Using Statistical Deformable Models to Reconstruct Vocal Tract Shape from Magnetic Resonance Images

Maria João M. Vasconcelos
Faculty of Engineering, University of Porto
Institute of Mechanical Engineering and Industrial Management
Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
e-mail: maria.vasconcelos@fe.up.pt

Sandra M. Rua Ventura
Radiology, School of Allied Health Science – IPP
R. Valente Perfeito 322, 4400-330 Vila Nova de Gaia, Portugal
e-mail: smr@estsp.ipp.pt

Diamantino Rui S. Freitas
Department of Electrical Engineering and Computers, Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
e-mail: dfreitas@fe.up.pt

João Manuel R. S. Tavares
Department of Mechanical Engineering, Faculty of Engineering, University of Porto
Institute of Mechanical Engineering and Industrial Management
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
e-mail: tavares@fe.up.pt

Corresponding author:
João Manuel R. S. Tavares
Department of Mechanical Engineering, Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
Phone: +351 22 508 1487, fax: +351 22 508 1445, e-mail: tavares@fe.up.pt, url: www.fe.up.pt/~tavares
Using Statistical Deformable Models to Reconstruct Vocal Tract Shape from Magnetic Resonance Images

Abstract: The mechanisms involved in speech production are complex and have thus been subject to growing attention by the scientific community. Magnetic Resonance Imaging (MRI) has been proven to be a powerful means in the understanding of the morphology of the vocal tract. Over the last few years, statistical deformable models have been successfully used to identify and characterize bones and organs in medical images and Point Distribution Models (PDMs) have gained particular relevance. In this work, we have studied the suitability of these models to characterize and further reconstruct the shape of the vocal tract in the articulation of Portuguese European (EP) speech sounds, one of the most spoken languages worldwide, with the aid of MR images. Therefore, a PDM has been built from a set of MR images acquired during the artificially sustained articulation of 25 EP speech sounds. Following this, the capacity of this statistical model to characterize the shape deformation of the vocal tract during the production of sounds was analyzed. Next, the model was used to reconstruct 5 EP oral vowels and the EP fricative consonants. As far as a study on speech production is concerned, this study is considered to be the first approach to characterize and reconstruct the vocal tract shape from MR images by using PDMS. In addition, the findings achieved permit one to conclude that this modelling technique compels an enhanced understanding of the dynamic speech events involved in sustained articulations based on MRI, which are of particular interest for speech rehabilitation and simulation.

Keywords: Medical Imaging; Image Analysis; Bioengineering; Modelling and Simulation; Speech Production; Vocal Tract Shape; Morphological Study.
1. Introduction

Image analysis may be considered to be a very active and promising research area, especially due to the numerous applications that may be found in distinct domains ranging from areas as diverse as the industry to medicine [1-26]. Over the last few years, many computational techniques have been developed to perform several image analysis tasks in a semi or fully automatic manner. Due to this, deformable model have received considerable attention [9, 11].

Active contours, commonly known as snakes [9, 11, 12], were the first deformable models used in image analysis. They consist of an elastic set of points which may be adapted to the shape of the object under study through the physical combination of internal and external forces. Recently, enhanced physical modelling approaches have allowed the integration of the objects’ physical behaviour in models built thanks to previously acquired [9, 11, 15]. In fact, physical modelling attempts to explore the real objects’ behaviour, frequently leads to unexpected results and the techniques which are very demanding in computational terms [9, 11]. Techniques of image analysis based on deformable templates, use geometrical shapes (templates) driven by parameterized functions in order to correctly identify the objects modelled in images [9, 11, 13].

