



# Uncontrolled Manifolds in vowel production: Assessment with a biomechanical model of the tongue

Andrew Szabados<sup>1</sup>, Pascal Perrier<sup>1,2</sup>

<sup>1</sup> University of Grenoble Alpes, Gipsa-lab, F-38000 Grenoble, France

<sup>2</sup> CNRS, Gipsa-lab, F-38000 Grenoble, France

andrew.szabados@gipsa-lab.grenoble-inp.fr, pascal.perrier@gipsa-lab.grenoble-inp.fr

## Abstract

Motor equivalence is a key feature of speech motor control, since speakers must constantly adapt to various phonetic contexts and speaking conditions. The Uncontrolled Manifold (UCM) idea offers a theoretical framework for considering motor equivalence. In this framework coordination among motor control variables is separated into two subspaces, one in which changes in control variables modify the acoustic output and another one in which these changes do not influence the output. Our work develops this concept for speech production using a 2D biomechanical model of the tongue, coupled with a jaw and lip model, for vowel production. We first propose a representation of the linearized UCM based on orthogonal projection matrices. Next we characterize the UCMs of various vocal tract configurations of the 10 French oral vowels using their perturbation responses. We then investigate whether these UCMs describe phonetic classes like phonemes, front/back vowels, rounded/unrounded vowels, or whether they significantly vary across representatives of these different classes. We found they clearly differ between rounded and unrounded vowels, but are quite similar within each category. This suggests that similar motor equivalence strategies can be implemented within each of these classes and that UCMs provide a valid characterization of an equivalence strategy.

**Index Terms:** speech motor control, motor equivalence, perturbations

## 1. Introduction

Motor equivalence is a key characteristic of motor control in animals and in humans. It corresponds to the capacity to accomplish the same motor task with different control strategies, associated with different body positions, or the use of different effectors, or different coordinations between some specific effectors. This capacity is essential for the Central Nervous System to be able to deal with changes in external constraints applied to the body before or during the execution of a movement, or with perturbations of the normal achievement of the task due in particular to pathologies.

Motor equivalence phenomena exist in speech production and they have been documented in number of studies [1],[2],[3]. They are crucial for speech production, since they enable speakers to produce intelligible speech under a broad spectrum of conditions, including various phonetic contexts, speaking with a pen in the mouth, speaking while eating, or speaking while running, etc. Motor equivalence is at the basis of the famous phenomenon of coarticulation that makes speech segments highly variable even for an individual speaker depending on the preceding and following sounds [4],[5] and

on the speaking rate [6]. In this paper we consider a mathematical framework, the Uncontrolled Manifold [7], in order to deal with motor equivalence mechanisms, and evaluate them in the context of the production of vowels.

The Uncontrolled Manifold concept was proposed by Scholz and Schöner [7] in the context of arm motor control. They defined it as “the set of [shoulder, elbow, and wrist] joint configuration[s] that leaves the end-effector position invariant” [7, p. 291]. It is conceived as a way to interpret patterns of variability that have been observed as early as Bernstein [8] in various motor tasks. The hallmark of this pattern is that variability is highly structured, specifically in a way that enables the control of its interference with task-dependent objectives. In this work we seek to address the appropriateness of the UCM idea for speech motor control, and specifically whether UCMs could provide a strategy for dealing with motor equivalence in the context of speech. To do this we take one particular potential utility of motor equivalence - that is, the facilitation of an efficient perturbation response - and test whether UCMs are useful in the context of speech motor control.

The idea that underlies our methodology is that, in order for UCMs to be useful in speech motor control, they must 1) exist and 2) generalize to a certain degree so as to be compatible with the particular nature of speech motor control that is linked with the fact that goals are related to linguistic categories such as phonemes or syllables. One way in which UCMs could be useful is in providing an advantage in perturbation response, for example by providing a correction to a command perturbation that restores the original output without returning to the initial configuration that the perturbation could prevent from reaching.

