A Preliminary Application of Principal Components Analysis to Tongue Motion Patterns

Maureen Stone
Dept of Neural and Pain Sciences
Dept of Orthodontics
U. of Maryland Dental School
650 W. Baltimore St, Baltimore MD 21201
mstone@umaryland.edu

Xiaofeng Liu
Dept of Computer Science
Johns Hopkins University
Baltimore, MD
xfliu@cs.jhu.edu

Hegang Chen
Dept of Epidemiology
U of Maryland Medical School
Baltimore, MD
hchen@epi.umaryland.edu

Jerry L. Prince
Dept of Electrical and Computer Engineering
Johns Hopkins University
Baltimore, MD
prince@jhu.edu

Abstract

In the tongue, complex internal tissue deformations create motion during speech. Motion of tissue points can be measured using tagged MRI, after which such patterns can be visualized. To better categorize tongue motion it is useful to reduce the dimensionality of these patterns. The goal is to determine how homogeneous tongue motions for a given gesture are across subjects, including those that vary in native language and due to disorders. This paper examines midsagittal tongue motion (i.e., velocity fields) for 8 data sets during the motion from /i/ to /u/. Three subjects are American English (AE) speakers, one is a native Japanese speaker, one is a native Tamil speaker (for whom we have three data sets) and one is an AE speaker post glossectomy surgery (i.e., removal of part of the tongue). Two sets of analyses are being done, one on the tongue and one on the tongue-plus-jaw, to determine which better defines the data. In addition, a cluster analysis was done to compare commonalities among subjects.

Keywords: tongue, PCA, cluster analysis.

1. Introduction

The tongue contributes to speech by shaping the vocal tract. Vocal tract shapes result directly from changes in tongue shape. Tongue shape changes occur whenever tongue muscles contract, because the tongue is devoid of bones. Therefore, a better understanding of tongue function requires us to learn about its internal deformation patterns and their relationship to underlying muscle activity. The long term goal of this project is to determine the best quantities to represent and distinguish the complex tongue motion patterns across subjects and groups. The long term goal of this work is to study the effects of tongue cancer surgery on tongue motions during speech. At present we have available a small heterogeneous data set with only one glossectomy speaker, therefore, this paper will present a preliminary exploration of two statistical quantities, Principal Components Analysis (PCA), which measures variability, and Cluster Analysis, which measures commonality. The data are velocity fields extracted from tagged-MRI movies in the midsagittal plane of the tongue. Even the velocity fields in a single plane, for a simple vowel-to-vowel motion, show the deformations typical of this volume preserving, soft tissue structure.

Principal Components Analysis (PCA) is an excellent method to extract and represent patterns in high-dimensional data for which no expectations or a priori models are available. PCA reduces the dimensionality of a data set by quantifying its variability and is typically applied to a data set after removing the common component, that is, the mean. Each PC represents a vector of variability that categorizes the motions patterns of groups or individuals. PCA, or its close relative Factor Analysis, has been used to characterize speech related motion of the midsagittal tongue surface (Harshman et al, 1977, Jackson, 1988, Hoole et al, 1999) and the coronal tongue surface (Stone, et al., 1997, Slud et al, 2002). The speech of tongue cancer patients, pre and post glossectomy surgery, has also been characterized using PCA which distinguished tongue surface motions resulting from different reconstruction procedures (Bressman et al, 2005, 2007). One question asked by the present paper was whether the motion of the tongue alone is sufficient to represent the key tongue motions of the patient, or whether a tongue-plus-jaw region of interest (ROI) is needed. To answer this question several PCA's were performed. PCA 1 compared an ROI that included the tongue-plus-jaw muscles for seven normal data sets plus one patient at the onset of motion from /i/ to /u/ (see Figure 1a). PCA 2 was performed on a smaller region of interest: the tongue-only to determine the contribution of the jaw muscles to the pattern (see Figure 1b). Hold-one-out analyses were also performed for each of the individual subjects to see if the patient differed from the other subjects.

