The tongue stops here: Ultrasound imaging of the palate

Melissa A. Epstein[1], Maureen Stone[1, 2]
[1]Department of Biomedical Sciences, Room 5A12
[2]Department of Orthodontics, Room 3E08
University of Maryland Dental School
666 W. Baltimore Street
Baltimore, Maryland 21201

Running title: Ultrasound imaging of the palate

ABSTRACT

This paper presents a method for imaging the palate and extracting the palate contour from ultrasound images. Ultrasound does not usually capture the palate because the air at the tongue surface reflects the ultrasound beam back to the transducer. However, when the tongue touches the palate during a swallow, the ultrasound beam is transmitted through the soft tissue until it reaches and is reflected by the palate. In combination with tongue contours, the palate contour has the potential for disambiguation of the tongue surface, registration of images within and across subjects and for calculation of phonetically important measures.

PACS number: 43.70.Jt

I. INTRODUCTION

Ultrasound has been used to image the surface of the tongue during speech and swallowing for more than 20 years, beginning with Sonies *et al.* (1981). Ultrasound provides real-time images of the tongue surface as a video sequence. Data are easy to collect and the procedure is non-invasive making the methodology suitable for use with any subject, including patients and children. One of ultrasound's limitations is that only the upper tongue surface appears in the image. Information on other vocal tract structures, such as the palate, would enhance interpretation of the tongue data by providing a vocal tract reference. This paper presents a method for using ultrasound to collect palate contours for use as a reference for tongue movements during speech and swallowing.

Once palatal contour information is available, it allows important applications of ultrasound imaging which are not otherwise possible. For example, Wrench and Scobbie (2003) and Mielke *et al*. (2005) use a palate tracing to better describe subjects' articulations. The present paper addresses three applications. First, the palate contour can be overlaid on raw tongue images to disambiguate the location of the tongue surface. Second, by providing a reference within headspace, the palate contour may be used for within and across subject registration of ultrasound images. Finally, in combination with tongue contours, the palate contour allows the computation of phonetically important measures, such as the location and degree of constrictions in the oral cavity.

II. BACKGROUND

When imaging the tongue the ultrasound transducer is placed beneath the chin. The ultrasound beam travels upward through the tongue body and reflects back from the upper surface because the air has a different acoustic impedance than the tissue. This reflection produces a white line on the ultrasound image. Weaker echoes occur between boundaries of similar densities such as tissue-to-tissue or tissue-to-water. Tissue-to-tissue interfaces occur between muscles, fat and connective tissue within the tongue. A tissue-to-water interface occurs between the tongue surface and the bolus during a swallow. When either the tongue or a bolus of liquid makes contact with the hard palate or velum the ultrasound beam reflects off the palate (see Fig. 1). The beam reflects off the near side of the bone, or, in the case of the velum, the air on the nasal side of the soft tissue. Ultrasound images of the tongue surface and palate are limited anteriorly and posteriorly by “acoustic shadows” (black regions) created by refraction of the ultrasound beam off the mandible and hyoid bones (see Fig. 1).

The palate is normally not seen on an ultrasound image because the ultrasound beam reflects off the air in the vocal tract and never reaches the palate. However, during a swallow the tongue makes full contact with the palate as the bolus is propelled backward. The palate contour may be observed during either dry (saliva) or wet (liquid bolus) swallows. Wet swallows allow the sound to pass through the water and reflect from the palatine bone. However, wet swallows may introduce an artifact because the ultrasound beam can reflect off air ingested with the bolus of water. Since the air stays above the water in the mouth, a bright reflection occurs and can be mistaken for the palate. Fig. 2 shows an ultrasound video frame with a reflection off a bolus of water that could be mistaken for the palate.

Data from 2 subjects executing 3 to 5 different kinds of command swallows (e.g., dry, wet, soda, different size boluses) demonstrate that palatal images can be collected from all types of swallows (Epstein *et al.*, 2004). Variability among measured palates is fairly small. There is a 4 mm maximum difference between palate traces for Subject 1 across 10 swallows; there is a 2 mm maximum difference for Subject 2 across 3 swallows. Measurement and instrumental error account for up to 1 mm. The rest of the variability may be due to air trapped in the bolus.

