Principal Component Analysis of Internal Tongue Motion in Normal and Glossectomy Patients with Primary Closure and Free Flap.

Maureen Stone[1,2], Xiaofeng Liu[5], Jiachen Zhuo[4], Rao Gullapalli[4], Andrew Salama[3], and Jerry L. Prince[6]

Departments of [1]Neural and Pain Sciences, [2]Orthodontics,[3]Oral and Maxillofacial Surgery
University of Maryland Dental School, Baltimore, MD, USA.

[4]Department of Diagnostic Radiology,
University of Maryland School of Medicine, Baltimore, MD USA

[5]Departments of Computer Science, [6]Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, MD, USA

This study examines the motion patterns of the midsagittal section of the tongue during elevation of the tongue body. The motion from /u/ to /k/ is a very small one. Therefore, glossectomy patients who can make a /u/ should be able to make a /k/. However, because of the altered anatomy it appears that their execution of this gesture differs from the normal speaker due to scar accommodation and motion of the flap. The word "a souk" or “disouk” was spoken by six normal speakers, and five post-glossectomy patients with T1 or T2 tumors, N0M0; four with primary closure and one with radial forearm free flap. All patients produced a normal sounding /k/. The goal of this work was to quantitatively characterize any motion differences between the normal and patient speakers.

INTRODUCTION

After tongue cancer surgery speech quality is primarily affected by changes in the motion of the post surgical tongue. One of the main factors that determine the effectiveness of tongue motion is the size of the residual tongue left after removal of the tumor (cf. Matsui, et al., 2007). For fairly small tumors, T1 and T2, there are some variations in surgical closure procedure that may also affect tongue motion. Two of these are: primary closure, in which the tissue is sutured closed; and free flap reconstruction, in which distal tissue is added to the tongue to increase bulk (cf. Urken et al, 1994). Reduced muscle mass after surgery results in a reduction of total muscle force, the need for additional supporting muscles, and possibly reduced closure area. Flap tissue adds dead weight whose positioning will further tax the remaining muscle tissue, but may provide mass for improved vocal tract shaping. Scar tissue adds regions of rigidity that need to be incorporated into the motion. These effects may reduce tongue range of motion, closure accuracy, and speed of motion all contributing to unclear speech.

Research results vary as to whether one of these methods produces better speech quality. Hsiao et al., 2002, found that speech intelligibility was better after primary closures than flaps, whereas Terai and Shimahara (2000) found the opposite. Nicoletti, et al, (2004) found no difference in speech between flaps and primary closures. Even in cases where speech is unaffected, the differences in these two closure procedures are likely to create different motion patterns.

In order to quantify motion patterns of post-glossectomy patients and determine how and where they differ from normal it is necessary to have a normal database of motions, a metric that represents the motions, and a statistical method to quantify the patients' motions relative to the database. One metric that represents internal tongue motion is the velocity vector field of the mid-sagittal tongue. This field details the motion pattern of each point in slice at any moment during speech. In the present paper we examined the velocity field at the onset of motion from /u/ to /k/ for 5 patients. Although these patients produced perceptually normal /u/ and /k/, the /k/ onset motions were unusual for several patients. To detail and quantify the patients' motions relative to the normal speakers, a Principal Components Analysis (PCA) was performed to extract the underlying components of the normal tongue motion patterns and determine how well these components represented each patient's motion.

Principal Components Analysis (PCA) is an excellent method to extract and represent patterns in high-dimensional data for which no expectations or models are available. PCA has been applied to image analysis in many different contexts in the past. Assisting in object recognition using active appearance models (Cootes et al, 1998), predicting "average images" in a database (Moghaddam and Pentland, 1997), and retrieving dominant modes for fast imaging (Zientara et al, 1994), are just a few examples. PCA has also been applied in images representing deformation or motion; for example, they have been used to build a normal atlas of cardiac motion (Chandrashekara et al, 2003) and to study shape variations in the normal brain (Xue et al, 2006).