Cluster analysis is used to form subgroups such that subjects within a subgroup are more similar (patterns) to

one another than subjects in different subgroups. In this study we clustered subjects based on the global tissue point trajectories in the tongue and their x and y component motions. A hierarchical clustering algorithm was used for classifying subjects. The clustering trees were generated to explore the features that could be useful for categorization. With a larger data set, such analysis could reveal different velocity field patterns among AE speakers, non-native speakers, or patients. The question asked by this analysis was whether patient would be in a different group from the rest of the normal subjects.

These two statistical techniques, which perform point-to-point comparisons, assume that each data set contains the same coordinate points. However, the tongue varies considerably across subjects in MRI images due to differences in anatomy and resting posture of the head and tongue during image collection. Therefore, registration and scaling of data sets must be done prior to analysis of group data, otherwise the variability among humans will create error. In the long run, we aim to create a 3D registration system for the tongue. The present paper, however, begins with a registration method based on points chosen on the surface of the tongue.

2. Methods

2.1. Data used in the analyses. Subjects and Speech Material

Eight data sets were available for this analysis; demographics appear in Table 1. This was a non-homogeneous data set in that it contained (1) three data sets spoken by the same subject (subj 1, 2, and 3) recorded after two months, and one year, respectively; (2) three different native languages; (3) one speaker who underwent glossectomy surgery about 1 year prior to the study (subj 8). The surgery removed 1/3 of his tongue on the left side and replaced it with a radial forearm free flap, while preserving the tongue tip. All subjects were male. Differences in slice thickness and tag separation, which are always the same to create square voxels, change the resolution of the data, but should not significantly affect the goals of this preliminary study.

To record the data, the subjects repeated /i/-/u/ to the first two beats of a four beat metronome set at 0, 333, 800, 1400 ms in a 2 second repeat time. The last two beats were used for a controlled inhalation and exhalation. The timing was coordinated to the trigger of the MRI machine so that the first beat occurred at the onset of the MRI

acquisition and tags were applied 16 ms before the beat. This method is based on that of Masaki and colleagues (Masaki et al.1999, Shimada et al., 2002).

Table 1. Subject Demographics

Subj	Language	health	Tesla	ST / tag sep
1	Tamil	normal	1.5	7mm
2	Tamil	normal	1.5	7mm
3	Tamil	normal	1.5	7mm
4	English	normal	1.5	7mm
5	English	normal	1.5	7mm
6	English	normal	1.5	7mm
7	Japanese	normal	3.0	5mm
8	English	patient	3.0	6mm

2.2. Data Collection

To acquire each tagged cine seriesthe 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.

2.3. Pre-Processing of Data: Registration of subject data using Cine-MRI images
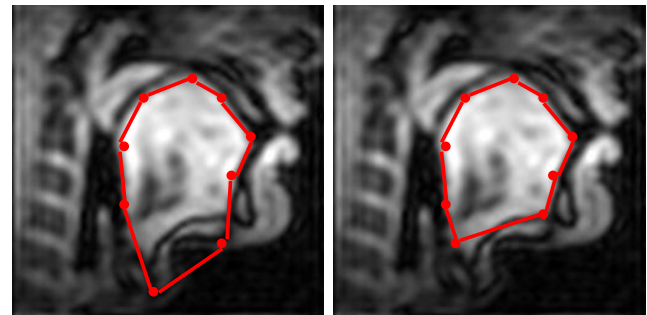
Figure 1. Nine landmark points were used to align the ROIs for all subjects, jaw muscles were included (left) or omitted (right).