Anecdotally, palate images are sometimes clearer with wet rather than dry swallows. Consequently, it is useful to collect several types of command swallows and measure the clearest one. In addition, continuous swallowing is especially good because after several swallows the initial air has been swallowed and only water boluses remain. There is also anecdotal evidence that drinking a glass of water before the experiment brightens the ultrasound image of the tongue, possibly because of greater hydration of the tissue (Archangeli, p.c.). Encouraging subjects to drink water *during* the experiment may also brighten the image. Although it is possible for sound to pass through to the palate by having the subject "press" his/her tongue against the entire palate at once, subjects often have difficulty doing this well, and full contact cannot be guaranteed especially for the velum.

Features of the palate image may vary during the swallow and the palate contour may need to be extracted from multiple frames over time as the bolus passes through the vocal tract. The data points for the contour are collected from the front to the back, corresponding to the passage of the bolus through the tract. Initial frames show good alveolar edges and later frames good vault and velar edges. To increase accuracy of intra- and inter-subject registration, particular attention should be paid to tracing of palatal landmarks. Subjects tend to have a bend between the rugae and the vault (this is also seen on the dental cast). If no bend is seen on the palatal contour, the tongue – not the palate – may have been measured and careful inspection of the images may show a frame in which the bottom of the white line is curved, and the top bent (see Fig. 3). This line contains both tongue and palate data. The lower edge of the palate (between the two lines) should be extracted if at all possible. The bend at the velar/palatal junction can also be used as a landmark by observing frames with velar lowering.

Ultrasound images of the hard palate and velum may not be directly comparable by visual inspection to a digital midsagittal dental cast representation of the palate for some subjects for several reasons. First, the amount of mucosa covering the palatine bone is thicker posteriorly than anteriorly (see Fig. 4). Ultrasound represents the palatine bone, whereas the dental cast represents the mucosa. Anteriorly, the mucosa is quite thin, 1 mm or less. Posteriorly, the mucosa is thicker and there is a larger distance between the bone and soft tissue surface. Second, the velum is minimally captured, if at all, on a dental cast to prevent a gag reflex; however, the velum is visible on the ultrasound image. Thus the ultrasound palate extends farther back than the dental cast. In addition, the velum may be more than 10 mm in thickness (Kuehn and Kahane 1990; Kuehn, p.c.). On the dental cast the oral side of the velum is captured, and the velum will be in an open position to accomodate the subject’s breathing. On the ultrasound image the velar reflection is usually the nasal side since the sound passes through the tissue and reflects off the air in the nasal cavity. Moreover, the velum is closed at certain times during the swallow. Thus, the cast reflects the more rounded surface of the unevenly thick mucosa and, if captured, a lowered velum. The ultrasound image reflects the more level palatine bone and the nasal side of the (possibly raised) velum. A final difference occurs because the anterior portion of the hard palate, including the alveolar ridge, is often not visible in the ultrasound palate, due to a posterior transducer angle or to obscuration by the acoustic shadow cast by the jaw or by air underneath the tongue. Therefore the ultrasound palate will likely be shorter anteriorly due to the jaw shadow, and longer posteriorly due to velar imaging, when compared to the dental cast. However, not all subjects show all the above described differences.

Three types of tongue-palate relationships can occur when comparing a palate contour with tongue contours of certain speech sounds, such as high front stops (e.g., /t,d/) and high back stops (e.g., /k,g/). First, the tongue contour may appear to merge with the palate contour. This is because when the tongue touches the palate the ultrasound beam travels through the tongue and reflects off the palate and not the surface of the tongue. Second, the tongue contour may lie above the palate contour. Then it is possible that at least part of the palate contour is an artifact (e.g., air in the bolus of water). Finally, the elevated tongue may not appear to touch the palate. For the back consonants this may be due to variations in the thickness of the palatal mucosa and velum. For front consonants the palate bone may be several millimeters higher than the mucosal palate or the point of contact for front consonants may be obscured by the shadow of the jaw and the air beneath the tongue. An additional confound exists for rapid motions, like stop consonants; they tend to be undersampled at the ultrasound frame rate of 30 Hz. Therefore, it is possible that the maximal frame imaged is not actually the maximal tongue position for the stop.