Principal components analysis (also factor analysis) has been very successful as well at representing the tongue surface shape while reducing its dimensionality. One of the first such studies was of American English vowels, Harshman et al, 1977 found that two main shape dimensions, front-raising and back-raising explained most of the variance in vowel tongue shapes seen on midsagittal X-rays. Front-raising is the elevation of the anterior tongue and fronting of the posterior tongue. Back-raising is the arching of the tongue body toward the velum. PCA studies in other languages have produced similar PC's suggesting that this is a basic principle of midline tongue shaping (Jackson, 1988 and Hoole et al., 1999). We studied the coronal tongue surface shape using PCA (Stone et al., 1997, Slud et al., 2002) and found that two PC's plus a mean y-level accurately represented coronal shapes for all 11 American English vowels, capturing vowel height, shape, asymmetry, and consonant context.

Bressman et al. (2005, 2007) used PCA to study glossectomy patients, with and without a flap. Using ultrasound they imaged 3D tongue surfaces during 9 sustained speech sounds (2 sec capture time) in 12 controls, and 12 glossectomy subjects pre- and post surgery. All patients before surgery had surface shapes similar to normal and required two PC's to explain their shapes. After surgery the no-flap patients were still similar to normal, but the flap patients needed 3 PC's to explain their data. The third PC appeared on the operated side and added left-to-right asymmetry to the surface as well as reducing the midline tongue groove (concavity). Measures of concavity and asymmetry were not correlated to each other, and interestingly the key correlate with speech acceptability was tongue concavity not asymmetry.

METHODS

Subjects

Subjects were five normal controls (4 male, 1 female) and 5 glossectomy patients (4 male, 1 female). The patients all had surgery at least one year prior to recording to remove a T1N0M0 tumor in the medial third of the lateral tongue. Primary closure was performed on 3 males and 1 female patient, and a radial forearm free flap reconstruction was done on 1 male. Pt. 1 received the flap; NL 2 and Pt. 2 were females. All subjects were native American English (AE) speakers.

Speech Materials

The subjects all repeated the word /əsuk/, except NL 5, who repeated the utterance /disuk/. These pseudo-words were chosen for several reasons. They are within the MRI repeat time of 2 sec per utterance, which includes a breath. They maximize tongue deformation by engaging the jaw very little. They cover a large range of AE positions and shapes (cf. Stone and Lundberg, 1996), and they contain a range of difficulties for glossectomies. The word was changed to begin with a neutral vowel after the first two subjects, in order to approximate as well as possible the resting tongue during hMRI and DTI for comparative analyses (not presented here). Since the latter portions of the words were common across subjects, PCA's were viable for velocity fields at the onset of motion into the /k/. This task was of interest because all the patients produced perceptually normal /k/ sounds, allowing us to assess how similar their tongue motions were to normal.

Data Collection Procedures

Four types of MRI data sets were recorded in the same session using a head and neck coil: Diffusion Tensor Imaging (DTI), high-resolution MR images (hMRI), Cine-MRI, and tagged cine-MR images (tMRI). Only the latter two will be presented here. Both the tMRI and the Cine-MRI data sets were collected with a 6 mm slice thickness and had an in-plane resolution of 1.875mm/pixel resolution. To acquire each tagged cine series, (two CSPAMM image sequences in two tag directions) the subject repeated each speech task 16 times per slice resulting in 80 to 100 repetitions including four pauses. The non-tagged cine-MRI images were used to register the data sets across subjects prior to the PCA.

The UMD MRI facilities now have an MRI trigger system that uses acoustic cues to synchronize speech utterance repetitions with MRI acquisition. The protocol for synchronized auditory cueing is based on the method of Masaki and colleagues (Masaki et al.1999, Shimada et al., 2002). In this method, the audio system of the MRI console delivers short white-noise pulses through headphones at predetermined intervals to the subject, and triggers the MRI acquisition synchronously. The subject utters syllables and breathes in time with the pulses. A 15-minute training protocol, with feedback from the experimenter, was developed using nine normal subjects. Because of the training, excellent cine and tagged images are now obtained for naive subjects and patients even with long repetition sets (Stone and Murano, 2007, Stone et al, 2008). Recording time can take up to 1 hour and 15 minutes.