In this aim we have used a realistic model of the vocal tract associating a biomechanical model of the tongue coupled with functional models of the jaw and of the lips. This model is controlled by a number of control variables either generating muscle forces in the tongue model or controlling the spatial position in the jaw and the lips. In the context of this model we propose a way to represent UCMs and then run an experiment in which we test their efficacy in dealing with local perturbations during the production of French oral vowels.

In this paper we will first provide details about the methodology, before presenting results related to the efficacy of various accounts of the UCMs in dealing with perturbations. Finally we will discuss the usability of the UCM concept in the context of the specificities of speech motor control.

## 2. Methodology

The methodology is based on the use of a model of the vocal tract that enables finding realistic motor command patterns for vowel production. Uncontrolled Manifolds were estimated on the basis of these motor command patterns by applying local perturbations to the commands. Different UCMs were proposed by grouping vowel configurations into various categories, for example phonemic categories or phonetic categories based on such distinguishing features openness, rounding, or front/back position. Finally the various UCMs were evaluated with respect to their capacity to compensate for local perturbations of the command patterns.

### Calculating Uncontrolled Manifolds (UCMs)

#### 2.1.1. Simulating vowels

We used a 2D model of the vocal tract that includes a 2D biomechanical model of the tongue [9],[10], and 2D geometrical models of the lips, and jaw [11] to simulate the tongue-jaw-lip configuration resulting from the application of 9 motor commands. The tongue is controlled with 6 variables determining the forces generated by the 6 muscles that most significantly influence the tongue shape is the mid-sagittal plane, namely the Posterior and Anterior Genioglossus, the Hyoglossus, the Styloglossus, the Inferior Longitudinalis and the Verticalis. The force generation mechanism is based on the  $\lambda$ -model proposed by Feldman in the context of the Equilibrium Point Hypothesis for motor control [12]. The muscle commands are called  $\lambda$  commands henceforth. The additional commands control the lip protrusion, lip rounding, and the jaw aperture. The characteristics of the 9 control parameters are listed in Table 1. The model uses the 2D shape of the vocal tract formed by the resulting configuration to estimate the volume of the vocal tract (also called area function, [13]), and then acoustic features (the first three spectral peaks, also called “formants” – denoted F1, F2, F3) thanks to a harmonic model of the vocal tract [14]. Specifically, the first three formants are known to be sufficient to distinguish between the vocalic phonemes of a language. The 2D vocal tract model was used to generate around 50000 vocal tract configurations, based on a random sampling of the 9 dimensional motor space (see [15] for details about the methodology). For each of these configurations the corresponding formant pattern was computed, in such a way that it was possible to label them in terms of one of the 10 French oral vowels. Using this method we randomly selected a sample set of 8 representative tongue-jaw-lip configurations of each of the 10 French oral vowels.

| Motor Command                      | Units   | Dynamics  |
|------------------------------------|---------|-----------|
| Posterior genioglossus (GGP)       | mm      | EP        |
| Anterior genioglossus (GGA)        | mm      | EP        |
| Hyoglossus (Hyo)                   | mm      | EP        |
| Styloglossus (Sty)                 | mm      | EP        |
| Verticalis (Vert)                  | mm      | EP        |
| Inferior longitudinalis (IL)       | mm      | EP        |
| Jaw opening (Jaw)                  | degrees | set point |
| Lip rounding (Lip <sub>r</sub> )   | degrees | set point |
| Lip protrusion (Lip <sub>p</sub> ) | cm      | set point |

Table 1. Summary of motor commands. EP stands for equilibrium point dynamics following the  $\lambda$ -model.

#### 2.1.2. Calculating Jacobians and linearized UCMs

Mathematically UCMs are described by Scholz and Schöner [7] in their local linearized form, as “the null space” of the Jacobian. The Jacobian is a matrix composed of the partial derivatives of the output values of the motor plant with respect to the control variables. The null space is the subspace of the motor command space in which changes to the commands or combinations in changes of commands result in no change in output values. Simply, this is the space in which commands can be changed with no effect on output values and is thus related to the notion of motor equivalence in that a point’s null space is the set of points it shares an output with.