To register the images across subjects, nine data points were identified for each subject on the surface of the tongue. The points were extracted from the Cine-MRI frame that corresponded to the tagged MRI frame used in the PCA (Figure 1). On the right, the points were: the base of the valleculae, the upper tip of the epiglottis (projected 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. On the left, the jaw
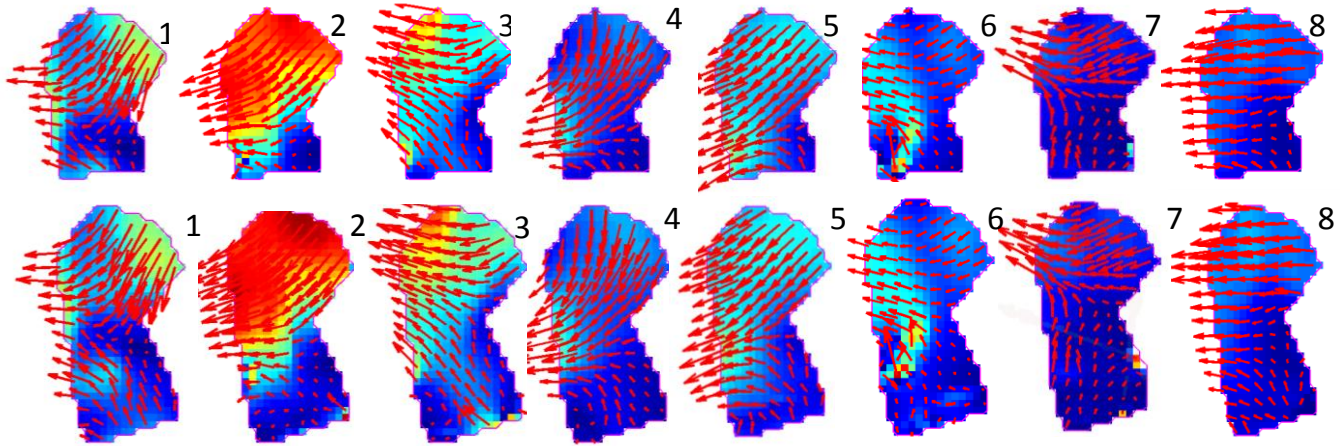
Figure 2. Velocity fields for all subjects' tongue-only and tongue-plus-jaw regions of interest

muscles were included by moving the two lowest points below the soft tissue of the chin. The points were registered across subjects and the rest of the voxels in the image were then repositioned and interpolated accordingly. Velocities were calculated for each tissue point and PCA's were done on both ROIs.

2.4. PCA Methods

PCA's 1 and 2 looked at directions of maximum variability in the motion data (velocity field) of the midsagittal slice for the 8 subjects at the onset of the transition from /i/ ('ee') to /u/ ('oo'). The tag extraction procedure involved: (1) HARP analysis to compute velocity fields; (2) alignment of all 8 data sets (rigid + scalar) (based on Figure 1A or 1B); (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 (Parthasarathy et al, 2007).

Hold-one-out Analyses. To determine how well each subject was represented, and to consider whether the method might distinguish normal from patient subjects, we performed a "hold-one-out" experiment. Eight PCA's were performed, each using 7 different subjects. The PC's of each analysis were then fit to the "held-out" data set to determine how well it was represented by the PC's of the other seven data sets.

2.5. Cluster Analysis

For each subject the data point consisted of its global tissue point trajectories and its x and y component motions, so data points are velocity field profiles for subjects. Using Pearson correlation as a distance metric and average linkage, agglomerative hierarchical clustering begins with each subject as its own cluster and at each stage chooses the "best" merge of two subjects or two clusters of subjects

until, in the end, all subjects are merged into a single cluster. The end result of hierarchical clustering is a tree structure or dendrogram (seen in Figures 4 and 5 below). At the bottom of the tree, each subject constitutes its own cluster and, at the top of the tree, all subjects have been merged into a single cluster. Merges between two subjects or between two clusters of subjects, are represented by horizontal lines connecting them in the dendrogram (Duda et al, 2001).

3. Results

3.1. Velocity fields.

Figure 2 depicts the velocity fields for each subject during the maximum /i/-to-/u/ motion. Although the common direction of motion is backward and converging, there are considerable subject differences. The first three subjects are actually the same subject at different dates and show considerable difference in motion. The patient (S8) has the least convergence. His tongue moves directly backward. For all subjects the tongue-plus-jaw images do not incorporate an additional motion direction they have an enlarged the lower region.