Data Reduction

The motion of the internal tongue tissue was extracted from the tMRI data using Harmonic Phase Analysis (HARP) (Osman, et al., 1999, Parthasarathy, et al. 2007). HARP extracted and tracked the tags in the words, and visualized speaker specific tongue motion patterns. All analyses presented here are in the midsagittal plane, and visual inspection of the velocity patterns was used to determine the frame of initial motion into /k/. The criterion was a change in direction from the /u/ motion, or the onset of motion after a pause for /u/.

The data were normalized across subjects using 9 tissue points on the surface of the tongue, as shown in Figure 1. They include: the base of the valleculae, the upper tip of the epiglottis (projected

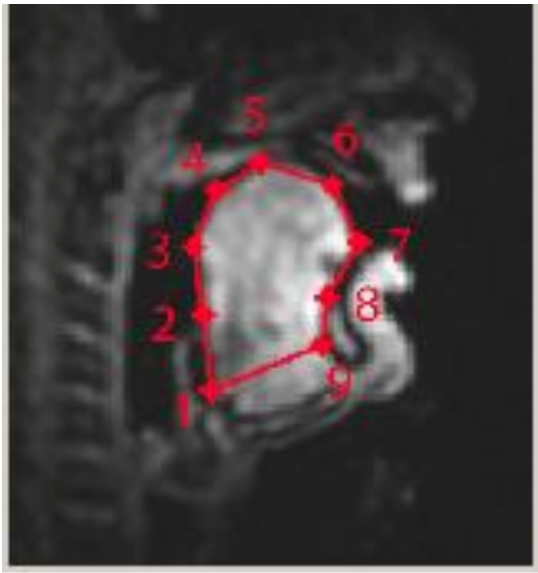

Figure 1. Nine landmark points were used to align all the tongues.

onto the tongue surface), the point on the tongue surface that lies between the elbow of the velum and the lower edge of the mandible, the mid palate, the tongue tip, the origin of genioglossus, and several additional points equidistant between these landmarks. These points were extracted, for each subject, from the Cine-MRI image that corresponded to the velocity frames used in the PCA. These points were registered across subjects and the rest of the voxels in the image were then repositioned and interpolated accordingly.

Principal Component Analysis

The PCA was performed on the normal subjects and quantified the component motions of the midsagittal velocity patterns for each task. For each patient we determined the how well the motion patterns were explained by the normal principal components (PC's). The motions represented by each PC would not only highlight unusual patterns, but might provide insight into patient and group difficulties.

The PCA procedure involved: (1) HARP analysis to compute velocity fields; (2) alignment of all 6 data sets (rigid + scalar)using the landmarks; (3) selection of a common tongue region; (4) creation of a velocity field data vector for each subject (lexicographic ordering); (5) singular value analysis of sample covariance matrix. The largest to smallest singular values correspond to eigenvalues and eigenvectors that represent the most dominant to the weakest PCs.

RESULTS

/k/-Onset Motion in the Normal Database.

A large amount of variability was seen in the 5 normal subjects at the onset of /k/ (see Figure 2), indicating that there were a number of onset strategies. Subjects 1 and 2 elevated the tongue body straight upward, and the tongue root anteriorly. Subject 3 had similar motion but with greater root movement, less body elevation, and tip retraction. Subject 4 added a backward component to the upward motion of the body, and subject 5 moved primarily backward. Thus the /k/ onset differed by whether the gesture began with an upward or backward motion, and that depended on where their /u/ was positioned.

