Optimal speech motor control and token-to-token variability: a Bayesian modeling approach

Jean-François Patri, Julien Diard, Pascal Perrier

Noname manuscript No.
(will be inserted by the editor)

Optimal speech motor control and token-to-token variability: A Bayesian modeling approach

Jean-François Patri · Julien Diard · Pascal Perrier

Received: date / Accepted: date

Abstract The remarkable capacity of the speech motor system to adapt to various speech conditions is due to an excess of degrees of freedom, which enables producing similar acoustical properties with different sets of control strategies. To explain how the Central Nervous System selects one of the possible strategies, a common approach, in line with optimal motor control theories, is to model speech motor planning as the solution of an optimality problem based on cost functions. Despite the success of this approach, one of its drawbacks is the intrinsic contradiction between the concept of optimality and the observed experimental intra-speaker token-to-token variability. The present paper proposes an alternative approach by formulating feedforward optimal control in a probabilistic Bayesian modeling framework. This is illustrated by controlling a biomechanical model of the vocal tract for speech production and by comparing it with an existing optimal control model (GEPPETO). The essential elements of this optimal control model are presented first. From them the Bayesian model is constructed in a progressive way. Performance of the Bayesian model is evaluated based on computer simulations and compared to the optimal control model. This approach is shown to be appropriate for solving the speech planning problem while accounting for variability in a principled way.

Keywords Speech motor control · Speech sequence motor planning · Bayesian modeling · Optimal motor control

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 339152, "Speech Unit(e)s", PI: Jean-Luc-Schwartz).

Jean-François Patri
Univ. Grenoble Alpes, Gipsa-lab, F-38000 Grenoble, France
CNRS, Gipsa-lab, F-38000 Grenoble, France
11 Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France
Tel.: +33 (0)4 76 57 48 48
E-mail: Jean-Francois.Patri@gipsa-lab.grenoble-inp.fr

Julien Diard
Univ. Grenoble Alpes, LPNC, F-38000 Grenoble, France
CNRS, LPNC, F-38000 Grenoble, France
1251 Avenue Centrale, 38400 Saint-Martin-d'Hères

Pascal Perrier
Univ. Grenoble Alpes, Gipsa-lab, F-38000 Grenoble, France
CNRS, Gipsa-lab, F-38000 Grenoble, France

1 Introduction

Motor control aims at finding patterns of activation that agents should generate in their articulatory chain in order to achieve desired motor goals. This is in essence an ill-posed problem, since degrees of freedom of articulatory chains often largely exceed the degrees of freedom of the task. Therefore there is a multiplicity of possible solutions for achieving the desired motor goal. Optimal motor control theories aim at resolving this well-known redundancy problem (Jordan, 1996; Uno et al, 1989) by considering a cost function that attributes to each possible solution a certain performance value. The redundancy is resolved by this mean, if there is a unique solution that optimizes the value associated with this criterion.

A crucial consequence of this approach is that the resulting behavior of the controlled system is stereotyped. This means that for a given task and in a specified condition, the optimal solution is always the same and no trial-to-trial variability can be obtained. While this may be desirable in engineering applications, it is

a major drawback for models that try to reproduce behavior of biological agents. In that sense, most optimal motor control models have been successful in accounting for average patterns of behavior at the expense of not being able to model trial-to-trial variability (see Todorov (2004) for a review). Even though this issue has been addressed by stochastic optimal feedback control (Todorov and Jordan, 2002) by explicitly considering feedback in the planning process (leading to a closed loop control), this approach only concerns movements that can use on-line feedback information. In the case of very fast movements though, ongoing control is unlikely to rely on feedback, due to delays in afferent signals (for example 30–100 ms for visual motion in Schmolesky et al (1998); 99–143 ms for auditory feedback in speech processing in Tourville et al (2008)), and control is rather assumed to be performed through an open loop planning (see for example Kawato (1999)).

Nevertheless, two approaches are usually considered for recovering variability under a feedforward optimal control model. The first approach assumes that optimal planning is stereotyped, but that variability arises from noise in the pathway of the control signal or in the dynamics of the articulatory chain. This is the approach followed by stochastic optimal control theory. The second approach assumes that planning is driven by optimality, but that its realization does not systematically lead to the unique optimal situation. It should be noted that even if the control process is certainly subject to stochastic dynamics, the first approach alone is not satisfactory for explaining situations where variability results from different specific patterns of behavior (see for example Shim et al (2003) for prehension tasks). Rather, the second approach better accounts for systematic deviations from a single optimal solution. For instance the role of motor memory on convergence to locally optimal solutions rather than a global optimum has been suggested as a crucial aspect of motor control (Ganesh et al, 2010).

These questions are of particular interest in speech motor control. Indeed, it is unlikely that speech control relies primarily on feedback signals, due to the speed of tongue movement (Perkell et al, 1997). Yet, trial-to-trial variability is observed in phoneme production at the acoustic, articulatory and muscle activation levels (Perkell and Nelson, 1985). This variability is underpinned by the presence of redundancy at the three levels described above: 1) a particular phonemic goal does not correspond to a unique point in the acoustic domain since different acoustic signals are perceived as a unique phonemic category, 2) a particular acoustic signal can be produced by different vocal tract configurations, and 3) a particular vocal tract configuration can be attained by different patterns of muscle activation.

Attempts at modeling feedforward speech motor control based on optimal control theory have been able to reproduce a number of experimental speech patterns (Ma et al, 2006; Perrier et al, 2005; Guenther et al, 1998; Guenther, 1995). However, while these results are consistent with average values among and across subjects, they fail at accounting for individual trial-to-trial variability from a theoretical point of view.

The present work aims at addressing this issue. By formulating optimal control in a Bayesian modeling framework we suggest that both variability and selection of motor control variables in speech production can be obtained in a principled way, from uncertainty at the representational level and without resorting solely to stochastic noise in the dynamics. We illustrate this approach by presenting a Bayesian formulation of an optimal control model for speech motor planning (Perrier et al, 2005).

The remainder of this paper is divided into four sections. Section 2 describes the essential ingredients of GEPPETO, the optimal control model that we aim to reformulate. From these ingredients the Bayesian model is then introduced in Section 3, in two steps. The first step consists in a Bayesian model inferring motor control variables for the production of a single phoneme. The second step consists in the complete Bayesian formulation of GEPPETO, that is, a Bayesian model planning optimal motor control variables for the production of sequences of phonemes. In Section 4 we compare and discuss the main results of the Bayesian formulation with respect to its optimal control version.

2 Main ingredients of GEPPETO

This section summarizes the key components of GEPPETO, the optimal control model for speech motor planning that we aim at reformulating in the Bayesian framework. The following description focuses on the main hypotheses and we refer the reader to Perrier et al (2005) for a more detailed description of the model. We structure them as follows.

***H*$_1$:** GEPPETO computes the motor control variables of a biomechanical model of the tongue (Payan and Perrier, 1997; Perrier et al, 2003) in order to produce a desired sequence of speech gestures. GEPPETO defines a speech sequence as a succession of fundamental phonological units, corresponding to phonemes of the considered language. As the model only includes an account of the tongue, only phonemes that do not

require lip rounding are considered. The set of considered phonemes is therefore {/i/, /e/, /ɛ/, /a/, /oe/, /ɔ/, /k/}. The first hypothesis of GEPPETO is therefore:

(H_1) The motor control of a speech sequence is organized on the basis of the specification of motor goals that are related to phonemes.