III. APPLICATIONS

To demonstrate the use of palates in ultrasound-based speech research, three applications are described based on palates and speech articulations collected from twelve archived data sets (5 females and 7 males; ages range from 16 to 34 years). Due to the archival nature of the data, only dry swallows are available for analysis. The same ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, Model AI5200S) is used for all 12 subjects. The transducer for all subjects is a 2.0-4.0 MHz multifrequency convex-curved linear array transducer that produces 30, 90 degree wedge shaped scans per second. Focal depth is 10 cm. Ultrasound and audio data are collected while subjects are seated in a dental chair positioned supine, then upright. Subjects wear a cervical collar with ultrasound transducer attached to restrain head motion and stabilize the transducer. Unless otherwise specified, the data discussed here are from the upright position. Palates are collected from either the frame with the most anterior hyoid bone shadow or reconstructed based on palatal landmarks (e.g. the palatal bend posterior to the rugae) over several frames of the swallow. For validation of the extracted palate contours for Subjects 1-5, word-initial /d/ and /g/ are collected from the words “dash” (“dack” for Subject 2) and “golly” to locate the palates relative to the tongue contours. Tongue contours for /d/ and /g/ are extracted from the frame where the tongue reaches the maximum constriction height before the release burst seen in the simultaneously collected acoustic signal. Tongue and palate contours are measured using the Maryland Tongue Analysis Package (MTAP), consisting of EdgeTrak, a semi-automatic system for the extraction and tracking of tongue contours (Li, *et al.*, to appear), and Surfaces, a contour-sequence display and analysis program (Parthasarathy, *et al*., to appear). Palate and tongue contours for Subjects 1-5 are displayed in Fig. 5.

The first application is the use of the palate as a reference for tongue contours. Fig. 5 displays tongue and palate contours for Subjects 1-5. Note how the palate contour disambiguates the location for the constriction for the /g/ tongue contour. Furthermore, the posterior edge of the palate contour provides a landmark for the division between the oral and pharyngeal regions of the tongue contour.

The second application is using the palate for registration across subjects or sessions. Data collected from two subjects or sessions may differ in transducer angle. Imposing the palate contour on the tongue data makes this rotation clear. For example, in Fig. 5, the palate for Subject 5 appears rotated further forward compared to the other subjects, and rotated forward. This indicates that for Subject 5 the ultrasound transducer is rotated backward more than the other subjects when the data was collected. This is because the ultrasound video display always positions the transducer as if it were pointed upward; a backward transducer rotation when transformed in this way rotates all the imaged data anteriorly. To determine what registration is needed to overlay two data sets, landmarks on palatal contours may be aligned. For a single subject's data, two landmarks can be used: the intersection of the velum with the hard palate (in a frame where the velum is lowered); and the angle between the rugae and the vault. Across subjects, caution should be exercised due to different sizes of the palate and differences in vault angle.

Once these landmarks are identified, they can be used to spatially align two palates by rotating and translating the subject's palate at time 2 to overlay the palate with time 1. The alignment parameters are then applied to the tongue data at time 2 to align the two tongue data sets. The following describes the results of a study aligning palate traces from Subjects 1-12 in upright and supine position. The rugae landmark is identified in both palate traces. Palate 2 (supine) is then rotated and translated to overlay palate 1 (upright). The alignment parameters are then applied to the tongue data at time 2 to align the two tongue data sets. Results of the alignment indicate that of the 12 palate pairs, eight differed in anterior-posterior length. Four of these 8 have virtually identical palatal traces (maximal difference of less than 1 mm) and 4 differ only in velar angle. Three additional subjects have a maximal difference between 2-5 mm, and one has two entirely different shapes. There is only one collected swallow per subject, therefore, the nature of the error for this last subject could not be determined or rectified.

The final application discussed here is to use the palate contour for the computation of phonetically important measures, such as the constriction degree and location. For example, Fig. 6 shows local differences between the palate and the tongue contours for Subject 1. The local differences are calculated by first cutting the contour and the palate to the same length in the x direction. The contours are then resampled to have 50 points and a nearest neighbor difference is calculated for each point. Finally, the contours are divided into equal length segments (in this example 2 segments – anterior and posterior – are used) and average nearest-neighbor differences are calculated between the palate and the tongue contour for each segment. As can be seen in the figure, these differences indicate that for /d/ the palate-tongue distance decreases from back to front and for /g/ the distance increases from back to front. Without a palate contour, it would only be possible to make comparisons in location directly between /d/ and /g/ and none in absolute vocal tract constrictions. The advantages of comparing tongue contours to a single reference palate increase when analyzing a large data set (e.g. several phonemes or an entire sentence), the alternative being comparing each phoneme to all the others or to an arbitrary reference phoneme or rest position.