In the context of the 2D model of the vocal tract, for each configuration we applied a series of perturbations of magnitude 1 (in the respective units mentioned in Table 1) to the motor commands that allowed us to calculate using backward difference method the Jacobian matrix,  $J$ , which represents how changes in each of the motor commands affects each of the formants. Since our motor command space is 9-dimensional and the acoustic output space is 3-dimensional, the Jacobian is a  $3 \times 9$  matrix each entry of which is the partial derivative of the corresponding formant with respect to the corresponding motor command. For each set of motor commands we then calculated the null space of the Jacobian. It represents the linearized Uncontrolled Manifold [7] estimated at a given point in motor command space. We refer to this as the IUCM to distinguish it from the global, non-linear UCM which it locally approximates. Specifically, for each vocal tract configuration we obtained an orthonormal basis set,  $u$ , of the IUCM using singular value decomposition. For all configurations the dimensionality of the IUCM was 6, meaning it was a 6-dimensional hyperplane embedded in the 9-dimensional command space. Thus  $u$  is set of 6 9-dimensional vectors which can be represented as a  $9 \times 6$  matrix.

#### 2.1.3. Projection matrices

When a perturbation, for example one affecting the position of the jaw, is applied to the usual articulatory pattern of a sound, its compensation requires the Central Nervous System to find another set of motor commands that will take into account the new jaw position and still generate the same formants as the original command pattern. Mathematically, this compensation means projecting, in the space of the control variables, the perturbed commands back into the IUCM of the original position. We thus devised a strategy to characterize the IUCM itself via its orthogonal projection matrix. Using the orthonormal basis of the IUCM,  $u$ , we calculated the orthogonal projection matrix  $P$  for each configuration as  $P = uu^T$ , which is a  $9 \times 9$  matrix. This representation of the IUCM has several advantages. First, it is a unique representation of the IUCM, which is not the case with the basis-set representation. Indeed, swapping columns of  $u$  would lead to a different matrix that represents the same space. Second, its geometric interpretation fits nicely with the proposed utility of a UCM - application of  $P$  is equivalent to embedding a perturbation into the IUCM at the point in the IUCM that is closest to the perturbation. Third, its columns (or, by symmetry, rows) can be interpreted as the response to a perturbation of magnitude one in the corresponding dimension of the command space. The role of the projection matrix in the compensation for a perturbation to the motor commands is summarized in Figure 1 in the simple case of a 2D motor command space.

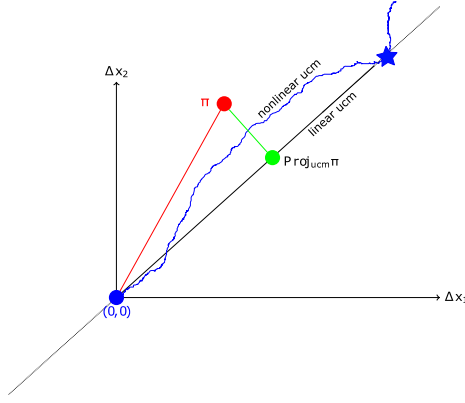


Figure 1: Diagram summarizing the role of the projection matrix for a 2D example. The plane represents the command space (space of the control variables). The origin (0, 0) is the original point in command space and the axes characterize the deviations along each of the motor variables from the unperturbed command pattern. The IUCM at this point is represented by the straight line passing through the origin and the blue star, the point at which the IUCM was estimated. A command perturbation,  $\pi$ , results in change in acoustic output (since  $\pi$  does not fall on the 1D IUCM shown here as ucm). Orthogonal projection matrix (here  $\text{Proj}_{\text{ucm}}$ ) projects  $\pi$  onto ucm at the closest point on ucm to  $\pi$ . At the resulting point  $\text{Proj}_{\text{ucm}}\pi$ , the formant pattern should be the same as at the origin, if the linearized UCM coincides with the nonlinear UCM. Already  $s$  might not be a good approximation of the UCM for that point, and, even if it falls perfectly on the non-linear UCM, the linearized UCM it forms may, as shown above, not be optimal (i.e. minimize the area between IUCM and UCM for a given neighborhood around the origin).