H_2: GEPPETO supposes that phonemes are characterized and controlled in the acoustic domain and that the acoustic signal is characterized by the value of the first 3 peaks of the spectral envelope (these peaks are called "formants"). Therefore:

(H_{2a}) The acoustic signal is represented by a point in a 3 dimensional space.

(H_{2b}) Phonemic goals are represented by specific simply connected regions of the 3-dimensional acoustic space. These regions are accounted for by ellipsoids defined through dispersion regions measured from phoneme production experiments (Robert-Ribes, 1995; Ménard, 2002; Calliope, 1984). These regions are represented in Figure 1 by their projections on the (F_2, F_1) and (F_2, F_3) planes.

H_3: The biomechanical model on which GEPPETO is based consists of a finite element structure representing the projection of the tongue on the mid-sagittal plane. Six principal muscles are considered as actuators for shaping the tongue. Figure 2 represents the tongue configuration at rest as well as the fibers of one of the considered muscles (the posterior genioglossus). The resulting tongue shape corresponds to the mechanical equilibrium of the forces generated by each muscle; the activation of each muscle is specified through a λ parameter, which specifies the muscle length above which active muscle force is generated [1], in agreement with the Equilibrium Point Hypothesis (Feldman, 1986). Hence:

(H_3) Achieving a particular configuration of the tongue consists in specifying a point in the 6-dimensional control space $(\lambda_1, \ldots, \lambda_6)$.

[1] Muscle force F generated by the biomechanical model is specified as

$$F = \rho[\exp(cA) - 1], \quad (1)$$

where c is a form parameter accounting for the gain of the feedback from the muscle to the motoneurons pool and ρ a magnitude parameter directly related to force-generating capability. A is the muscle activation corresponding to

$$A = l - \lambda + \mu\dot{l}, \quad (2)$$

where l is the actual muscle length, $\dot{l}$ the muscle shortening or lengthening velocity and μ a damping coefficient due to proprioceptive feedback (Payan and Perrier, 1997).

H_4: For every tongue configuration the resulting acoustic signal is generated from the computation of the vocal tract volume (via a model that links mid-sagittal views and cross-sectional areas from the glottis to the lips (Perrier et al, 1992)). For every point in the 6-dimensional control space there is therefore a unique associated point in the 3-dimensional acoustic space. Furthermore, GEPPETO assumes that:

(H_4) The knowledge of the mapping from the control variables to the acoustic domain is stored in an internal model in the Central Nervous System (CNS). It is assumed that this model results from a learning process that generalizes the relation between motor control variables and formants from a limited number of examples. This model is considered to be "static" as it associates motor control variables and outputs at targets. It is implemented through a Radial Basis Function (RBF) network (Poggio and Girosi, 1989). This neural network is learned through classical supervised learning.

H_5: GEPPETO assumes that articulatory trajectories between two successive targets emerge from the interactions between the motor control variables at targets, the specified duration of the transition between targets and the biomechanical properties of the tongue. It is important to note that GEPPETO does not assume any kind of specification of desired trajectories or any optimization of a cost at the level of the trajectories (such as maximum velocity, jerk, total amount of force). The specification of the appropriate control variables for the generation of a sequence does not involve inverse dynamics. Once the motor control variables at the targets are specified, the time variations of these variables are assumed to proceed from the value at target n to the values at target $n+1$ at a constant rate of shift. This is consistent with the suggestion made by Laboissière et al (1996).

H_6: The aim of GEPPETO is to specify in the 6-dimensional control space a discrete sequence that generates a sequence of acoustic goals at the targets that are inside the ellipsoids characterizing the motor goals of the different phonemes. This is an ill-posed problem as there is an infinity of possible trajectories reaching the desired phonemic targets. To resolve this redundancy, GEPPETO assumes that the controller selects the trajectory that is optimal in the displacement of the corresponding λ variables (i.e. motor control space) . To this end, GEPPETO defines a cost function measuring the distance between control variables over the whole set of targets within the sequence. For a three-phoneme sequence, which will be our focus for the re-

mainder of the text, this corresponds to the perimeter of the triangle defined by the three control points [2]. A subtlety about the interpretation of this cost function should be mentioned here. The cost function introduces a term relating the first and last phonemes in the sequence. This could be interpreted as sequence planning where only neighbors phonemes would influence each other, and where there would be a return, from the last phoneme, to the first one. However this is not what it is intended here. The term relating the first to the last phoneme is introduced in order to model forward and backward planned coarticulation influences. In this sense, in a sequence of 3 phonemes, this cost function introduces dependencies between items independently of their relative order in the sequence, allowing influence of every phoneme on every other one.

(H_{6a}) For a three-phoneme sequence, the planning problem consists in finding a set of three points in the 6-dimensional control space that minimizes the perimeter of the triangle they define, under the perceptual constraint that the corresponding spectral properties of the signal are within the ellipsoids regions assigned to each phoneme in the sequence.

(H_{6b}) The optimization process is performed by a gradient descent algorithm where the perceptual constraint is specified as an additional cost that vanishes whenever the corresponding acoustic signal falls within the correct ellipsoid region and goes to infinity (in practice, a large number) otherwise.

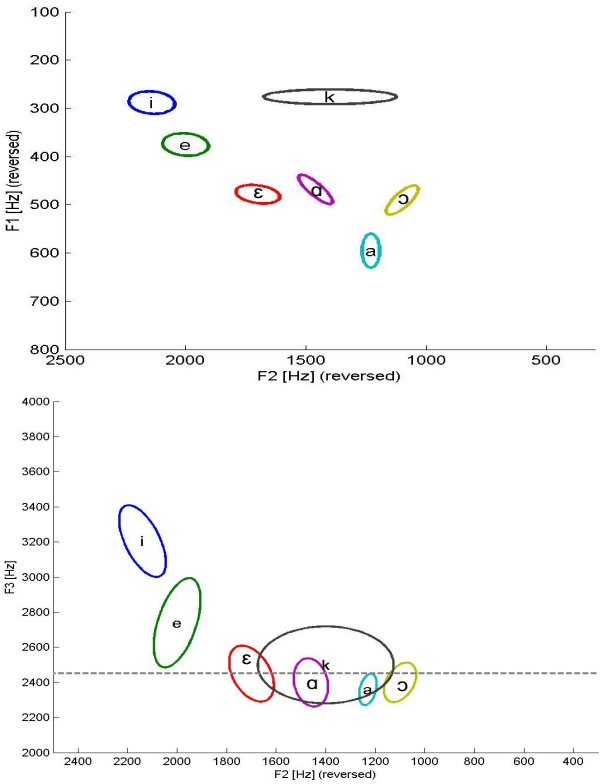

Fig. 1 Projections of the 3-dimensional dispersion ellipsoids corresponding to each target region characterizing phonemes. Top: (F_2, F_1) plane; Bottom: (F_2, F_3) plane. The dotted line on the bottom image indicates the F_3 value specified in Figures 4 and 5. (/oe/ is noted /ɑ/ in all figures for sake of simplicity in graphic representation)

3 Bayesian formulation of GEPPETO

This section presents the Bayesian framework within which the reformulation of the optimal control model described in the previous section is derived. The approach is based on the Bayesian Programming methodology (Lebeltel et al, 2004; Bessière et al, 2013, 2008) that proposes a precise structure for the definition of a Bayesian model. In order to illustrate the framework and to derive the general model stepwise, we first expose a sub-model, aiming at generating the control variables for the production of a single phoneme, before focusing on the generation of sequences of phonemes.