IV. CONCLUSIONS

The inclusion of a representation of the hard and soft palate in an ultrasound image has the potential to increase our ability to interpret and assess tongue contours. Although there are some limitations in the collection and interpretation of ultrasound images of the palate, palate contours are useful for disambiguation of the tongue surface, across and within subject registration and for the calculation of phonetically interesting measurements in the oral cavity. The addition of palate information facilitates the reduction and enhances interpretation of the data by providing an oral cavity reference, subject and session registration, and a second vocal tract structure for quantitative analyses. With the inclusion of the palate, ultrasound imaging can present the tongue in its proper position in the vocal tract.

ACKNOWLEDGMENTS

This research was supported in part by NIDCD/NIH Grant RO1-DC01758 and by NIH Grant T32-DE07309. The authors would like to thank Greg Stock, Marianne Pouplier, Kevin Bunin, and Andy Browne for their assistance in data collection and analysis.

REFERENCES

Epstein, M. A., Stone, M., Pouplier, M. and Parthasarathy, V. (**2004**). "Obtaining a palatal trace for ultrasound images," J. Acoust. Soc. Am. **115**, 2631 (Abstract).

Kuehn, D. P. and Kahane, J. C. (**1990**). "Histologic study of the normal human adult soft palate," Cleft Palate J., **27**, 26-34.

Li, M., Kambhamettu, C. and Stone, M. (**to appear**). "Tongue motion averaging from contour sequences," Clin. Linguist. Phon.

Mielke, J., Baker, A., Archangeli, D., Richtsmeier, P. and Racy, S. "Aligning tongues and palates with Palatron," Ultrafest III, Tucson, Arizona, 14-16 April, 2005. Available at http://dingo.sbs.arizona.edu/~apilab/ultrafest/Palatron.pdf

Parthasarathy, V., Stone, M. and Prince, J. L. (**to appear**). "Spatiotemporal visualization of the tongue surface using ultrasound and kriging (SURFACES)," Clin. Linguist. Phon.

Sonies, B., Shawker, T., Hall, T., Gerber, L., and Leighton, S. (**1981)**. ‘‘Ultrasonic visualization of tongue motion during speech,’’ J. Acoust. Soc. Am. **70**, 683–686.

Wrench, A.A. and Scobbie, J. M. (**2003**). “Categorising vocalisation of English /l/ using EPG, EMA and ultrasound,” Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia, 7-10 July, 2003, 314-319. Available online at http://sls.qmuc.ac.uk/RESEARCH/Ultrasound/child_home_pages/wrench_scobbie.pdf

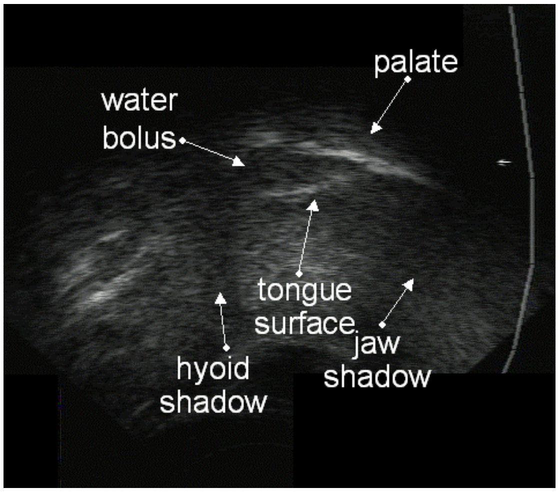

Figure 1. Ultrasound image of the palate and tongue surface during a 3cc water swallow. Both the palate and the tongue are visible because of the partial transmission/reflection of the ultrasound beam by the bolus of water. The shadow of the jaw bone obscures the most anterior portion of the tongue.

