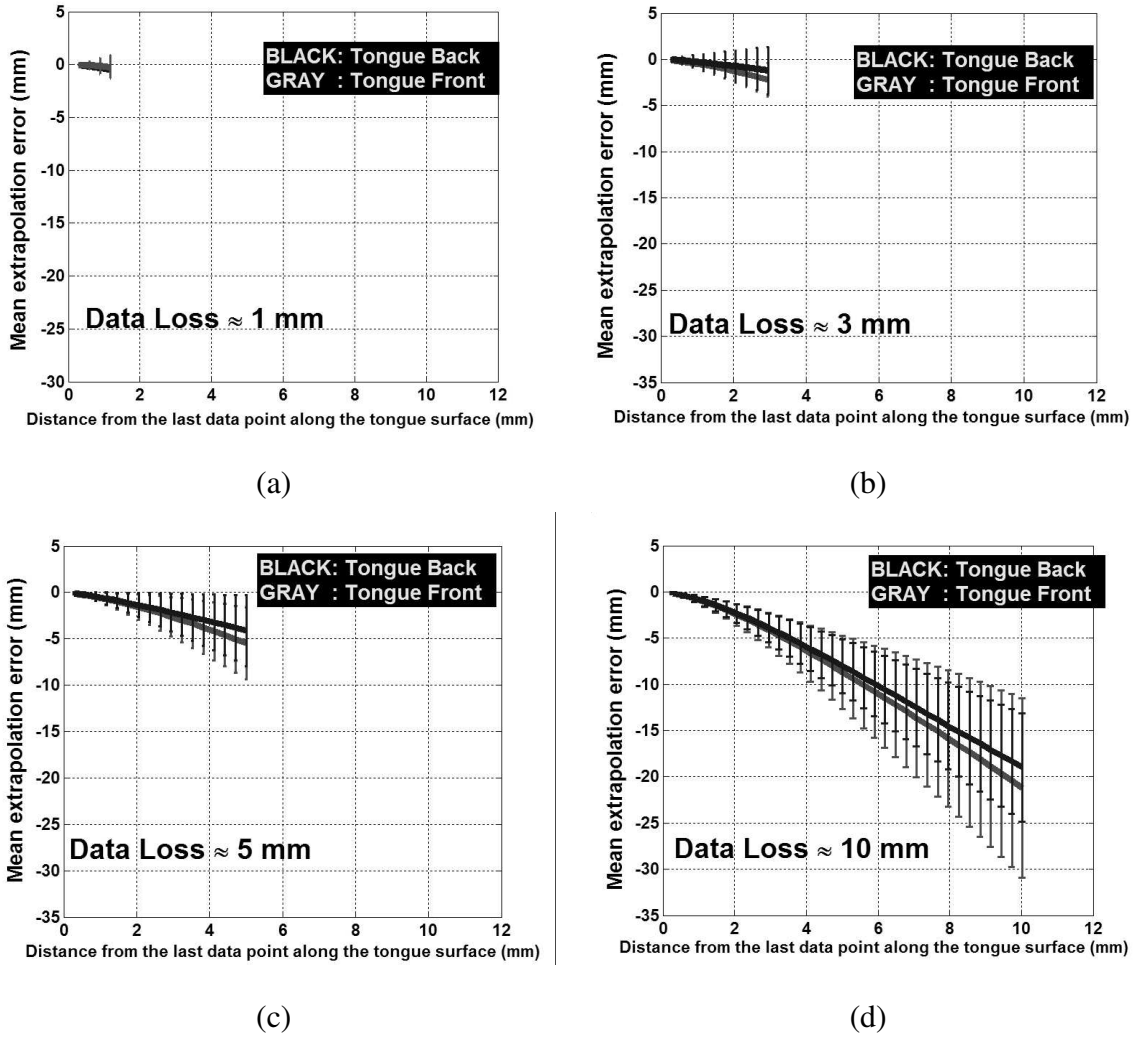


Figure 6: Validation Results: Mean errors in estimation when the length of tongue contour cut is (a) 1 mm (b) 3 mm (c) 5 mm and (d) 10 mm. The error bars represent standard deviation. The x-axis denotes distance along the surface of the tongue, not the distance along the spatial axis (x-axis). The gray curves denote the errors in the front of the tongue and black denotes the errors in the back of the tongue.

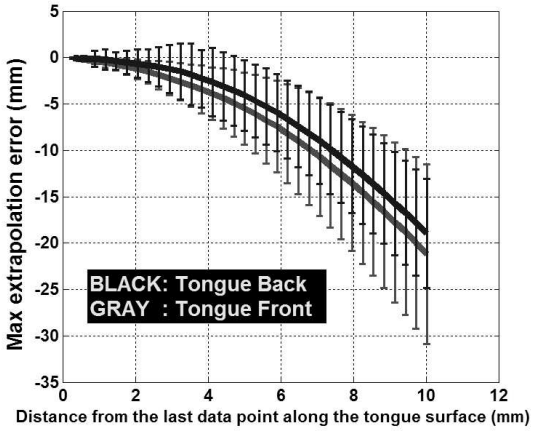


Figure 7: Maximum Errors: Maximum expected errors in estimation as a function of the length of tongue cut.