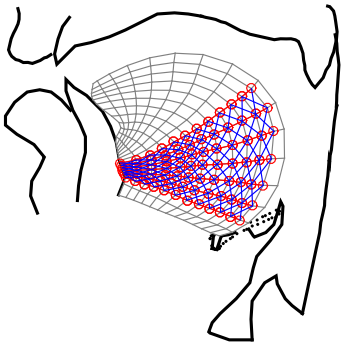

Fig. 2 Biomechanical model of the tongue. Colored lines correspond to fibers of the posterior genioglossus muscle. Crossed elements are the muscles elements and their elastic properties change with muscle activation.

[2] For simplicity, the main text presents the case of sequences of 3-phonemes, without loss of generality. For a general n-phoneme sequence the proposed cost function would correspond to the perimeter of the corresponding $(n-1)$-simplex defined by the n control variables in the 6-dimensional control space. For the present 3-phoneme case the 2-simplex corresponds to the triangle introduced in the text. Rigorously, influence of every phoneme of the sequence on every other one would be rather modeled by a cost function involving distances between every pair of phonemes. In order to avoid the corresponding quadratic combinatorial growth of the number of terms in the cost function, its definition has been simplified into the one presented here.

3.1 Bayesian model for the production of a single phoneme

3.1.1 Description

This section presents the description of the Bayesian model, obtained by translating into probabilistic terms the hypotheses and knowledge introduced above.

Variables The variables of the Bayesian model are simply extracted from the key ingredients of GEPPETO described in Section 2.

- $\boldsymbol{\Phi}$ is the variable representing all the categories of phonemes. It is a discrete variable composed by all the phonemes specified in hypothesis H_1 of Section 2. An additional "no-phoneme" category (denoted by /00/) is further assumed in order to take into account all acoustic configurations that do not fall within any of the above phonemic categories. The values taken by this variable are labeled by:

  $\Phi = \{$ /i/, /e/, /ɛ/, /a/, /oe/, /ɔ/, /k/, /00/ $\}$.

- $\boldsymbol{S}$ represents the spectral characteristics of the acoustic signal. Hypothesis H_{2a} specifies this signal as a point in a 3-dimensional formant space. Therefore, S corresponds to a continuous vector variable, $S = (F_1, F_2, F_3)$. The domain for this variable is the same as in GEPPETO and corresponds to the acoustic values attained by the simulations of the biomechanical model of the tongue.
- $\boldsymbol{M}$ represents the motor control variables controlling the articulatory configurations of the tongue. According to H_3, these control variables correspond to the six λ parameters specifying the activation threshold length of each muscle. M is therefore a continuous 6-dimensional vector variable defined by $M = (\lambda_1, \ldots, \lambda_6)$. The domain of M is specified as in GEPPETO and corresponds to the values of each λ for which the bio-mechanical model attains its equilibrium configurations, constrained by the vocal tract boundaries.

Decomposition We now define the structure of the Bayesian model, by specifying the joint probability distribution over the three above variables.

Following the chain rule, an exact decomposition of the joint probability distribution $P(M\ S\ \Phi)$ is given by:

$$P(M\ S\ \Phi) = P(M)\ P(S \mid M)\ P(\Phi \mid S\ M). \tag{3}$$

In order to avoid confusions, we draw attention to the notation that is employed here. The domain of the joint probability distribution constructed here is composed of discrete and continuous variables. Usually, one writes P for probability distributions over discrete variables and p for probability densities over continuous variables. For simplicity, we chose not to make this distinction here. Similarly, all summations and integrals are denoted by the sign $\sum$, even when rigorously it is the $\int$ sign that should be used for continous variables.

Now, due to hypothesis H_2 described in Section 2, the last term of the decomposition of Equation (3) can be simplified. Indeed, according to this hypothesis phonemes are assumed to be fully characterized by their characteristics in acoustic space. Therefore, it can be assumed that Φ is independent of M conditioned on the knowledge of S. Under this assumption the joint probability distribution becomes

$$P(M\ S\ \Phi) = P(M)\ P(S \mid M)\ P(\Phi \mid S). \tag{4}$$

Figure 3 illustrates the Bayesian network representing this decomposition.

Parametric forms

- $\boldsymbol{P(M)}$ is the prior probability distribution over motor control variables M. Since no prior knowledge is assumed about this variable, $P(M)$ is defined as a uniform probability distribution over domain D_M:

  $$P(M) = \begin{cases} \frac{1}{|D_M|} \text{ if } M \in D_M \\ 0 \quad \text{otherwise.} \end{cases} \tag{5}$$

- $\boldsymbol{P(S \mid M)}$ represents the knowledge about the spectral characteristics of the acoustic signal produced, given the equilibrium configuration of the tongue attained for the motor variables M. This knowledge is described in H_4 and corresponds to the internal model implemented by the RBF network of GEPPETO. Denoting by $S^*(M)$ the spectral properties of the acoustic signal associated to M by this RBF network, the corresponding probability distribution is assumed to be deterministic and is given by:

  $$P(S \mid M) = \delta_{S^*(M)}(S) \tag{6}$$

  where δ_a denotes the Dirac distribution centered in a. It translates the fact that $P(S \mid M)$ is zero unless $S = S^*(M)$.
- $\boldsymbol{P(\Phi \mid S)}$ corresponds to the probability of assigning phoneme Φ to the given spectral property S. It can therefore be interpreted as the confidence on a phonemic categorization of the acoustic signal. This knowledge is formulated in H_{2b} as the acoustic regions corresponding to each phoneme. Probability distribution $P(\Phi \mid S)$ is inferred by a sub-model based on a Gaussian probability distribution for $P(S \mid \Phi)$, the probability distribution of the produced spectral property S given each phoneme Φ. These probability distributions are specified by the corresponding ellipsoids described in H_{2b}. The variance of each

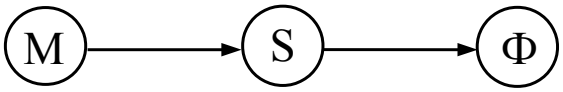

Fig. 3 Bayesian network representing the decomposition of the joint probability distribution given by Equation (4).

Gaussian distribution is controlled by a parameter, denoted by κ_S, that multiplies its variance. This allows to control the precision of the categorization task performed by the probability distribution $P(\Phi \mid S)$. This sub-model is not further described here and we refer the reader to the supplementary material for a more detailed description. For illustration, Figure 4 presents an example of probability distribution $P(S \mid [\Phi = k])$ as well as the resulting likelihood function $P([\Phi = k] \mid S) = f(S)$.