One of the main problems resulting from this is that the construction of deformable templates is a complex process which is highly dependent on the object under consideration. As such, it has difficulties in adapting itself to the configurations of new objects [11, 14]. On the other hand, statistical modelling techniques [9-11], such as the one used to build Point Distribution Models (PDMs) [4, 5, 10, 16], consider a set of training images representing the possible configurations of an object in question to extract its main characteristics with the aid of statistical procedures, namely, the mean shape and possible deformations of the model built. The main advantages of these modelling approaches are usually considered to be the reduced computational requirements needed and high reliability, which are allied to the well defined behaviour of the model built [11]. Due to the advantages presented, PDMs, as well as their further extensions, in particular Active Shape Models (ASM) [4, 10] and Active Appearance Models (AAM) [4, 10], have been extensively used in different image analysis tasks, such as: visual inspection [1, 2], biometric recognition and automated surveillance [3] in addition to the localization and characterization of bones and organs in medical images [4, 5].
Since the beginning of studies of this nature, the process of speech production has attracted human interest aiming at reaching a deeper understanding and modelling of all the mechanisms involved by taking both morphological and speech acoustic aspects into consideration. The main anatomic aspects and the physiology of the human vocal tract are common to all individuals. However, speech production is an exceptionally complex and individualistic mechanism, which implies that the modelling developed, needs to be flexible so as to permit accurate individual characterizations [6-8].

Due to the important contribution that such studies represent in terms of the human quality of life, many studies have been conducted to study the vocal tract’s shape as well as the articulators during speech production. For example, a 3D geometrical database of the speech organs of one speaker, which has considered a set of 46 sustained French allophones, was recently built based on the linear component analysis of Magnetic Resonance Imaging (MRI) and computer tomography (CT) images [17]. On the other hand, in [18] a semi-automatic method for extracting vocal tract movements from X-ray films is presented.

From the imaging modalities that have been taken into consideration in order to study the vocal tract’s shape and articulators, MRI has been the most commonly accepted [6-8]. In fact, the use of MRI allows for the study of the whole human vocal tract and consequently, the calculation of area functions for a better understanding of the speech production mechanisms [6-8]. In addition, the quality and resolution of soft-tissues and the use of non-ionizing radiation are key advantages presented by MRI [6-8, 19, 20]. Moreover, static MRI measurements have been revealed to be representative of dynamic speech and have demonstrated that the articulations in the MRI data are hyperarticulated [7, 21].

According to the International Phonetic Alphabet (IPA), the European Portuguese (EP) language, which is one of the most widely spoken languages, consists of a total of 30 sounds: 9 vowels, 2 semivowels and 19 consonants [7]. In the current work, the suitability of statistical deformable models to characterize and reconstruct the shape of the human vocal tract in the articulation of EP speech sounds, by using a set of Magnetic Resonance (MR) images acquired during artificially sustained articulations of 25 Portuguese sounds, has been studied.

Since the mid seventies, statistical methods have been applied to the analysis of speech production data. For example, in [27] the authors used a component analysis to identify an adequate set of articulatory features for
the tongue shapes of ten English vowels. One year later, in [28] the statistical analysis of real data to describe
the position of the articulatory organs was applied. In [29] a factor analysis of the lateral shapes of the vocal
tract is described. Thereafter, [30] a principal component analysis was used to examine sagittal tongue
contours for five English vowels constructed from ultrasound images. Despite all the work that has been
extensively done, the application of statistical deformable models to characterize and reconstruct speech
sounds with MR images has only just become a reality.

This paper is organized as follows: in the following section, the methods adopted by the current study are
introduced; Following this, the statistical model built is described and its employment in terms of the
characterization and reconstruction of EP speech sounds from MR images is studied; finally, in the final
section of this paper conclusions are drawn.

2. Methods

In this section, the methodologies adopted in this work to characterize and reconstruct EP speech sounds
from MR images with a statistical deformable model are described. Thus, the MRI protocol adopted in the
images acquisition is presented. Following this, an explanation as to the modelling of objects with PDMs is
provided.