3.2. PCA 1 and PCA 2

These two analyses examined the tongue-plus-jaw vs. the tongue-only ROI's for all eight subjects, after subtracting the mean, and looked at the percent variance accounted for by the PC's. Table 2 shows the eigenvalues and variance explained by all the principal components (PC's) in both conditions. The first four PC's accounted 93% and 95% of the variance respectively. The biggest difference between the two analyses occurred in PC's 1 and 2. Although PC1 plus PC2 had similar explanatory power for both ROI's (72% vs. 74%), PC 2 explained more variance in the tongue-plus-jaw data (24%) than the tongue only data (15%) and PC1 explained less data (47% vs. 58%).

Table 2. PCA 1 and 2. Jaw vs. Tongue data for 8 subjects.

	Eigenvalues		Percent Explained		Cumulative percent	
	Jaw	Tongue	Jaw	Tongue	Jaw	Tongue
PC 1	195	145	47%	58%	47%	58%
PC 2	99	38	24%	15%	72%	74%
PC 3	59	36	14%	15%	86%	88%
PC 4	30	16	7%	6%	93%	95%
PC 5	12	5	3%	2%	96%	97%
PC 6	10	4	2%	2%	99%	99%
PC 7	6	3	1%	1%	100%	100%

The associated hold-one-out analyses in Table 3 showed this same effect of PC1 and PC2, despite the variability caused by three native languages, one subject with three data sets, and one glossectomy patient. PC1 plus PC2 explained almost the same amount of variance (1-4%) for both R0Is, and in the tongue-only condition PC1 explained more variance (9-13%) and PC2 less variance (1-10%). Figure 3 (explained below) shows that this is because PC2 represents, in part, the degree of upward motion occuring in the lower tongue and jaw, more of which is captured in the tongue-plus-jaw analysis. Therefore, the rest of the analyses were based on PCA 1, the tongue-plus-jaw data, as it included more information.

Table 3. Hold one out analyses for PCA 1 and 2. Percent variance explained by the first two PC's for the jaw+tongue (J) and the tongue-only (T).

		PC1	PC2	PC1+2
no S2	J	48%	25%	73%
	T	58%	16%	74%
no S3	J	46%	26%	71%
	T	55%	20%	75%
no S4	J	40%	29%	69%
	T	52%	19%	71%
no S5	J	47%	26%	72%
	T	58%	16%	74%
no S6	J	54%	28%	82%
	T	67%	18%	85%
no S7	J	60%	20%	80%
	T	63%	19%	83%
no S8	J	50%	26%	76%
	T	63%	16%	79%
min	diff	9%	-1%	1%
max	diff	13%	-10%	4%