Parameter identification The last step in order to completely specify the joint probability distribution given by Equation (4) is the identification of parameter values characterizing its probability distributions. κ_S is the only parameter that remains unspecified. Its specification reflects the integration of particular assumptions in the model and its value will be given for each result. The effect of parameter κ_S on the likelihood function $P(\Phi \mid S)$ is illustrated in Figure 5.

3.1.2 Inference of control variables M for the production of a given phoneme Φ

Having specified the joint probability distribution $P(M\ S\ \Phi)$, we now formulate the question to be solved by the Bayesian model. As the problem is to infer motor control variables M producing a desired phoneme Φ, the approach consists in computing the probability distribution over M, conditioned on the specified value of Φ. The corresponding probability distribution, $P(M \mid \Phi)$, is obtained by standard Bayesian inference as

$$\begin{aligned} P(M \mid \Phi) &= \frac{P(M\ \Phi)}{P(\Phi)} \\ &= \frac{\sum_S P(M\ S\ \Phi)}{P(\Phi)} \\ &\propto \sum_S P(M\ S\ \Phi), \end{aligned} \tag{7}$$

where the proportionality symbol "$\propto$" on the last line accounts for the term $P(\Phi)$, which does not depend on M for a given Φ value. Now, using the decomposition of Equation (4) we have:

$$\begin{aligned} P(M \mid \Phi) &\propto \sum_S P(M\ S\ \Phi) \\ &\propto \sum_S P(M)\ P(S \mid M)\ P(\Phi \mid S) \\ &\propto \sum_S P(S \mid M) P(\Phi \mid S) \\ &\propto P(\Phi \mid S^*(M)), \end{aligned} \tag{8}$$

where the third line followed because $P(M)$ is assumed to be a uniform distribution and the last line is derived by recalling that $P(S \mid M)$ is zero unless $S = S^*(M)$ (see Equation (6)).

3.1.3 Implementation of the model

The aim of the model is to generate motor control variables M performing a desired phoneme Φ. The probability distribution $P(M \mid \Phi)$ characterizes every control variable M with its probability for achieving the desired phoneme Φ. The best choice would be to select control variables that maximize this probability. However this would eliminate any possible variability and lead to the stereotyped situation encountered in GEPPETO. As the aim is to preserve variability, we adopted a decision policy based on a random sampling of the control variables space from $P(M \mid \Phi)$. Accuracy of the obtained solutions is ensured in average, since with this sampling the most probable control variables correspond to the ones having high probability of achieving the desired phoneme. This sampling is implemented by a standard Markov Chain Monte Carlo algorithm (MCMC), which performs a random walk that has the desired probability distribution as its equilibrium distribution. We draw attention on the interpretation of this particular implementation of the Bayesian model. We are not assuming that the biological system is indeed performing MCMC sampling. In terms of a biological implementation of this process, one would imagine that the brain stores information about $P(M \mid \Phi)$ in some ways and would use it to optimize the mapping from phoneme to motor space.

3.1.4 Results

Figure 6 shows histograms of control variables samples M, obtained from $P(M \mid \Phi)$ as described in the previous section. As they are 6-dimensional probability distributions, they are represented by their six marginal distributions. It can be noted that each phoneme corresponds to a specific set of distributions of control variables λ. Some of these control variables appear to be constrained within small ranges of values, for instance

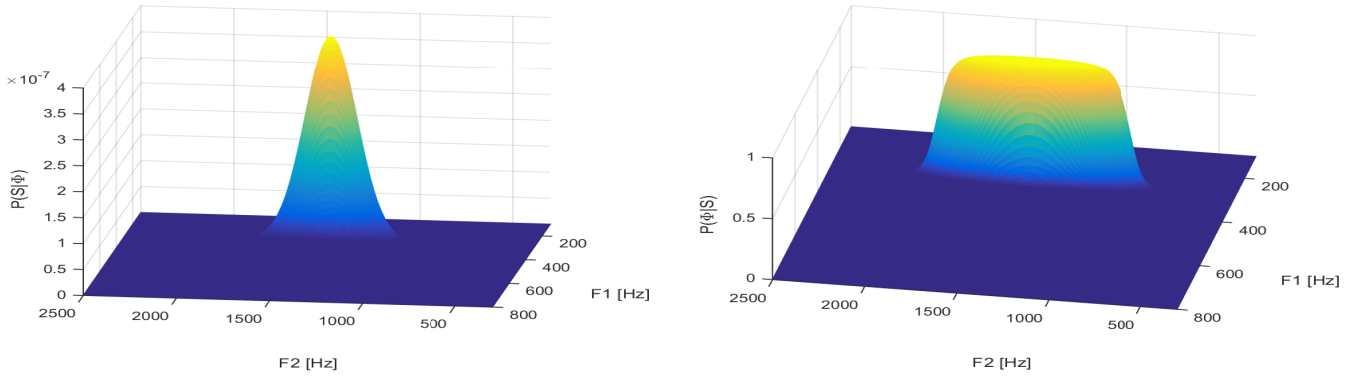

Fig. 4 Values of the probability distribution $P(S \mid [\Phi = k])$ (left) and the corresponding likelihood function $P([\Phi = k] \mid S)$ (right), projected on the (F_1, F_2) plane defined by $F_3 = 2,450\ Hz$.

