Identification of Independent Kinematic Regions of the Face during Speech Production

Jorge C. Lucero[1] and Kevin G. Munhall[2]
[1]Dep. Mathematics, Univ. Brasilia, [2] Deps. Psychology and Otolaryngology, Queen's University
E-mail: [1]lucero@unb.br, [2]kevin.munhall@queensu.ca

Abstract

This paper reports our progress on the empirical modeling of facial biomechanics during speech production. Our model is based on decomposing the facial surface into a finite set of linearly independent kinematic regions, which is used as a basis to represent the total facial motion. The main algorithm, based on the column-pivoted QR factorization, is reviewed and compared to other techniques commonly used in statistical regression. The results show that the former technique is more robust to variations in the speech data, and detects regions with a higher measure of independency.

1 Introduction

In recent works [6, 7], we have presented an empirical strategy for building a 3D model of facial physiology for applications to speech production and perception studies. The model is based on the analysis of facial movement data recorded from a subject producing speech, to detect regions which follow independent motion patterns. The total motion of the face is next expressed as the linear combination of the movement of the independent regions.

We argue that this modeling approach offers a number of advantages because it focuses the data analysis on the generating mechanism for speech gestures: the facial musculature. The action of muscles is obviously spatially concentrated and thus facial regions can be found that are associated with individual muscles or synergies of muscles that are in close proximity or whose actions are spatially localized. This muscle-based approach is consistent with a productive tradition in the analysis of facial expression and the study of perception of expressions [2] and this approach has also been a powerful tool in facial animation [9]. However, instead of setting a model by defining biomechanical properties of skin tissue and muscle structure based on *a priori* theoretical reasons, and estimating its parameters from available measures from the literature, we propose to infer a possible model just by looking at the measured motion patterns of the facial surface. For each individual, we let the motions define what regions of the face move as an independent unit, what the boundaries of the surface regions are, and where the spatial peak of motion is located. This can be seen as a lumped representation of the muscular actions and their influence on the facial tissue biophysics.

Another advantage is that the independent regions can be animated in arbitrary facial configurations. There is considerable interest in speech perception research on the role of individual talker characteristics in speech perception [3]. Studies that involve the use of animating a generic face or the animation of one talkers morphology with another talkers motion must solve a registration and morphing problem [5]. The identification of key features and spatial regions is one form of solution to this correspondence problem.

The next sections will review and discuss the modeling algorithm using two sets of speech motion data collected from a subject.

2 Data

The data consist of the 3D position of 57 markers distributed on a subject's face, recorded with a Vicon equipment (Vicon Motion Systems Inc., Lake Forest, CA) at a 120 Hz sampling frequency, and expressed in head coordinates. The approximate lo-

cation of the markers is shown in Fig. 1.

The data were recorded while the subject was producing 50 selected sentences from the Central Institute for the Deaf Everyday sentences [1], listed in `http://www.mat.unb.br/lucero/facial/qr2.html` (subject S2). The set of 50 sentences was recorded twice, forming two datasets which will be denoted as S2a and S2b.

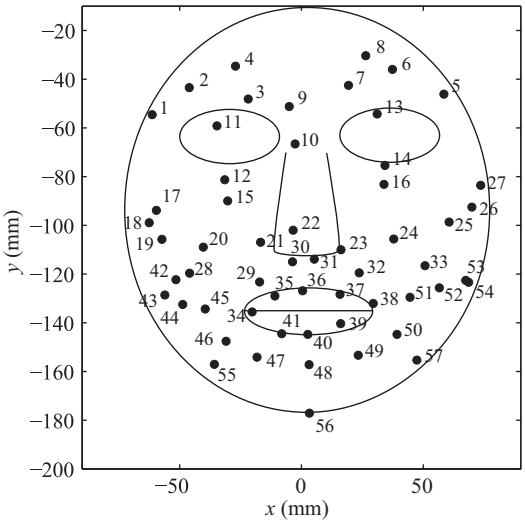

Figure 1: *Spatial distribution of the marker positions superimposed on a schematic face*

For each data set, the displacement of each marker was computed relative to an initial rest (neutral) position. Next, the markers' displacements for all sentences where concatenated and arranged in a displacement data matrix A.

3 Modeling algorithm

The model is based on the so-called *subset selection problem* of linear algebra [4]. Assume a given data matrix A and the observation vector b, and that a predictor vector x is sought in the least squares sense, which minimizes $\|Ax-b\|_2^2$. Instead of using the whole data matrix A to predict b, it may be desirable to use only a subset of its columns, so as to filter out the data redundancy.

To solve the subset selection problem, the most linearly independent columns of matrix A must be identified. Let A_k denote a subset of k columns of A. A measure of "independency" of the subset is provided by the smallest singular value of A_k, σ_k, which measures the distance of A_k to the set of k-rank singular matrices, in the 2-norm. In principle, the subset selection problem could be solved by testing all possible combinations of k columns from the total of n columns of A. However, the number of possible combinations could be prohibitively large.