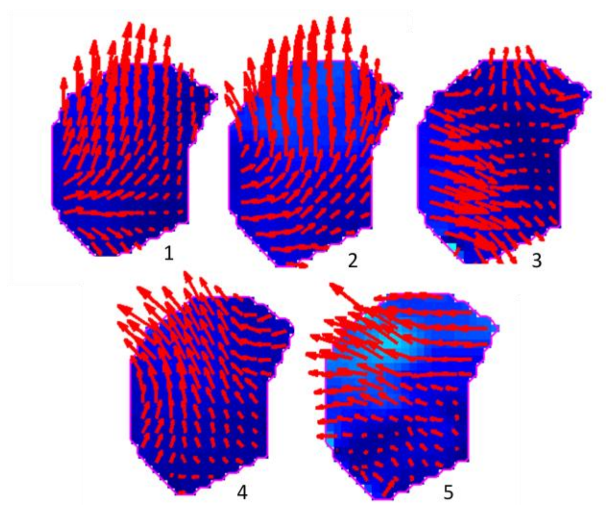

Figure 2 Velocity fields at the /k/-onset motion for 5 normal subjects.

The raw PC shapes can be seen in Figure 3. PC 1 represents the variability in the location of the upward motion. PC 2

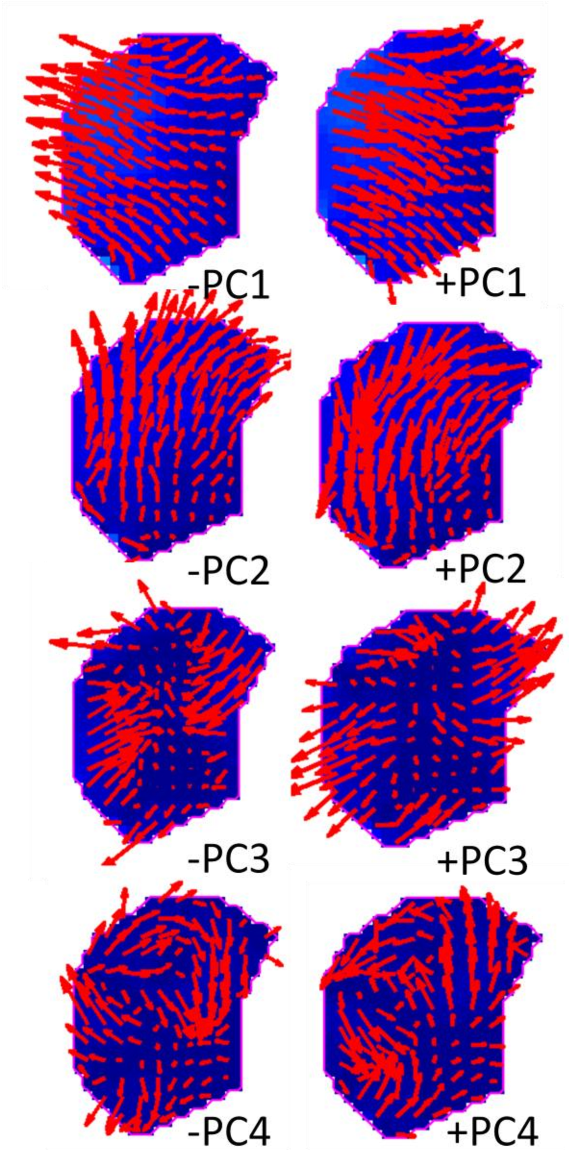

Figure 3. The PC shapes.

represents lesser or greater upward velocity and a curved motion trajectory. PC3 represents additional lengthwise expansion/compression, and PC4 represents additional sagittal rotation. Eigenvalues for the first 4 PC's were 372.24, 157.05, 39.27 and 11.03 respectively, indicating that PC's 1 and 2 accounted for most of the variance.