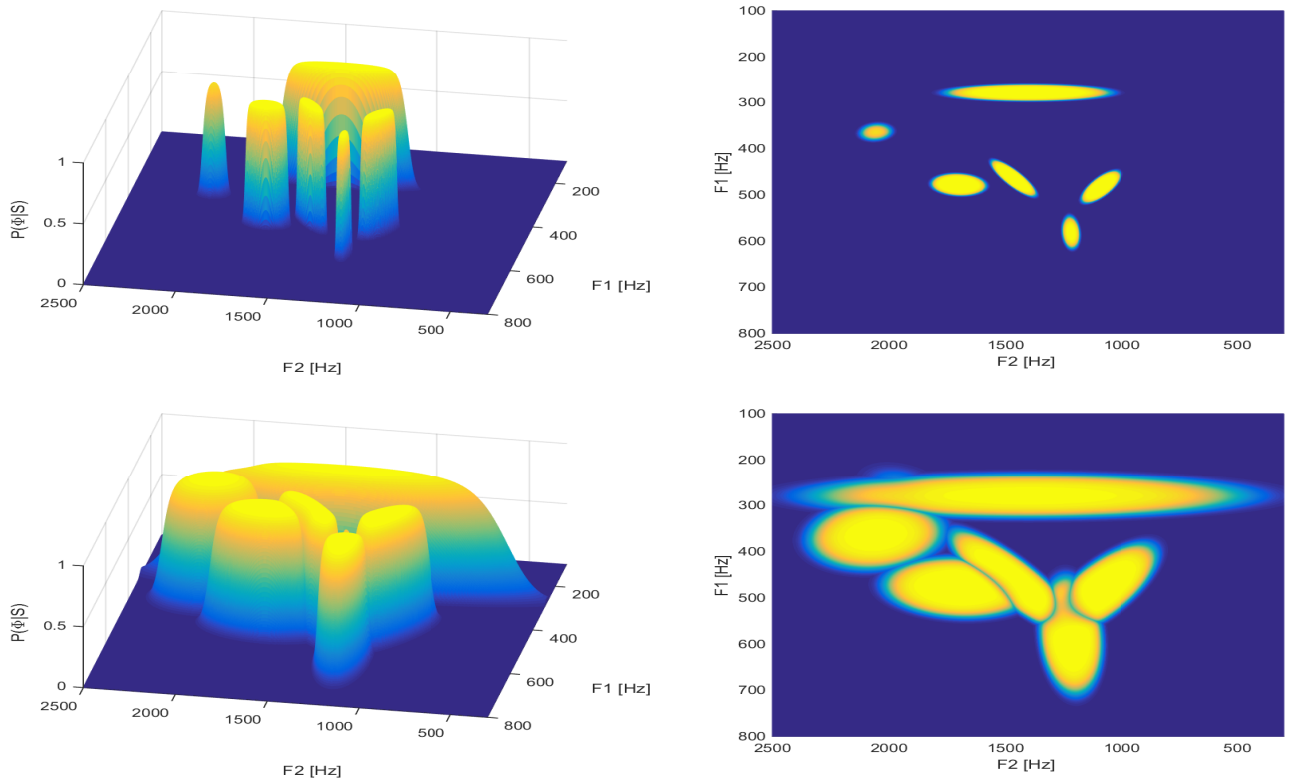

Fig. 5 Effect of the κ_S parameter on the likelihood functions $P(\Phi \mid S)$. We superpose the likelihood functions for all phonemes, and project them on the plane (F_1, F_2) defined by $F_3 = 2,450\ Hz$. **Top:** $\kappa_S = 0.3$. **Bottom:** $\kappa_S = 1$. Right panels are top-views of the left panels. Smaller values for κ_S narrow the confidence regions in the categorization task.

λ_3 in phoneme /ɔ/. Some other appear to have a wide range of variation, for instance λ_5 for all phonemes. This indicates the importance of the role of each muscle in performing each phoneme. In particular, we notice that control variables λ_1 and λ_3 negatively correlate for phonemes /i, e, ɛ, a/. Smaller values of λ_1 are related to higher values for λ_3 and *vice versa*. The values taken by these control variables specify the activation level of the Posterior Genioglossus and Hyoglossus muscles. Small values of the λ control variables correspond to high levels of muscle activation and *vice versa*.

We can thus see that the Bayesian model correctly extracts the antagonist interaction of these two muscles in the front/high and back/low movement direction. This antagonism has been found in electromyographic measures of muscle activity during speech production (Honda (1996) Figure 2). This direction of movement is thus coherent with the variation of position of the

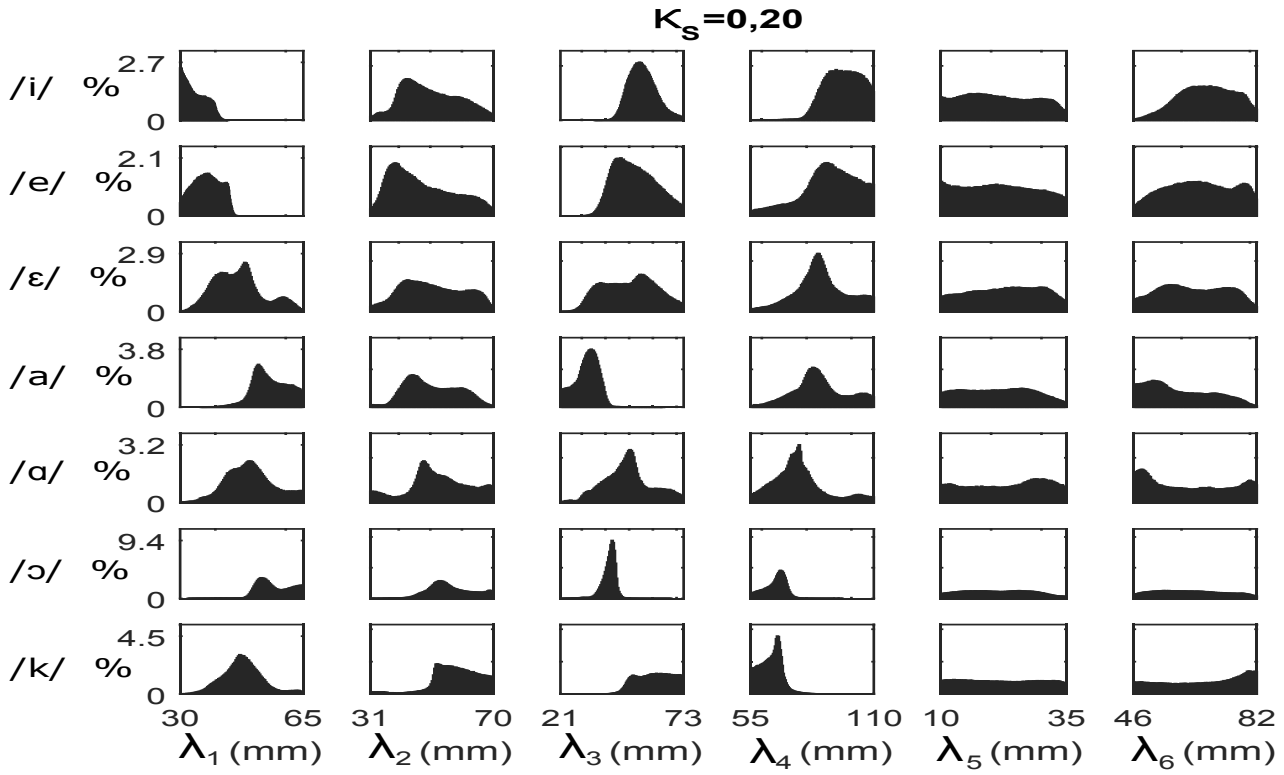

Fig. 6 Histograms of 2.10^6 control variables samples, obtained through MCMC algorithm according to $P(M|\Phi)$ for $\kappa_S = 0.2$. Lines correspond to phonemes and columns to each control variable λ. The corresponding muscles controlled by each control variable are: λ_1: Posterior Genioglossus, λ_2: Anterior Genioglossus, λ_3: Hyoglossus, λ_4: Styloglossus, λ_5: Verticalis, λ_6: Inferior Longitudinalis.

tongue for the production of these four phonemes, qualitatively confirming the good adequacy of the Bayesian model with experimental results.

In order to assess the performance of the Bayesian model it is necessary to evaluate its capacity to effectively generate spectral properties that distribute around the correct areas defined in the acoustic space for each phoneme. Figure 7 represents the histograms of the first three formants of the acoustic signals corresponding to the samples M of Figure 6. These values were obtained through the RBF network that models the mapping between M and S, as described in Section 2. It can be seen that the obtained formants correctly distribute inside the goal regions.

3.2 Bayesian model for planning a sequence of phonemes

The previous section described a Bayesian model inferring motor control variables for the production of a unique phoneme. We now turn to planning sequences of phonemes under a "minimum effort" assumption. The concept of effort in motor control is not uniquely defined (see for example Nelson (1983) for some possible acceptations). In this paper "effort" is evaluated in terms of global change in motor control variables along the sequence. GEPPETO implements this assumption with a cost function favoring small variations of the motor control variables across the planned sequence. The present section formulates a Bayesian model that aims at performing the same planning task as GEPPETO. In this Bayesian version, the cost function implementing the "minimum effort" assumption is cast as an additional constraint on the transitions between motor control variables.