Figure 3 depicts the mean velocity for the tongue-plus-jaw data set (middle image) and the addition of +/- 1SD of PC1 (horizontal images) and PC2 (vertical images). The mean velocity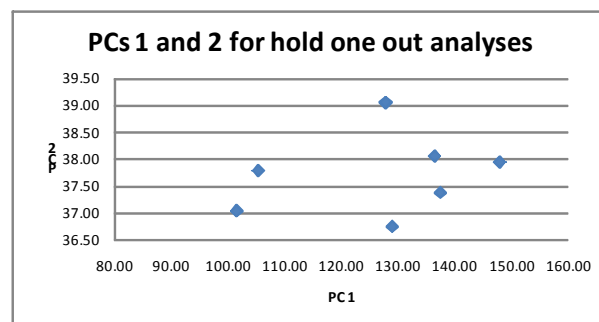 indicates that the predominant motion direction from /i/ to /u/ was backwards with anterior tongue lowering. The addition of PC 1 made this motion more downward/oblique. Subtraction of PC1 made the motion
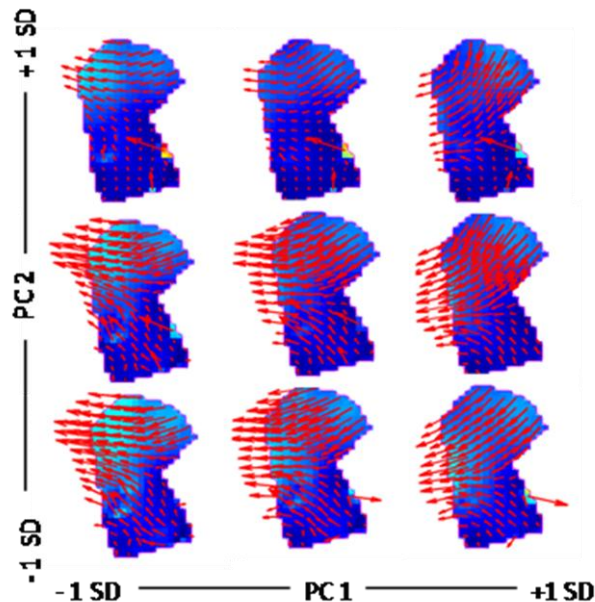
Figure 3. Mean velocity (middle) and +/- 1SD of PC 1 and PC2 for the tongue-plus-jaw data

Figure 4. Eigenvalues for PC 1 x PC2 for the hold-one-out analyses. The largest PC1 (147.82) is for the PCA that held out the patient.

more directly horizontal. The eigenvalues and percentages in Table 2 show that PC1 accounted for the bulk of the variance (47%). PC2, which accounted for 24% of the variance, affected the magnitude of the motion and the amount of upward motion in the lower tongue and jaw. Figure 4 graphs PC1 and 2 for each of the hold-one-out analyses. When the patient was withheld PC1 had its maximum value (147.82).

3.3. PC FITS

We performed PC fits to each observed velocity field by adding the mean value of all subjects, to the PC 1 and PC2 loadings for each subject (see Table 4). The velocity data of four subjects (3, 4, 5, 7) were very well fits by the mean plus PCs 1 and 2 (82% - 100%). Six subjects were represented primarily by the mean plus PC 1, which was

Table 4. Fits for each subject in PCA 1. Percent variance explained by the first two PC's in the jaw+tongue data

PCs	S1	S2	S3	S4	S5	S6	S7	S8
1	63%	76%	85%	95%	79%	47%	48%	40%
2	1%	0%	8%	0%	3%	8%	52%	21%
1+2	64%	76%	93%	95%	82%	55%	100%	61%

backward motion of the tongue, either directly back or obliquely down and back. PC2 increased (or decreased) the convergence in the anterior tongue; when added, the lower tongue and jaw tissue moved up and back. Subjects 7 and 8 loaded more heavily on PC2; reflecting that they had little to no upward motion in the lower tongue. Subject 8 also loaded on PC3 (24%) and PC4 (11%), which reduced the downward component in the anterior tongue.

3.4. Clustering.

Results of the cluster analysis on the combined xy velocity data can be seen in Figure 5. The normals-only
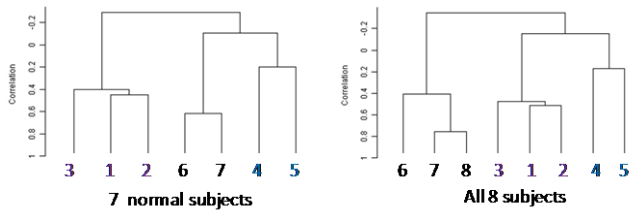

Figure 4. Dendrograms of clusters for tongue-plus-jaw data for the normal (left) and all (right) speakers.

analysis (left) shows that the three datasets by the same speaker (1,2,3) clustered together, as did two of the three AE speakers (4,5). The third AE speaker (6) grouped with the Japanese speaker (7). Adding the patient to the analysis (right) did not change the cluster alliances; the patient grouped with the Japanese and one AE speaker.