2.1. Magnetic Resonance Imaging protocol

The MR images were acquired by using a Siemens Magneton Symphony 1.5T system and a head array coil,
with the subject lying in the supine position. Due to the use of this experimental setup, the acquisition of T1-
weighted sagittal slices of 5 mm thickness was obtained by using Turbo Spin Echo Sequences, with
recording durations of approximately 10 s and the adoption of the following parameters: a field of view 150
mm, an image matrix 128x128 and an image resolution of 0.853 px/mm. The speech corpus consisted of a
set of 25 MR images collected during sustained articulations of 25 EP speech sounds; that is, one image was
acquired for each sound considered.
This static study was designed to obtain the morphologic data of the vast majority of the range of the articulators’ positions with the aim of attaining imaging characterization and the reconstruction of EP speech sounds. Since, as may be verified in Figure 1, the sagittal data is particularly useful in the study of the entire vocal tract anatomy, thereby demonstrating the main aspects of the shape and positions of some articulators, e.g. the tongue, lips and velum.

2.2. **Statistical Deformable Modelling**

As has been previously mentioned, PDMs are used in the statistical modelling of objects by analysing their shape configurations from a set of training images. Thus, a PDM describes the mean shape of the object modelled, as well as the admissible variations in relation to this same mean shape.

In this work, one has been statistically modelled the vocal tract from 25 MR images partially depicted in Figure 2. From the images included in this figure, one may observe various vocal tract configurations for EP vowels and consonants production, as well as some oral and nasal sounds.

In the building process of a PDM, each shape of the object to be modelled which is presented in the training imaging set should be described by a set of labelled landmark points. These points should reflect important aspects of the object’s boundaries or interior, Figure 3. (In the current and subsequent images the landmark points appear connected by fictitious line segments so as to enhance their visualization.) It should be noted that the manual labelling of landmark points in all training images requires a comprehensive knowledge of the object in question, as the resultant model behaviour greatly depends on the landmark points selected.

In this work, the manual selection of the landmark points has been carried out by one of the authors who has medical imaging knowledge of the MR images sequentially displayed on the computer screen in addition to being cross-checked by another author in accordance with the following rule.

The labelling process took the anatomic location of the vocal tract articulators into consideration by adopting the following landmark points:

- Four points in the lips (front and back of the lip margins);
- Three points corresponding to the lingual *frenulum* and tongue’s tip;
- Seven points equally spaced along the surface of the tongue;
- Seven points along the surface of the hard palate (roof of the oral cavity) placed in symmetry with the tongue points;
- One point at the velum (or soft palate);
- Three points equally spaced at the posterior margin of the oropharynx (behind the oral cavity).

Thus, in each of the 25 MR images, 25 landmark points which are used in the building process of the PDM of the vocal tract were defined.

In order to study the variation of the coordinates of the landmark points of the training shapes, it is initially necessary to align them, by using dynamic programming for example [31]. Hence, given the co-ordinates \((x_{ij}, y_{ij})\) of each landmark point \(j\) of the shape \(i\) of the modeled object, the shape vector is:

\[
\begin{bmatrix}
    x_0 & y_0 \\
    x_1 & y_1 \\
    x_2 & y_2 \\
    \vdots & \vdots \\
    x_{n-1} & y_{n-1}
\end{bmatrix}
= \begin{bmatrix}
    x_0 \\
    x_1 \\
    x_2 \\
    \vdots \\
    x_{n-1}
\end{bmatrix}
\begin{bmatrix}
    1 \\
    0 \\
    0 \\
    \vdots \\
    0
\end{bmatrix}
+ \begin{bmatrix}
    y_0 \\
    y_1 \\
    y_2 \\
    \vdots \\
    y_{n-1}
\end{bmatrix}
\begin{bmatrix}
    0 \\
    1 \\
    0 \\
    \vdots \\
    0
\end{bmatrix},
\]  

(1)

where \(i = 1 \ldots N\), with \(N\) representing the number of shapes in the image training set and \(n\) the number of landmark points. Once the shapes are aligned, the mean shape and the variability of the modelled object may be found. The modes of variation characterize the manner in which the landmarks of the modelled object tend to move together, the results of which may be obtained by applying a Principal Component Analysis (PCA) to the deviations from the mean. Thus, it is possible to rewrite each shape vector \(x_i\) as:

\[
x_i = \bar{x} + P_s b_s,
\]  