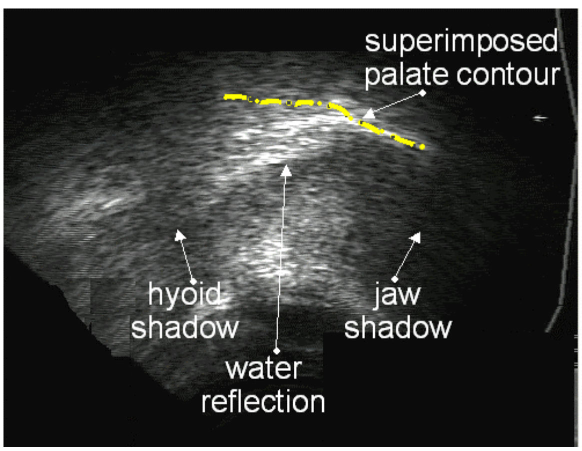

Figure 2. Reflection of a bolus of water. A palate contour from a dry swallow is superimposed on the ultrasound image (small black and white dots). Note how the water bolus reflection is below the palate contour along the posterior edge.

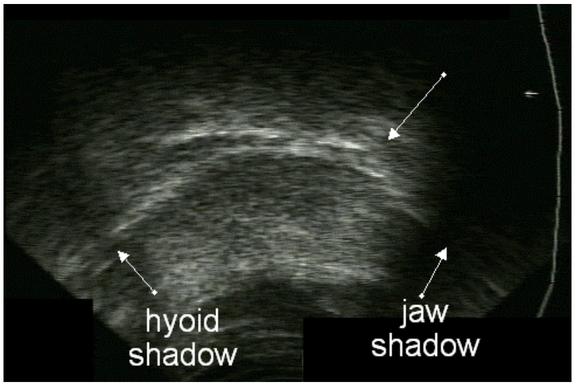

Figure 3. Ultrasound image illustrating bend between rugae and vault of the palate for Subject 2 during a dry swallow. In this frame the tongue contour is curved and the palate contour has an angular bend (marked by white arrow). The lower edge of the palate contour should be extracted.

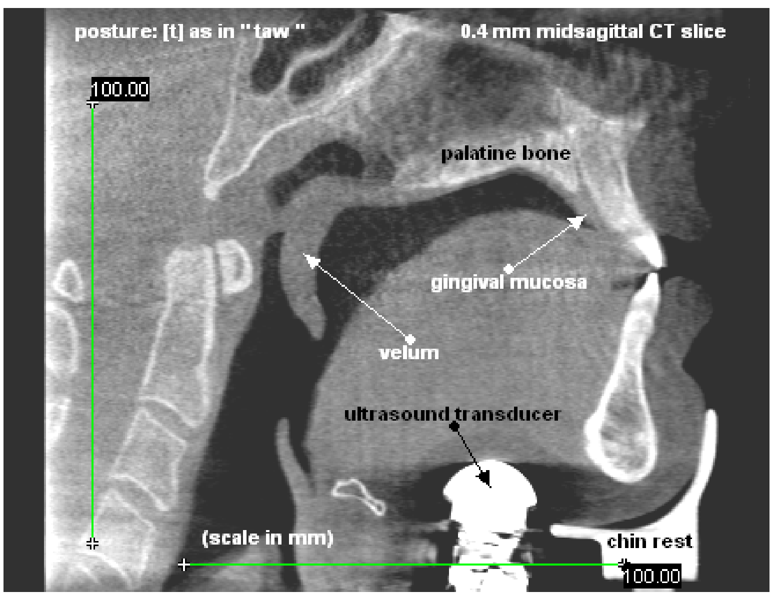

Figure 4. 0.4 mm midsagittal CT slice (iCAT, Imaging Sciences International) of subject sustaining /t/. Note how the palatal mucosa thickens from anterior to posterior and the thickness of the velum. Figure courtesy of Ian Wilson, University of British Columbia.