The effect of adding PC's 1 and 2 to the mean velocity vectors is shown in Figure 4. Panel 5 is the mean velocity, which is flanked by the mean + or - 1 standard deviation for PC1 (horizontal) and PC2 (vertical). The four corners sum PC1 and 2. Panel 5 indicates that the mean tongue body motion is upward and backward, with inward compression at the tip and root . The mean plus positive loading on PC 1 (panel 6) places the primary pattern of motion as straight upward, and inward at root and tip. The mean plus negative loading on PC 1 (panel 4) places the primary direction of motion as up and back. In addition, PC1 distinguished which portion of the tongue was in motion. Positive loading indicated that a more anterior portion of the tongue was elevating. Thus, the motion for this group was primarily upwards, with the greatest difference between subjects being whether the motion was straight up or up and back. The effect on the mean of positive loading on PC 2 (panel 2) results

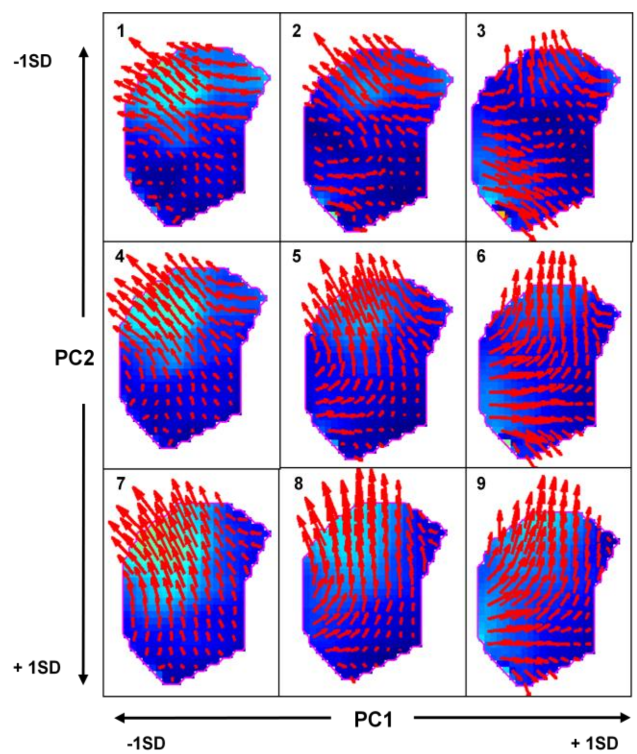

Figure 4. The mean motion pattern (panel 5) and modifications due to addition of PC1 and PC2.

in less upward motion and a small rotation in the posterior tongue; negative PC2 loading meant straight upward motion and no inward compression in the upper posterior tongue (panel 8).

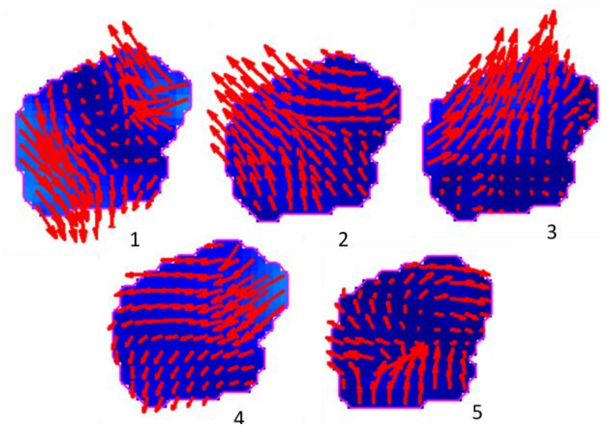

Figure 5. Velocity fields at the /k/-onset motion for 5 glossectomy patients.

/k/-Onset Motion in the Patient Samples.