A good solution to the subset selection problem is provided by the QR factorization technique with column pivoting [4]. That algorithm decomposes A in the form $A\Pi = QR$, where Π is a column permutation matrix, Q is an orthogonal matrix, and R is an upper triangular matrix with positive diagonal elements. The permutation matrix Π reorders the columns of A to make its first columns as well conditioned as possible. Therefore, the first k columns of $A\Pi$ may be then adopted as the sought subset of k least dependent columns.

A number of other algorithms have been proposed to find solutions to the subset selection problem. In statistical regression, a different criterion is commonly used [8]. Instead of looking at the independency of the columns, the subset A_k that minimizes the residual $\|A_k x - b\|_2^2$ is sought. Since the objective of the model is to predict an arbitrary observation b, one would normally want to do such prediction with the smallest possible error. However, a disadvantage has been pointed out for this approach: it looks at the output of the model, instead of its structure. If a redundant marker has a large motion, it might be included in the selected subset because of its large contribution to the total output [10]. In fact, there can be a trade-off between the independence of the selected columns and the total error at the model output [4]. It has been shown that minimization of the output error might also lead to unstable solutions that are highly sensitive to perturbations in the data set.

Once a subset of independent columns has been selected, the remaining (redundant) columns are approximated as linear combinations of the independent ones by using a least square algorithm. The coefficients of the linear combinations define the independent kinematic regions, and may be extended to arbitrary facial points by interpolation [6, 7].

4 Analysis of the facial data

Table I shows the index of the first 10 markers selected by the column-pivoted QR factorization (CPQR), when using the first 30 sentences in the

two datasets. For comparison with the second approach mentioned above, the Table also shows results when using a forward selection with sequential replacement algorithm (FSSR)[8]. This algorithm selects the markers one by one, while keeping the error at the output of the model (i.e., the residual when fitting the remaining data columns to the selected ones) to the minimum possible. After each marker is added to the subset, the previously selected markers are reviewed and replaced, if such a replacement leads to a lower error. Therefore, the order in which the markers appear in Table I is meaningless. The CPQR algorithm, on the other hand, orders the markers according to their relative independency.

Table 1: *Selected subset of markers. CPQR: column-pivoted QR factorization. FSSR: forward selection with sequential replacement algorithm.*

	QR		FSSR	
Order	S2a	S2b	S2a	S2b
1	40	40	39	35
2	34	34	52	41
3	13	13	35	50
4	38	38	47	11
5	47	48	7	20
6	42	47	11	43
7	6	11	43	56
8	56	36	56	37
9	11	42	38	7
10	36	49	13	13

In case of the CPQR algorithm, the first selected marker is the 40th, at the center of the lower lip, which has the largest displacement. The second is marker 34, at the lip's left corner, next the left eyelid (13), and the lip' right corner (38). The next selected markers include markers at the lower-right portion of the face (42, 47, 48, or 56), depending on the dataset, the right eyelid (11), the upper lip marker (36), and left eyebrow (6). When the 10 markers selected for S2a and S2b are compared, 8 of them appear in both sets.

When the CPQR vs. the FSSR results are compared, we note differences in the selected subset of markers. However, the differences come from markers that are located close together in the face. For example, comparing the results for set S2a, we see that the FSSR algorithm selects marker 7 instead of 6, both at the left eyebrow, 43 instead of 42, both at the right cheek, and so on.

Finally, comparing the markers selected by FSSR for S2a vs. S2b, we note that only 4 markers are common to both sets, vs. 8 for the CPQR algorithm. This fact implies a higher sensitivity to variations in the input data (or to data perturbations) of the FSSR.

In the case of the subset selected by the CPQR algorithm, the smallest singular values σ_{10} are 92.2 and 88.5, and the residuals are 578.7 and 531.5, for datasets S2a and S2b, respectively. In the case of the FSSR algorithm, the smallest singular values are 91.7 and 74.3, and the residuals are 532.7 and 496.0, respectively. Therefore, the subsets selected by the CPQR algorithm are more independent (larger σ_{10}), at the expense of larger error at the model output.

In addition to the forward selection with sequential replacement, we tested backward selection and other statistical techniques from the literature [8]. The results were similar to those reported above.

As an illustration of the independent regions computed by the CPQR algorithm for both datasets, the regions corresponding to markers at the center of lower lip, both lip corners, and center of upper lip are shown in Figs. 2 and 3. In the figures, the red and blue areas represent regions with positive and negative weights, respectively. Particularly, note that although the upper lip marker appears in different positions in the list of independent markers in Table I, the associated regions have similar shapes.

Following this analysis, facial animations of arbitrary speech utterances may be next produced by driving the selected subset of independent markers with collected records. As an example, animations for sentences 31 to 50 of both datasets are available in `http://www.mat.unb.br/lucero/facial/qr2.html` in AVI format.

5 Conclusion

The QR factorization with column pivoting algorithm provides a simple and robust technique for facial motion analysis and animation. It identifies a subset of independent facial regions, whose combined motion defines the total motion of the whole facial surface.