The model follows the Bayesian Programming approach illustrated in the previous section. As it will be seen, the previous single-phoneme model appears to be nested as a substructure of the three-phoneme model.

3.2.1 Description

Variables Planning a sequence of phonemes involves the same M, S and Φ variables considered in the previous single-phoneme model. They correspond to the motor control variables, spectral properties of the acoustic signal and phonemes, respectively. However, as we are considering a sequence instead of a single phoneme,

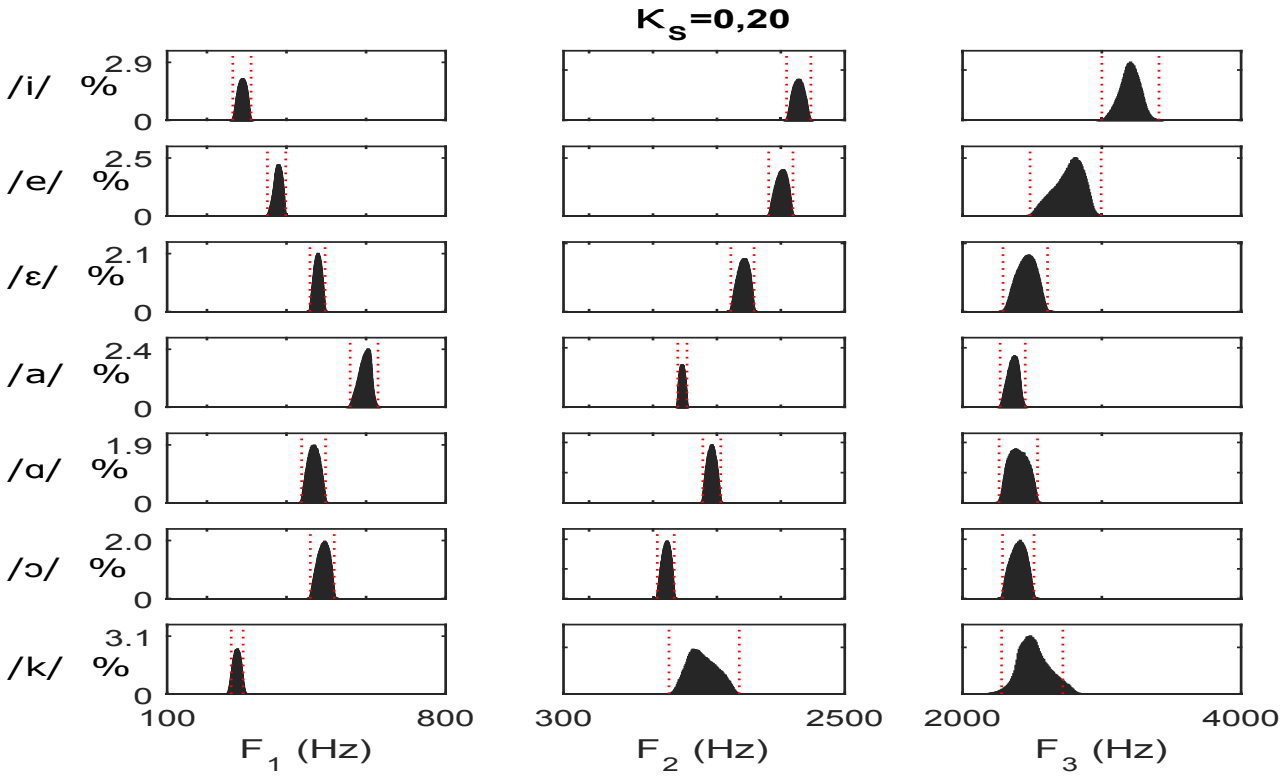

Fig. 7 Histograms of the acoustic signals resulting from the sampling of control variables sampled from the inferred probability distribution $P(M \mid \Phi)$ for $\kappa_S = 0.2$. The vertical dotted lines indicate borders of the acoustic regions characterizing each phoneme in the original formulation of GEPPETO.

each variable is repeated as many times as there are elements in the sequence. We therefore distinguish each different instance of the variables by an index specifying its position in the sequence. Thus, variables become M^i, S^i and Φ^i, with $i \in \{1:3\}$. For simplicity, we will denote by $Y^{1:3} = \{Y^1, Y^2, Y^3\}$ the conjunction of different instances of a given variable Y at different positions in the sequence.

An additional variable is also introduced in order to take into account the "minimum effort" assumption made in GEPPETO. This variable is denoted by C_m, standing for "motor constraint". C_m is a binary variable that acts as a switch, being either in the position L for "Lazy" (corresponding to the minimum effort requirement) or H for "Hyperactive" (corresponding to its opposite, a "maximum effort" requirement).

Decomposition The joint probability distribution is $P(M^{1:3}\ S^{1:3}\ \Phi^{1:3}\ C_m)$. Defining $X^i = \{M^i, S^i, \Phi^i\}$ as the set of all the variables at position i in the sequence, the joint probability distribution can be written

$$P(M^{1:3}\ S^{1:3}\ \Phi^{1:3} C_m) = P(X^1\ X^2\ X^3\ C_m). \tag{9}$$

Applying the chain rule, the right term of Equation (9) can be decomposed as

$$P(X^1\ X^2\ X^3\ C_m) = P(X^1)\ P(X^2|\ X^1)\ P(X^3\ |X^2\ X^1)\ P(C_m \mid X^3\ X^2\ X^1). \tag{10}$$

This expression can now be simplified thanks to the hypotheses made in GEPPETO. First, besides the cost function implementing the minimum effort assumption in GEPPETO, there is nothing creating any dependencies relating variables at different positions in the sequence. In real speech production, this type of constraint does exist, and corresponds to what is called "phonotactic" rules in linguistics. These rules are language dependent. It is not the purpose of the present study to address this type of high level linguistic constraints. We see though that the Bayesian Programming framework would be appropriate to account for this kind of constraint. According to this independence of variables at different positions in the sequence, the second and third factors in the decomposition of Equation (10) can be simplified such that:

$$P(X^{1:3}\ C_m) = P(X^1)\ P(X^2)\ P(X^3)\ P(C_m \mid X^3\ X^2\ X^1). \tag{11}$$

The last factor in the decomposition corresponds to the dependence of the variable C_m on the other variables. According to H_{6a}, the cost function only takes into account control variables $M^{1:3}$ at all positions in

the sequence by computing the perimeter of the triangle defined by these control variables. No other variable directly influences variable C_m. The last term in the decomposition of Equation (11) can therefore be further simplified as

$$P(C_m \mid X^3\ X^2\ X^1) = P(C_m \mid M^3\ M^2\ M^1), \tag{12}$$

Taking into account these simplifications, the joint probability distribution becomes

$$P(X^{1:3}\ C_m) = P(X^1)\ P(X^2)\ P(X^3)\ P(C_m \mid M^3\ M^2\ M^1). \tag{13}$$