(2)

where \(x_i\) represents the coordinates of the \(n\) landmark points of the new shape of the modelled object, \((x_k, y_k)\) are the coordinates of the landmark point \(k\), \(\bar{x}\) has the mean coordinates of all landmark points, \(P_s = \begin{pmatrix}
    p_{s1} & p_{s2} & \cdots & p_{st}
\end{pmatrix}\) is the matrix of the first \(t\) modes of variation, \(p_{si}\) corresponds to the most significant eigenvectors in a Principal Component Analysis applied to the coordinates of all landmark points, and \(b_s = \begin{pmatrix}
    b_{s1} & b_{s2} & \cdots & b_{st}
\end{pmatrix}^T\) is a vector of weights for each variation mode of the modelled object. Each eigenvector describes the manner in which linearly correlated \(x_i\) move together over the training set, referred to as a mode of variation. The equation above represents the PDM of an object and may be used to generate its new shapes. Further details about the construction of PDMs are be found in [10, 16].
Despite the fact that the following has not been analysed by this work, the local grey-level environment of each landmark point may also be considered in the statistical modelling of objects from images. Thus, statistical information is obtained in relation to the mean and covariance of the grey values of the pixels around each landmark point [10]. Furthermore, the combination of a PDM in addition to the grey level profiles of each landmark point which has been used may be considered to segment the object modelled in new images. This approach leads to the Active Shape Models (ASMs) by including an iterative optimisation scheme with PDMs permitting an initial estimated shape, that is, the mean shape $\bar{x}$, of the modelled object to be refined so as to match the shape represented in a new image [9-11, 22]. This refining process may be summarized by the following steps: 1) The necessary displacement that dislocates it to a better position, that is, closer to the final position, is calculated at each landmark point of the model ; 2) The calculation of the changes in the overall position, orientation and scale of the model that best satisfy the displacements found in 1); 3) In conclusion, by calculating the required adjustments for the model parameters, residual differences are used to deform the model according to the new shape. This image segmentation process has been improved in [23] by adopting a multiresolution approach: first, a multiresolution pyramid of the input images is built by applying a Gaussian mask; Following this, the grey level profiles at the various levels of the pyramid built are studied. Consequently, the ASMs segment the input images in a more rapid and reliable manner. A similar approach was applied to the mixture of PDMs with information about image texture that has resulted in the creation of the Active Appearance Models (AAMs) [3, 9, 10]. These models have been successfully used in the segmentation of complex objects represented in images. However, when compared to ASMs, they present higher computational complexity.

3. Results and Discussion

A framework in MATLAB has been developed to build statistical deformable models, namely PDM and ASMs, which integrates the Active Shape Models software [32]. To build the PDM of the vocal tract, a training set of MR images acquired during artificially sustained articulations of a number of Portuguese sounds, which are partially depicted in Figure 2, was used. The 25 images were acquired from one young male subject in a similar manner to that which has been formerly used by other studies that use MRI to study
the vocal tract \cite{6, 7, 8, 17, 25, 26}, due to the variability of the speech subject in addition to the considerable number of sounds studied (speech corpus). The training of the subject was performed to ensure the proper production of the intended EP speech sounds and to reduce speech subject variability. Moreover, the subject in question had a vast knowledge of EP speech therapy.

As expected, the manual labelling process was revealed to be difficult and extremely time consuming. Furthermore, the considerable noise presented in MR images as well as the significant variability among the sounds involved, make the use of automatic approaches very difficult to accomplish successfully.

From Table 1, one may observe that the initial 7 modes of the statistical deformable model built, that is, 14\% of the modes of variation, are capable of explaining 90\% of all variance of the vocal tract shape. Additionally, one may conclude that the first 10 modes, i.e. 14\% of the modes of variation, illustrate 95\% of all variance and the opening 16 modes, that is to say, 32\% of the modes of variation, provide an explanation for 99\% of all variance. This indicates that the PDM which has been built is capable of considerably reducing the data set that is required to represent all shapes that the vocal tract held in the images training set.