Figure 5 shows the velocity vector fields at /k/-onset for the 5 patients. Three patients elevated the tongue body, upward and/or backward into /k/. Pt 1 began the motion with upward/backward rotation of the anterior tongue and downward rotation in the posterior tongue. PC1 represented 61% and PC2 26% of the motion. PC's 3 and 4 were less than 1% each. Pt. 2 moved the tongue straight backward and upward, so 88% of her variance was accounted for by PC1. The other 3 PC's explained less than 1% each of her variance. Pt. 3 also was fairly well explained by PC1 (44%) and PC2 (33%). His tongue moved upward and forward. Pt. 4's motion was explained almost entirely by PC2 which accounted for 70% of the variance, while PC1 explained 13%. PC 4 also explained 5%. This is because Pt. 4 began the gesture with backward tongue motion and downward tip motion creating a downward/backward rotation in the anterior tongue. Patient 5 had 7% of his variance explained by PC1, none by PC2, 44% explained by PC3 and 5% by PC4. His tongue moved upward in the back and forward in the front, rotating the tongue body up and forward.

DISCUSSION

The /k/-onset data resulted in 4 PC's that quantified the component motions of the normal subjects. The PC's were then used to explain the /k/-onset motion of the patients. The patients were chosen to determine whether similar tumor conditions and closure procedures would yield similar motion patterns. Patient 1 differed, however, in that he had a flap reconstruction. Three of the patients used motion patterns unique to the data set.

Patient 1 used a circular motion that was unlike the normals or patients seen in this data set, though the pattern has been seen before in other patients and speech tasks (see Figure 5). The motion loaded fairly heavily on PC2 (26%), but the large PC1 (61%) reflected the elevation of the anterior tongue that result from this rotation. Rotation is not unique to flap patients and has been observed in primary closure patients where the scar appears to be the center of rotation. In the case of a flap patient this may not be the case. We cannot identify the location of the most rigid scars, because they form while the flap heals. Moreover, in different speech tasks we have observed rotations occurring about other points for this patient. Therefore, the rotations may be due to rigidity of underlying scar tissue, extra forces used to move the flap, or accommodation to the flap or scars.

Tongue motions for patients 2 and 3 were upward, with a backward or forward angle (Figure 5). These motions were quite similar to the normal subject motions shown in Figure 2 and were heavily represented by PC's 1 and 2. These patients appeared able to recreate the pattern seen in normal speakers despite their lateral tongue resection.

This motion is in contrast to patients 4 and 5. Pt. 4 had an extremely small PC1 loading coupled with a huge loading on PC2 (70%). The tongue essentially moved backward and downward from the /u/ to the /k/. This is unusual as the motion is typically upward. The subsequent motion into the /k/ was mostly backwards in direction. This gesture appeared to be a larger adaptation to the resected tongue than seen in the other patients and may have represented difficulty with tongue body elevation. Patient 5 moved his tongue upward primarily by elevating the floor of the mouth. That is, the jaw muscles apparently contracted to elevate the tongue body. In addition, this method resulted in a salient expansion in the tip and root that was represented best by PC 3 (44%). The upward motion in the floor did not occur in the normal subjects and was not captured by the PC's.

Principal components are not always physically meaningful, because they are statistical entities that decompose data into mathematically optimal quantities. It is interesting and exciting, therefore, that the PC's uncovered by this small database appear to be physiologically meaningful. That is, they are modifications to the mean velocity field that can reasonably be made by the tongue muscles (see Figure 3). Since the PC's can change with different databases, it is possible that this is a coincidence, and that future datasets will result in PC's that are less intuitive. Moreover, the addition of more subjects will allow for additional PC's that may divide the motions more finely, and possibly represent better some of the patient motions. These results are a promising start which might lead to the determination of primitives of tongue motion.

Additional patients will allow us to determine better whether closure type has an effect on the motion patterns, or whether other factors are more important. Additional patients and speech tasks will allow us to determine whether the idiosyncratic motion patterns result in audible speech errors when used in more demanding speech tasks. In the present case, the patients all achieved good stop closure for /k/, and differences in location of the closure were not audible.

CONCLUSIONS

We were able to distinguish between patients whose motion patterns were more or less similar to normal by observing their velocity fields in the onset of the motion from /u/ to /k/. We were also able to describe those motions relative to normal by observing their loadings on the PC's extracted from the PCA of the normal subjects. For stronger results we plan to use a database with a greater number of normal subjects and a larger number of patients with each surgical procedure.