From this last expression it can be seen that the decomposition of the joint probability distribution $P(X^{1:3}C_m)$ contains three copies of the probability distribution $P(X^i)$. They correspond to the joint probability distribution of variables involved in the production of a single phoneme, derived in Section 3.1:

$$P(X^i) = P(M^i)\ P(S^i \mid M^i)\ P(\Phi^i \mid S^i). \tag{14}$$

Combining Equation (13) with Equation (14) gives the complete decomposition of the joint probability distribution:

$$\begin{aligned} P(M^{1:3}\ S^{1:3}\ \Phi^{1:3}\ C_m) = {} & P(M^1)\ P(S^1 \mid M^1)\ P(\Phi^1 \mid S^1) \\ & P(M^2)\ P(S^2 \mid M^2)\ P(\Phi^2 \mid S^2) \\ & P(M^3)\ P(S^3 \mid M^3)\ P(\Phi^3 \mid S^3) \\ & P(C_m \mid M^3\ M^2\ M^1). \end{aligned} \tag{15}$$

Figure 8 represents the Bayesian network corresponding to the decomposition of the generative model given by Equation (15). Although this decomposition does not show explicit dependencies between control variables, the posterior distributions – of the most likely motor sequence, given a sequence of phonemes – will be conditionally dependent. In other words, the most likely motor change towards the next phoneme depends on all the other phonemes of the sequence.

This decomposition can be interpreted as being composed of the likelihoods of producing the desired phonemes at each target point in the sequence (*i.e.* all $P(S^i \mid M^i)\ P(\Phi^i \mid S^i)$ terms in Equation (15)) and of a prior belief about the sequence of motor goals (other terms in Equation (15)). We will see below that the cost function in the control space, described in H_6, plays the role of this prior belief, while the perceptual constraints can be regarded as the corresponding likelihoods.

Parametric forms Having derived the decomposition of the joint probability distribution, Equation (15), it is necessary to determine the form taken by each of the factors in this expression. This was already done in Section 3 for the terms appearing in the first three lines in Equation (15). The last term, $P(C_m|M^3M^2M^1)$, represents the dependence of variable C_m on the control variables. The aim of the cost function in GEPPETO is to penalize patterns of control variables that are far from each other by attributing them a cost that increases with the perimeter of the triangle that they define in the control space (H_{6a}). The same motor constraint is implemented in $P(C_m \mid M^3M^2M^1)$ through:

$$\begin{aligned} & P([C_m = L] \mid M^3M^2M^1) \\ & = e^{-\kappa_M(|M^2-M^1|+|M^2-M^3|+|M^3-M^1|)}. \end{aligned} \tag{16}$$

The additional parameter κ_M is introduced in order to modulate the strength of the constraint on motor control variables M. The motor constraint given by Equation (16) is interpreted in the following way. The further the control variables are from each other, the smaller the probability for the variable C_m to be in state $L =$"Lazy". Therefore if the state of being lazy is desired, its realization would become more probable for motor control variables being close from each other.

For completeness, as C_m takes only two values, the corresponding expression for the probability of having $C_m = H$ is given by:

$$\begin{aligned} & P([C_m = H] \mid M^3M^2M^1) \\ & = 1 - P([C_m = L] \mid M^3M^2M^1) \\ & = 1 - e^{-\kappa_M(|M^2-M^1|+|M^2-M^3|+|M^3-M^1|)}. \end{aligned} \tag{17}$$

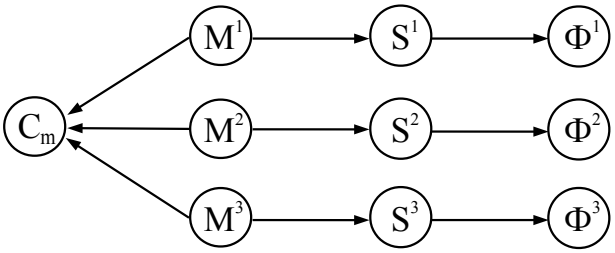

Fig. 8 Bayesian network corresponding to the decomposition of the joint probability distribution given by Equation (15).

3.2.2 Planning in the context of the Bayesian three-phoneme model

Considering the planning problem addressed in GEPPETO, the task assigned to the Bayesian three-phoneme model is to infer a sequence of motor control variables $M^{1:3}$ under the condition that the desired phonemic categories $\Phi^{1:3}$ are reached and assuming the "Lazy" state for variable C_m. This inference is formulated in Bayesian terms by $P(M^{1:3}\ |\Phi^{1:3}\ [C_m = L])$. This is again solved in a standard way through the knowledge

provided by the joint probability distribution of Equation (15). The corresponding expression is given by:

$$\begin{aligned}
&P(M^{1:3} \mid \Phi^{1:3}\ [C_m = L]) \\
&\propto \sum_{S^{1:3}} P(X^1)P(X^2)P(X^3)P([C_m = L]|M^3M^2M^1) \\
&\propto P([C_m = L] \mid M^3M^2M^1) \sum_{S^{1:3}} \prod_{i=1:3} P(X^i) \\
&\propto P([C_m = L] \mid M^3M^2M^1) \prod_{i=1:3} \sum_{S^i} P(M^i\ S^i\ \Phi^i) \\
&\propto P([C_m = L] \mid M^3M^2M^1) \prod_{i=1:3} P(\Phi^i \mid S^*(M^i)) \quad (18)
\end{aligned}$$

where the proportionality symbols account for normalization constants. The last line derives from the same observation as in Section 3 for the sum over S^i of the joint probability distribution $P(M^i\ S^i\ \Phi^i)$.

This completely specifies the solution to the inference problem.

3.2.3 Results

The Bayesian three-phoneme model is implemented through Monte Carlo sampling as described in Section 3.1.3 The performance of the previous single-phoneme model was evaluated in relation to its capacity to produce spectral properties that are located in the desired target regions in acoustic space . For the present three-phoneme model, performance is evaluated with respect to the minimization of the distance between inferred control variables along with the same perceptual constraint concerning the spectral properties. With respect to this perceptual constraint, Figure 9 illustrates the spectral properties of acoustic signals corresponding to 100 inferred motor control variables for the production of the sequence /aki/. The optimal solution obtained under similar conditions with GEPPETO is also displayed for comparison.

The first observation is the variability of the results obtained from the Bayesian three-phoneme model. This was expected given the probabilistic framework of the model. Secondly, it can be observed that the obtained spectral patterns effectively distribute inside the correct target regions. This illustrates that the three-phoneme model correctly satisfies the perceptual constraint. It will be shown below that the achievement of this constraint can be controlled by the values of parameters κ_S and κ_M. Note that these spectral and motor parameters are controlling the certainty or confidence of probabilistic mappings and are thus related to precision (or inverse variance) of the control in the corresponding space. Thirdly, it can be noted that point clouds characterizing the distributions of the resulting spectral properties are shifted from the center of the target regions toward their boundaries, with a clear tendency for the /a/ productions to be shifted to smaller F_1 values, and for the /k/ productions to be shifted toward higher F_2 values. This shows the influence of the motor constraint on the planned sequence at the acoustic level. This is also observed in the sounds obtained with GEPPETO. Finally, Figure 10 illustrates the role of parameters κ_M and κ_S in the fulfillment of the perceptual constraint. It can be seen that the stronger the weight of the motor constraint (κ_M), relative to the perceptual constraint (κ_S), the stronger the shift of the points from the central regions. At the extreme, targets are no longer reached if the value of κ_M becomes too large compared to κ_S as can bee seen in the two bottom panels of Figure 10.