The model has an empirical nature, however, it reflects the underlying biomechanical structure of the

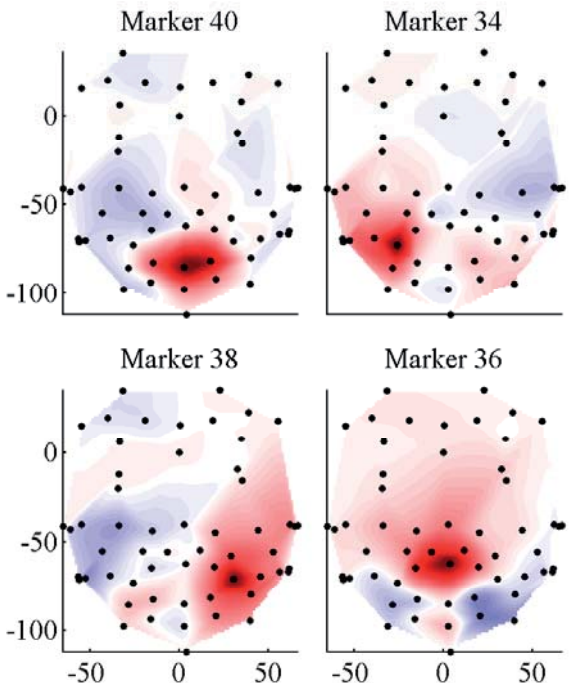

Figure 2: *Four independent kinematic regions for dataset S2a.*

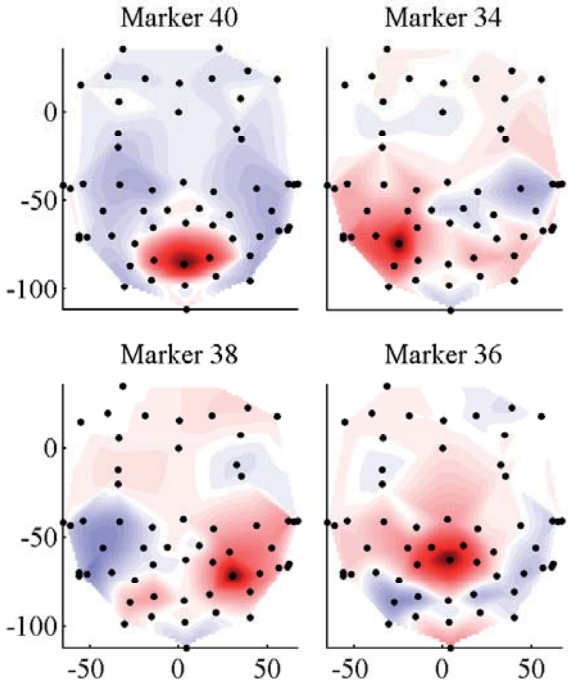

Figure 3: *Four independent kinematic regions for dataset S2b.*

face and may be used to infer aspects of that structure. The kinematic regions are the result of the interaction of the muscular driving forces and the the biophysical characteristics of skin tissue. Normally, when building a mathematical model of a given physiological system, one wants to separate out the plant characteristics from the control signals that are instantiated in the muscle activity. The present model, on the other hand, provides a lumped representation of the facial biomechanics.

Acknowledgments

This work has been supported by FINATEC, MCT/CNPq (Brazil), the National Institute of Deafness and Other Communications Disorders (Grant DC-05774), and NSERC (Canada).

References

[1] H. Davis and S. R. Silverman, eds. *Hearing and Deafness*. Holt, Rinehart and Winston, New York, 1970.

[2] P. Ekman, W. V. Friesen, and J. C. Hager. *The Facial Action Coding System*. Research Nexus eBook, Salt Lake City, 2002.

[3] S. D. Goldinger. Words and voices: Episodic traces in spoken identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cognit.*, 22:1166–1183, 1996.

[4] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996.

[5] B. Knappmeyer, I. M. Thornton, and H. H. Blthoff. The use of facial motion and facial form during the processing of identity. *Vision Research*, 43:1921–1936, 2003.

[6] J. C. Lucero, A. R. Baigorri, and K. G. Munhall. Data-driven facial animation of speech using a qr factorization algorithm. In *Proc. 7th Int. Sem. Speech Prod.*, pages 135–142, 2006.

[7] J. C. Lucero and K. G. Munhall. Analysis of facial motion patterns during speech using a qr factorization algorithm. *J. Acous. Soc. Am.*, 2008 (in press).

[8] A. Miller. *Subset Selection in Regression*. Chapman & Hall/CRC, Boca Raton,2002.

[9] D. Terzopoulos and K. Waters. Physically-based facial modeling, analysis, and animation. *J. Visual. Comp. Animat.*, 1:73–80, 1990.

[10] J. Yen and L. Wang. Simplifying fuzzy rule-based models using orthogonal transformation methods. *IEEE Trans. Syst. Man Cyb. B*, 29:13–24, 1999.