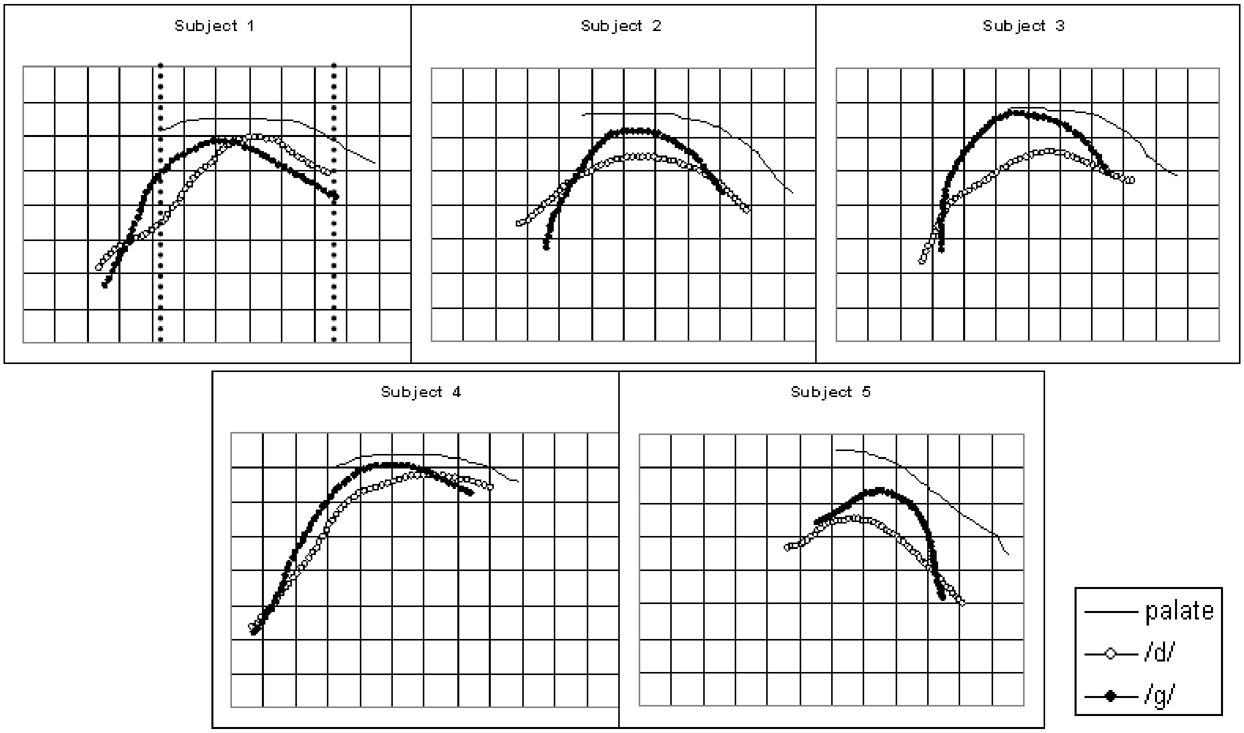

Figure 5. Tongue and palate contours for five subjects (centimeter scale). Tongue tips are to the right.

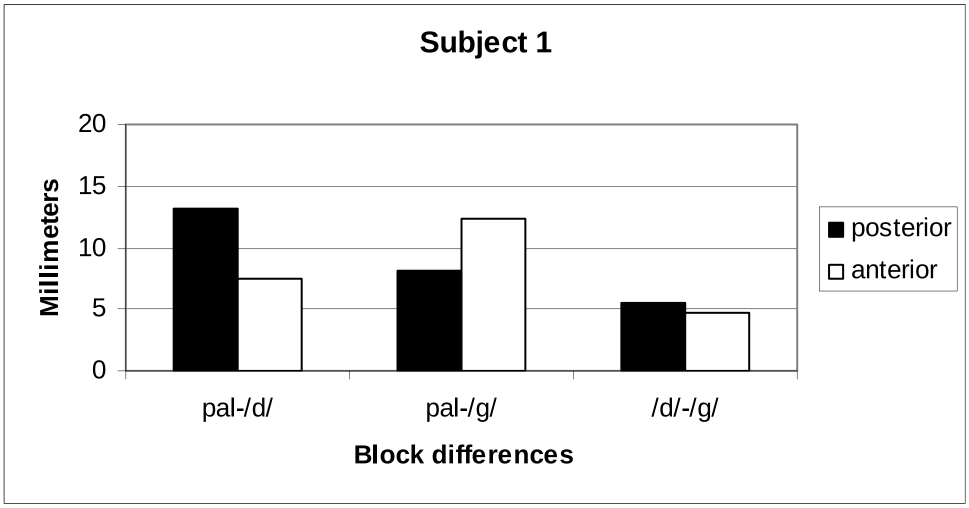

Figure 6. Block differences between the palate and the tongue contours or between the two tongue contours for Subject 1. The contours from Fig. 5 are segmented between the vertical lines. The contours are divided into 2 equal length segments (anterior and posterior) and average nearest-neighbor differences are calculated between the contours for each segment.