The effects of varying the initial 4 modes of variation are depicted in Figure 4. This figure allows one to become aware that the first mode is associated with movements from the high front to the lower back of the tongue in the oral cavity. With regards to the second mode of variation, it is possible to observe the vertical movement of the body of the tongue towards the palate. On the other hand, the variations of the third mode have been shown to be related with the lip movements. Finally, the fourth mode of variation reflects the approximation of the tip of the tongue to the upper alveolar region.

Following the construction of the vocal tract model, some sounds were chosen to be reconstructed by using the statistical deformable model built, namely 5 EP oral vowel sounds and the EP fricative consonants. Phonetically, the EP consonants are classified according to the obstruction of the vocal tract from the front to the back of the mouth; the EP vowels are regarded as being long and somewhat continuous sounds, classified from the front to the back of the mouth and from the higher to the lower tongue positions. EP consonants are classified according to the places where the articulators converge in order to obstruct the vocal tract – articulation points. In this manner, it is relatively easy to identify the distinctive features of the sounds produced. As far as the EP fricative consonants are concerned, the articulation points are: labiodental [f, v],
alveolar [s, z] and post-alveolar [ch, j]. With regards to the production of the EP vowels [i, e], the tongue moves to higher frontal positions, and in the case of the EP vowels [o, u], the tongue moves to more elevated backward positions. The EP sound [a] is produced when the tongue is to be found in a central and mid-low position.

The main goal of the present study in this second phase was to conclude whether the modes of variation of the statistical deformable model built could be combined in order to successfully reconstruct, that is, reproduce, an EP speech sound. The sounds that were revealed to be the easiest to reconstruct were the vowels [i] and [o] and the consonant [j], as they only required the combination of two variation modes of the model built. Thus, in order to obtain the shape of the vocal tract when articulating the vowel [i], it was necessary to merge the 1st and the 4th modes. On the other hand, the reconstruction of the vowel [o] meant that the 1st and the 3rd modes needed to be united. Finally, the combination of the 1st and 8th modes enabled the reconstruction of the EP sound of the consonant [j].

Through the union of three variation modes of the statistical deformable model built, it was possible to reconstruct the shapes of the vocal tract when articulating the EP consonants [ch] and [f]. Thus, the combination of the 1st, 3rd and 7th modes permitted the reconstruction of the vocal tract shape associated with the consonant [ch], and by using the 1st, 2nd and 3rd modes, the vocal tract shape related to the consonant [f] was reproduced.

In order to obtain the shape of the vocal tract when articulating the EP consonant [v], it was necessary to bring together the 1st, 2nd, 3rd and 5th modes of variation of the statistical deformable model built whereas the consonant [s], implied the union of the 1st, 2nd, 3rd and 8th modes. On the other hand, the reconstruction of the EP vowel [a], required the combination of the 1st, 3rd, 5th, 7th and 8th modes. In the case of the EP vowel [e], the articulation of 1st, 3rd, 5th, 8th and 9th modes was adopted. Finally, the reconstruction of the EP consonant [z], implied the union of the 1st, 2nd, 4th, 6th and 8th modes, Figure 5.

Contrary to all expectations, the EP vowel [u] required the combination of the highest number of variation modes to reconstruct the related vocal tract shape. Before initiating a reconstruction study, we held the belief that the EP vowels were the easiest sounds to be (re)produced since that the air flows without any obstruction on the vocal tract. However, this was proven not to be the case. Consequently, in order to rebuild the sound
of the EP vowel [u], the combination of the first ten modes of variation of the statistical deformable model built was required. This indicates that, from a morphological and dynamical point of view, the EP vowel [u] is not as simple to reconstruct as one would initially believe.

In Figure 5, the resultant reconstructions of the vocal tract shape relating to the EP consonants [s] and [z] and the vowels [u] and [i] is depicted. In order to assess the quality of the reconstruction of the shape of the vocal tract in the articulation of EP speech sounds, the minimum, maximum and mean errors and the standard deviation of the Euclidean distances between the landmark points of the original shape and that which is to be reconstructed must be calculated. Table 2 indicates these values for the reconstructions presented by Figure 5.