## Perturbation response testing

We now have a representation of the local, linearized Uncontrolled Manifold in the form of a projection matrix for each of the 8 configurations selected for the 10 French oral vowels. Even if we assume that IUCMs are good local approximations of the nonlinear UCMs around a given point in the motor command space, the question arises whether it is necessary to know the local UCM for a specific point or if it suffices to learn a less local IUCM by grouping configurations into phonetic categories based on phonemes or phonetic features such as front/back, high/low or rounded/unrounded lips. To test this idea, we propose an experiment in which we test the perturbation response performance of UCMs of different groupings of vowel configurations.

### 2.1.4. Perturbation response groupings

In order to test the efficacy of different methods of obtaining IUCMs we first came up with several ways in which IUCMs could be defined, ranging from a local IUCM for each point in the command space, to a single, global IUCM for all points. We included different groupings based on classical phonetic features or categories as listed in Table 2. In each group, the IUCM was characterized by the average projection matrix of all the configurations falling within that group.

|   | Nr of groups | Groupings                                     |
|---|--------------|---|
| 1 | 80           | individual configurations                     |
| 2 | 10           | phoneme average                               |
| 3 | 5            | {i e}, {ε a}, {y ø}, {œ ɔ}, {u o}             |
| 4 | 3            | {i e y ø}, {ε a}, {œ ɔ u o}                   |
| 5 | 3            | {i e ε a}, {y ø u o}, {œ ɔ}                   |
| 6 | 2            | front {i y e ø ε œ} vs. back {u o ɔ a}        |
| 7 | 2            | open {ɔ a ε œ} vs. closed {i y u e ø o}       |
| 8 | 2            | rounded {y ø œ u o ɔ} vs. unrounded {i e ε a} |
| 9 | 1            | global average                                |

Table 2. Groupings used for perturbation test

For each configuration and each IUCM grouping, we applied a series of 27 perturbations - one of each command at 3 different magnitudes, 0.1, 1.0, and 2.0. This gives  $10 \times 8 \times 27 = 2160$  different perturbed configurations in the motor space. We then applied the projection matrix to each perturbed configuration in order to compensate for the perturbation on the formant pattern and simulated the tongue configurations and acoustic outputs for each perturbed and the corrected configurations.

## 3. Results

The efficacy of the IUCMs determined for different groupings was assessed by computing the error in the 3 dimensional formant space between the formant pattern obtained for each of the selected configurations and the formant pattern obtained after the perturbation and the correction based on the projection matrix (i.e. the difference between the formant pattern at point (0,0) and at point  $\text{Proj}_{\text{ucm}}\pi$  of Figure 1). Since the error in F1 alone scales similarly, we just report F1 error for simplicity. Differences in grouping correspond to differences in the projection matrix used for the correction. We calculated the root mean square error (RMSE, Equation (1)) for each IUCM group once with all data points and once after removing outliers, identified as those simulations for which the F1 error was greater than  $2\sigma$ , where  $\sigma$  is the standard deviation of the residual error data for the group 1 (individual configuration) UCMs. Outliers here are considered to be points for which the grouped, non-local IUCM poorly represents the local IUCM. The more frequent their occurrence the worse the linear approximation.

$$RMSE_{F1Hz} = \sqrt{\sum_{v \in V} (v_{F1 \text{ unperturbed}} - v_{F1 \text{ corrected}})^2} \quad (1)$$

where V is the set of 2160 compensated vowels.