Acknowledgement:

The data used in this paper were collected by Dr. Emi Murano with support from her NIDCD grant # K99DC009279.

REFERENCES

1. Bressmann, T., Ackloo, E., Heng C., and Irish J.C., (2007) Quantitative three-dimensional ultrasound imaging of partially resected tongues Otolaryngology - Head and Neck Surgery Volume 136 (5), 799-805

2. Chandrashekara, R., Rao, A., Sanchez-Ortiz, G.I., Mohiaddin, R.H., and Rueckert, D. (2003) "Construction of a statistical model for cardiac motion analysis using nonrigid image registration," Info Proc Med Imag, Berlin: Springer, pp.599-610

3. Cootes, T., Edwards, G., and Taylor, C. (1998) "Active appearance models," in

Fifth European Conf on Comp Vis., Freiburg, Germany

4. Harshman, R., Ladefoged, P., and Goldstein, L. (1977) Factor analysis of tongue shapes, J Acoust Soc Am 62 pp., 693–713.

5. Hoole, P. (1999) ″On the lingual organization of the German vowel system″. J. Acoust. Soc. Am. 106(2), 1020-1032.

6. Jackson, M. (1988) Phonetic theory and cross-linguistic variation in vowel articulation. UCLA Working Papers Phonet. 71

7. Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (1999) "MRI-based speech production study using a synchronized sampling method." J Acoust Soc Jpn, 20, 375-379.

8. Matsui, Y. Ohno K., Yamashita Y. and Takahashi K. 2007) Factors influencing postoperative speech function of tongue cancer patients following reconstruction with fasciocutaneous/myocutaneous flaps- : a multicenter study. International journal of oral and maxillofacial surgery. 36(7), 601-609.

9. Moghaddam and Pentland. (1997) “Probabilistic visual learning for object recognition,” IEEE Trans Patt Anal Mach Intell, v.19, pp.696-710.

10. Nicoletti, G., Soutar, D.S., Jackson, M.S., Wrench, A.A., Robertson, G., and Robertson, C. (2004) Objective assessment of speech after surgical treatment for oral cancer: experience from 196 selected cases. Plast Reconstr Surg 113, 114-125.

11. Osman, N. F., Kerwin, W. S., McVeigh, E. R., and Prince, J. L. (1999) Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. Magn Res Med, 42,1048–1060.

12. Parthasarathy, V., Prince, J.L., Stone, M., Murano, E., and NessAiver, M. (2007) “Measuring tongue motion from tagged Cine-MRI using harmonic phase (HARP) processing,” Journal of Acoustic Society of America, vol.121, no.1, 491-504

13. Shimada, Y., Fujimoto, I., Takemoto, H., Takano, S., Masaki, S., Honda, K., and Takeo, K. (2002) "[4D-MRI using the synchronized sampling method (SSM)]." Nippon Hoshasen Gijutsu Gakkai Zasshi, 58(12), 1592-1598.

14. Slud, E., Smith, P., Stone, M., and Goldstein, M. (2002) Principal Components Representation of the Two-Dimensional Coronal Tongue Surface. Phonetica vol. 59, Nov. 2-3, pp. 108-133.

15. Stone, M., Goldstein, M., and Zhang, Y. (1997) “Principal component analysis of cross-sectional tongue shapes in vowels,” *Speech Communication*, 22, 173-184.

16. Terai, H., and Shimahara M. (2000) Articulatory function in patients who have undergone glossectomy with use of an artificial graft membrane. (Oral surg. oral med. oral pathol. oral radiol. endo.), 89(5), 560-562

17. Xue, Z., Shen, D., and Davatzikos, C. (2006) “Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping, Medical Image Analysis, v.10, pp.740-751.

18. Zientara (1994) “Dynamically adaptive MRI with encoding by singular value decomposition,” Magn Reson Med, 32(2), 268-274.