In terms of phonation, fricative consonants are classified as either being voiceless or voiced, implying that the sounds are produced with or without the vibration of the vocal cord. The process of reconstruction used throughout this work has also proven that the dynamics associated with the production of these sounds are distinct: on the one hand, the articulatory points are located in diverse positions. On the other hand, the combination of the variation modes has been proven to be a more complex phenomenon. This may be exemplified by the fact that the reconstruction of the voiceless consonant [s] implied the combination of 4 variation modes (1st, 2nd, 3rd and 8th) whereas the voiced sound [z] required five variation modes (1st, 2nd, 4th, 6th and 8th).

The sounds used to build the vocal tract shape model represent the vast majority of EP speech sounds. As has been previously mentioned, EP speech consists of a total of 30 sounds: 9 vowels, 2 semivowels and 19 consonants. Throughout, this work, 25 of the 30 existing EP speech sounds were used to construct the statistical deformable model built which was adopted to characterize and reconstruct the vocal tract shape during the production of sounds.

4. **Conclusions**

A PDM has been applied to MR Images in order to characterize and reconstruct the shape of the vocal tract in the articulation of 25 EP speech sounds. Up to the present moment, all research involving PDMs and
medical images have been related to the localization and characterization of bones and organs. Thus, in the field of the study of speech production, this is the first work to apply this modelling technique so as to characterise and reconstruct the vocal tract shape through the use of MR images. From the experimental results obtained, one may conclude that the statistical deformable model built is capable of efficiently characterising the behaviour of the vocal tract modelled from the MR images studied. In fact, the modes of variation of the model built could provide an explanation of the actual actions involved in the EP speech sounds considered, such as: the movement of the tongue in the oral cavity, the lip movements or the approximation of the tip of the tongue to the alveolar region. Additionally, it has been verified that the modelling performed could reduce the data set needed to characterize all variations of the vocal tract’s shape during the production of the EP speech sounds, as 99% of these are explained by just 32% of the total of all the modes of variation. The statistical deformable model built has also revealed that it could easily be used to reconstruct the vocal tract shape in the articulation of speech sound. For example, EP speech sounds such as the vowels [i] and [o] or the consonant [j] may be obtained through the combination of just two variation modes of the model built, whilst the vowel [u] required the union of the first ten modes of variation to be successfully reproduced. As a result of the assessment carried out on the reconstructions obtained through the use of the statistical deformable model built, one may analytically prove their elevated quality as all mean errors were inferior to 8 pixels. Prior to this study, it was believed that EP speech vowels were the easiest sounds to be reproduced as the air flows trough the vocal tract without any obstruction. However, the sound that was the most difficult to be successfully reconstructed was the vowel ([u]), thus indicating that it is morphologically more complex to reconstruct this vowel by using the model built. In comparison to the Articulatory Synthesizer (ASY) method developed by Haskins Laboratory (USA) [33] which proposed that a human vocal tract model with eight articulatory parameters be used for speech synthesis, our approach is based on a statistical deformable model that employs a larger amount of parameters (25 landmark points). However, it should be pointed out that fewer landmark points could be
used in the modelling process. However, the resultant model will not be as accurate and realistic, two essential aspects to be taken into consideration in speech rehabilitation and simulation.

Thus, considering all the results obtained and described in the previous section, one may conclude that the use of point distribution models allow for a clearer understanding of the dynamic speech events involved during sustained articulations. These vocal tract shape models are also be useful for speech rehabilitation and simulation purposes, namely to recognize and simulate the compensatory movements of the articulators during speech production.

Moreover, the data acquired and analysed throughout this study may also provide a valuable contribution for the construction of 3D articulatory models for EP speech synthesis. It may also be used for further studies on EP articulatory phonetics as well as for the modelling of the vocal tract for EP speech synthesis, with applications in fields such as speech pathology, linguistics and artificial speech.