| Group | RMSE <sub>F1Hz</sub> | RMSE <sub>F1Hz</sub> outliers removed | Outliers |
|-------|----------------------|---------------------------------------|----------|
| 1     | 15.9657              | 6.1046                                | 1.9%     |
| 2     | 13.9277              | 5.2069                                | 1.6%     |
| 3     | 11.3709              | 4.7433                                | 2.1%     |
| 4     | 53.1547              | 4.8963                                | 1.8%     |
| 5     | 9.6227               | 5.8048                                | 1.1%     |
| 6     | 73.2936              | 4.4579                                | 2.0%     |
| 7     | 10.7325              | 5.2120                                | 2.7%     |
| 8     | 7.4654               | 5.2031                                | 1.1%     |
| 9     | 37.1854              | 5.0515                                | 2.5%     |

Table 3. RMSE data computed for all the selected vowels and all the perturbations (i.e. for 2160 perturbed configurations)

## 4. Discussion

We are interested in evaluating whether UCMs could be useful in speech motor control, in particular as part of a perturbation response strategy. Trivially, returning to the original unperturbed point is a way to compensate for a perturbation. However, for most perturbations it is not possible to do so and even if it is, it may not be optimal. Hence, it could be advantageous to exploit the UCM - and thus the underlying principle of motor equivalence - to find another strategy which nullifies the acoustic effect of the perturbation, either because the perturbation may preclude returning to the original point or because this other strategy requires less effort. For this, we first clarify what we mean by *efficiency* of the IUCMs we have devised. Namely, it is the distance in the command space from the configuration after perturbation to the UCM (distance from  $\pi$  to  $\text{Proj}_{\text{ucm}}\pi$ , on Figure 1, called hereafter "compensation") relative to the magnitude of the perturbation. However, it is not enough to find the closest point on the UCM. This point must actually nullify the acoustic effects of the perturbation. And this capacity, that we call *effectiveness*, is highly dependent on the adequacy of the linear approximation of the UCM (see Fig. 1). In a second stage we evaluate effectiveness on the basis of the standard deviation of the residual error as given in Table 3. Ideally, a useful perturbation response strategy is both efficient and effective and here we evaluate whether this is the case for IUCMs in our experiment.

### Efficiency of IUCMs

A first question is whether linearized UCMs have any theoretical advantages over the trivial solution of resisting and opposing perturbations directly. To measure this, we can look at the projection matrices directly and calculate their efficiency, defined as the amount of compensation required to neutralize the effects of a perturbation as a fraction of the magnitude of the perturbation. Since the projection matrices represent the perturbation response assuming linear interactions between perturbations on individual commands the efficiency can be calculated independently of any simulations and depends only on the set of perturbations.

|   | Compensation | Distance |
|---|--------------|----------|
| 1 | 0.55777      | 1.3659   |
| 2 | 0.45469      | 1.2393   |
| 3 | 0.44137      | 1.2267   |
| 4 | 0.43628      | 1.2215   |
| 5 | 0.43033      | 1.2163   |
| 6 | 0.43107      | 1.2171   |
| 7 | 0.42905      | 1.2149   |
| 8 | 0.42847      | 1.2146   |
| 9 | 0.42534      | 1.2115   |

Table 4. *IUCM efficiency*. Compensation calculated as the magnitude of the correction (distance between red ( $\pi$ ) and green ( $\text{Proj}_{\text{ucm}}\pi$ ) points in Fig.1) divided by the magnitude of the perturbation (length of red line ( $\pi$ ) in Fig. 1). Distance is resultant perturbation magnitude calculated as distance from compensated point (green point,  $\text{Proj}_{\text{ucm}}\pi$  in Fig. 1) to the original point. Although the compensated point is on the linear UCM, it is not at the original point unless the UCM was inefficient and required complete compensation of the same magnitude as the perturbation. Data is for 10000 random perturbations.

### Effectiveness and robustness of IUCMs

The primary effect of interest is the ability of a IUCM grouping to allow for a perturbation response that is effective (reasonable average error) and robust (not prone to large errors). What we observe from the data in Table 3 is that Group 8 (rounded vs. unrounded) showed the lowest  $\text{RMSE}_{\text{F1Hz}}$  when outliers were left in and the lowest number of outliers. With outliers removed it was no longer the best, but was still comparable to the best performing groups. From this we can essentially conclude that all the groups have a similar effectiveness with the problem that some groups have more of a tendency than others to give inappropriate perturbation responses, which accounts for the major discrepancy between the performance of different groups. We can say that all the groupings work similarly well when they do work, but that group 8 and 5 fail to work least frequently.