A comparison of the xy-motion cluster analysis with ones for the y direction motion and x direction motion indicates that the dominant movement pattern was in the y direction. Figure 5 below shows that the y clusters are identical to the xy clusters and different from the x clusters.

Figure 5. Dendrograms of cluster results of the xy, y, and x motions for the tongue-plus-jaw data.
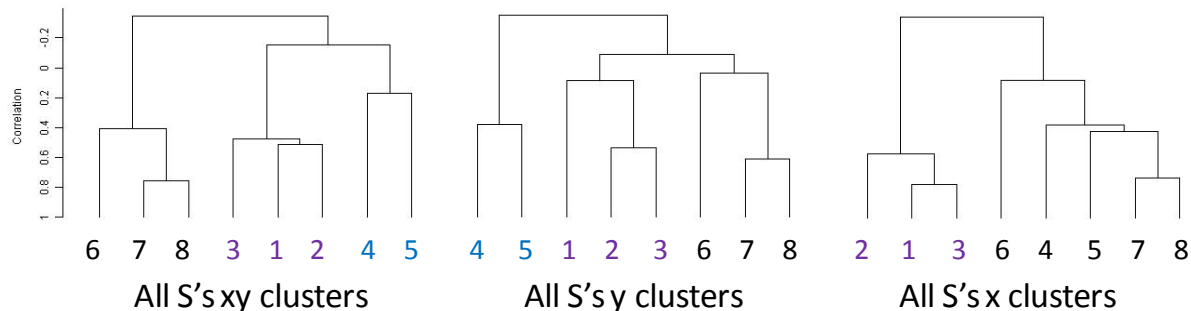
4.0. Discussion

4.1 Tongue-plus-jaw vs. tongue-only

The jaw muscles have a dual function in speech: to support the tongue and to move the jaw. PCA2, excluded this region in its ROI. Without the jaw muscles the velocity field variability was explained fairly well with a single PC. With them the second PC had a greater role, because PC2 represented, in part, the upward motion of the lower tongue. It was decided, therefore, that it is important to include the jaw muscle region in these PCA analyses.

4.2 PCA1 and individual subjects.

The datasets used in these analyses were very inhomogeneous there were replicates of one individual, multiple languages and one partial glossectomy. Because of this and the small number of subjects, the PCA fits varied widely across subjects. Although the patient had some unusual loadings, he did not present with a unique configuration. This may be true in any event, but in future analyses, with a larger and more homogeneous data set, this determination will be better made.

4.3 Comparison between Cluster Analysis and PCA.

Both analyses examined every voxel in the tongue or tongue-plus-jaw ROI, to determine similarities or differences among subjects. PCA looks only at variability among subjects, the mean is subtracted out. Cluster analysis includes all data, therefore, it includes the mean plus variability.

The PCA indicated that the maximum variability across subjects, PC1, was whether the backward motion from /i/ to /u/ was straight back or down and back. Similarly, PC2 indicated the degree of upward/backward motion of the lower tongue. The Cluster analysis grouped subjects according to raw tissue point motion. Athough it did not define the differences as PCA did, it created very interesting groupings. Firstly, subjects 1,2,3, who were the same subject, clustered together, reflecting some coherence in tissue point motion, despite their loading

quite differently on the PCs. Secondly, subjects 6, 7, 8 clustered together, reflecting similarity of motion (see Figure 2), despite their quite different history (AE speaker, Japanese speaker, glossectomy patient). Both techniques identified the backward motion as primary. PC 1 represented variation in backward motion. The clusters were identical for the xy motion and the y motion.

The clusters presented a different perspective from the PCA, they grouped subjects by shared voxel directions. The cluster did not detail the similarities across subjects, another analyses will be done to determine that. AS we collect larger data sets these techniques will be applied again to determine how the subjects are similar and different.

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] Duda, R.O., Hart, P.E. and Stork, D.G. (2001) Pattern Classification, 2nd edition John Wiley & Sons, New York.

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

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

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

[6] 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.

[7] 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

[8] 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.