5. Acknowledgments

This work was partially carried out under the scope of the project entitled “Methodologies to Analyze Organs from Complex Medical Images – Applications to the Female Pelvic Cavity”, with the reference PTDC/EEA-CRO/103320/2008, financially supported by the “Fundação para a Ciência e Tecnologia FCT”. The first author would like to acknowledge the support of the PhD grant SFRH/BD/28817/2006 from FCT, and the second author would like to express her gratitude for the support of the PhD grant SFRH/PROTEC/49517/2009 from the Instituto Politécnico do Porto (IPP) and FCT.

The images taken into consideration by this work were acquired at the Radiology Department of the Hospital S. João, Porto, due to the collaboration of Isabel Ramos (Professor at the Faculty of Medicine, University of Porto, and Department Director) and of the technical staff, which are most gratefully acknowledged. Additionally, the authors would like to express their deep gratitude to Dr. Vanessa Esteves from the University of Porto – Portugal for her valuable help with the English version of this paper.

6. References


FIGURE CAPTIONS

Figure 1: MR sagittal slice indicating the vocal tract organs.

Figure 2: 9 of the 25 training images used to build the vocal tract model.

Figure 3: a) Training image, b) landmark points selected, c) image labeled with the overlapped landmark points selected.

Figure 4: Effects produced by the variation of each of the first four modes of variation of the vocal tract model built ($\pm 2sd$).

Figure 5: Reconstruction of the EP speech sounds [s], [z], [u] and [i]: a) original shape, b) reconstructed shape and c) both shapes overlapped.
TABLE CAPTION

Table 1: First 16 modes of variation of the model built and their retained percents.

Table 2: Obtained errors of the reconstructed shapes.
FIGURES

Figure 1
Figure 2

Figure 3
Figure 4
Figure 5
TABLES

Table 1

<table>
<thead>
<tr>
<th>Mode of variation</th>
<th>Retained Percent</th>
<th>Cumulative Retained Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
<td>43.598%</td>
<td>43.598%</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>12.340%</td>
<td>55.938%</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>10.988%</td>
<td>66.926%</td>
</tr>
<tr>
<td>( \lambda_4 )</td>
<td>9.345%</td>
<td>76.271%</td>
</tr>
<tr>
<td>( \lambda_5 )</td>
<td>6.947%</td>
<td>83.218%</td>
</tr>
<tr>
<td>( \lambda_6 )</td>
<td>4.724%</td>
<td>87.942%</td>
</tr>
<tr>
<td>( \lambda_7 )</td>
<td>2.675%</td>
<td>90.617%</td>
</tr>
<tr>
<td>( \lambda_8 )</td>
<td>1.973%</td>
<td>92.590%</td>
</tr>
<tr>
<td>( \lambda_9 )</td>
<td>1.428%</td>
<td>94.018%</td>
</tr>
<tr>
<td>( \lambda_{10} )</td>
<td>1.312%</td>
<td>95.330%</td>
</tr>
<tr>
<td>( \lambda_{11} )</td>
<td>0.978%</td>
<td>96.308%</td>
</tr>
<tr>
<td>( \lambda_{12} )</td>
<td>0.797%</td>
<td>97.105%</td>
</tr>
<tr>
<td>( \lambda_{13} )</td>
<td>0.654%</td>
<td>97.759%</td>
</tr>
<tr>
<td>( \lambda_{14} )</td>
<td>0.537%</td>
<td>98.296%</td>
</tr>
<tr>
<td>( \lambda_{15} )</td>
<td>0.441%</td>
<td>98.737%</td>
</tr>
<tr>
<td>( \lambda_{16} )</td>
<td>0.334%</td>
<td>99.071%</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonant [s]</td>
<td>1.72</td>
<td>16.22</td>
<td>7.22 ± 4.13</td>
</tr>
<tr>
<td>Consonant [z]</td>
<td>1.04</td>
<td>18.25</td>
<td>7.11 ± 5.19</td>
</tr>
<tr>
<td>Vowel [u]</td>
<td>2.51</td>
<td>13.97</td>
<td>7.57 ± 3.58</td>
</tr>
<tr>
<td>Vowel [i]</td>
<td>1.91</td>
<td>18.64</td>
<td>9.10 ± 4.20</td>
</tr>
</tbody>
</table>