### Efficiency vs. Effectiveness tradeoff

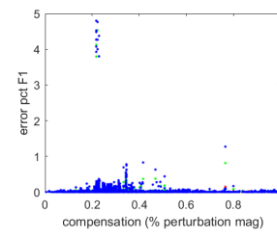


Figure 3: *Residual error in F1 (Hz) increases with decreased compensation, i.e. increased efficiency.*

Indeed, we see (Fig. 3) a tradeoff between efficiency and effectiveness, with the greatest error coming from outliers, which are actually failed compensations (due to inadequate linear approximation of the true, nonlinear UCM). Thus, ultimately, the best IUCM grouping is that which minimizes the likelihood of failed compensations, as in group 8.

### Secondary effect

We do notice a secondary effect in that the  $\text{RMSE}_{\text{F1Hz}}$  with outliers removed tends to be smaller for groups with fewer groupings. This seems to be related to the distance of the compensated point from the original point being lower with fewer groupings as well (Table 4, Distance) and can be explained as a consequence of the average compensation for different UCM groupings (Table 4, Compensation).

### Conclusion

The uncontrolled manifold concept has been shown to be very helpful to understand motor equivalence in pointing tasks. Our study shows that motor equivalence strategies based on a global average UCM for all the vowel configuration would also be relatively efficient in the large majority of the cases. This suggests that the same synergies and antagonisms between tongue muscles, lip shaping, and jaw positioning apply to most of the vowel configurations. We also observed that distinguishing rounded and unrounded vowels provides the most robust way to define UCM. This is consistent with the fact that formant sensitivity to changes in motor variables is highly dependent on the size of the lip area. This suggests that in terms of compensatory strategies the phonetic feature rounded/unrounded is the most appropriate one, along which motor equivalence strategies should be accurately defined.

## 5. References

- [1] Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica*, 33(3), 199-221.
- [2] Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel/u: A pilot "motor equivalence" study. *The Journal of the Acoustical Society of America*, 93(5), 2948-2961.
- [3] Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, 98(5), 2428-2442.
- [4] Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, 5(2), 115-133.
- [5] Whalen, D. H. (1990). Coarticulation is largely planned 7/3. *Journal of Phonetics*, 18, 3-35.
- [6] Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, 44(2), 340-353.
- [7] Scholz, J. P., & Schöner, G. (1999). The uncontrolled manifold concept: identifying control variables for a functional task. *Experimental brain research*, 126(3), 289-306.
- [8] Bernstein, N.A. The Co-ordination and Regulation of Movements. Oxford: Pergamon Press, 1967
- [9] Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, 114(3), 1582-1599.
- [10] Payan, Y., & Perrier, P. (1997). Synthesis of VV sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech communication*, 22(2), 185-205.
- [11] Zandipour, M., Perkell, J., Guenther, F., Tiede, M., Honda, K., & Murano, E. (2006). Speaking with a bite-block: Data and modeling. In *Proceedings of the 7th International Seminar on Speech Production* (pp. 361-368).
- [12] Feldman, A. G., & Levin, M. F. (1995). The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences*, 18(04), 723-744.
- [13] Perrier, P., Boë, L. J., & Sock, R. (1992). Vocal tract area function estimation from midsagittal dimensions with ct scans and a vocal tract cast modeling the transition with two sets of coefficients. *Journal of Speech, Language, and Hearing Research*, 35(1), 53-67.
- [14] Badin, P., & Fant, G. (1984). Notes on vocal tract computation. *Quarterly Progress and Status Report, Dept for Speech, Music and Hearing, KTH, Stockholm*, 53-108.
- [15] Perrier, P., Ma, L. & Payan, Y. (2005). Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. *Proceedings of Interspeech 2005* (pp. 1041 – 1044), International Speech Communication Association, Lisbon, Portugal.