We have seen the effect of the motor constraint on the planned sequence at the acoustic level: acoustic signals deviate from the center of the target regions and tend to be closer from each other. However, it should be noted that the minimization of the motor cost occurs in the motor space and not in the acoustic space. Hence, the closer proximity of spectral realizations of the phonemes in the sequence is a consequence in the acoustic space of the constraint in the motor space. This explain in particular that, in the upper panel of Figure 10, the spectral characteristics of the selected realizations of phoneme /i/ appear to deviate away from the two other phonemes, instead of being closer as one would expect from the form of the motor constraint. A tentative explanation for this phenomenon is the strong non-linearity of the mapping relating motor control variables to the spectral properties of the acoustic signal, observed in particular for vowel /i/.

In order to evaluate whether the motor constraint actually performs the minimization of the distance between motor control variables involved in the sequence, it is necessary to evaluate the actual perimeter of the triangle that they define in the motor space. Figure 11 shows the average value taken by this perimeter for 100 inferences of the Bayesian three-phoneme model for the sequence /aki/, as a function of the parameter κ_M and for different values of κ_S. The value obtained with GEPPETO is also presented for comparison. We first note that curves corresponding to different values of κ_S all merge for $\kappa_M = 0$. This corresponds to the situation where there is no constraint in the control variables, and therefore planning of the sequence is performed independently of the other phonemes in the sequence. The average value of the distance between control variables does not depend on κ_S in that case. Next, we observe that the average perimeter is clearly reduced when the strength of the motor constraint is raised with κ_M, and the capacity to minimize the motor cost is stronger for

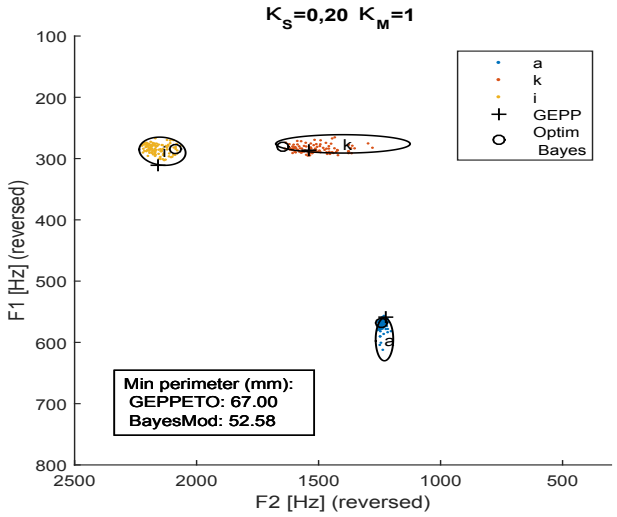

Fig. 9 Projection of the acoustic signal on the (F_2, F_1) plane, obtained by 100 motor control variables sampled from the inference probability distribution for the production of sequence /aki/. The acoustic signal obtained with GEPPETO and by the sample of the Bayesian three-phoneme model with minimum perimeter are also indicated. Values of the perimeters obtained by each model are indicated.

higher values of κ_S (i.e. for small perceptual constraints, see Figure 5). This illustrates the trade-off between the two constraints governed by κ_M and κ_S in the Bayesian three-phoneme model.

Now, how does the Bayesian three-phoneme model perform compared to GEPPETO? It can be noted in Figure 11 that for each value of κ_S (i.e. each level of perceptual constraint) there is a value of κ_M (i.e. a strength of the motor constraint) for which the average distance between control variables obtained with the Bayesian three-phoneme model coincides with the result obtained with GEPPETO. For instance, for $\kappa_M = 1$ the Bayesian three-phoneme model coincides with GEPPETO when $\kappa_S = 0.2$. Figure 9 confirms that for these specific parameter values, the perceptual constraint is correctly satisfied and the spectral characteristics obtained with the Bayesian three-phoneme model are close to the spectral characteristics obtained with GEPPETO. This suggests the equivalence of the two models for these specific values of the parameters. However, if we compare the optimal control variables obtained with the Bayesian three-phoneme model, i.e. those that minimize the perimeter in the motor control space, with the optimal commands obtained with GEPPETO, we realize that the optimal perimeter obtained by the Bayesian three-phoneme model is actually smaller than the one obtained by GEPPETO. This suggests that GEPPETO has not found the true optimal values. We will return to this issue in the next section.

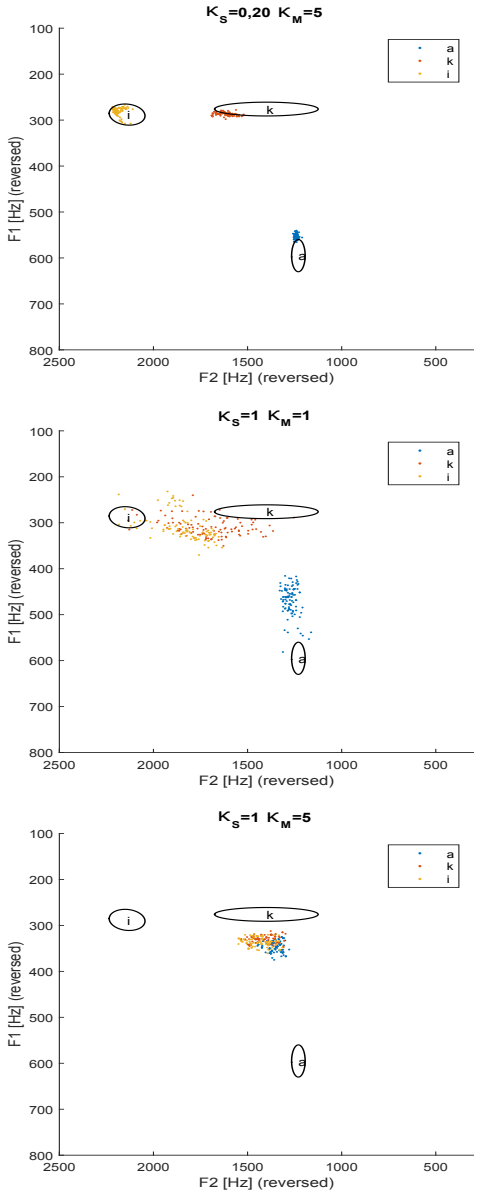

Fig. 10 Effect of parameters κ_S and κ_M on the spectral properties of the acoustic signals obtained by the Bayesian three-phoneme model. Refer to Figure 9 for comparison. **Top:** Keeping κ_S to the same value as in Figure 9 and augmenting κ_M by a factor 5. **Middle:** Keeping κ_M to the same value as in Figure 9 and multiplying κ_S by a factor 5 (remember that augmenting κ_S corresponds to relaxing the constraint, see Figure 5). **Bottom:** Augmenting the motor constraint and relaxing the perceptual constraint at the same time. Phonemic targets are attained as long as there is a correct balance between the strength of the motor constraint and the strength of the perceptual constraint.

4 Discussion

4.1 Equivalence of models

We have described a feedforward optimal control model of speech planning formulated within a Bayesian model-

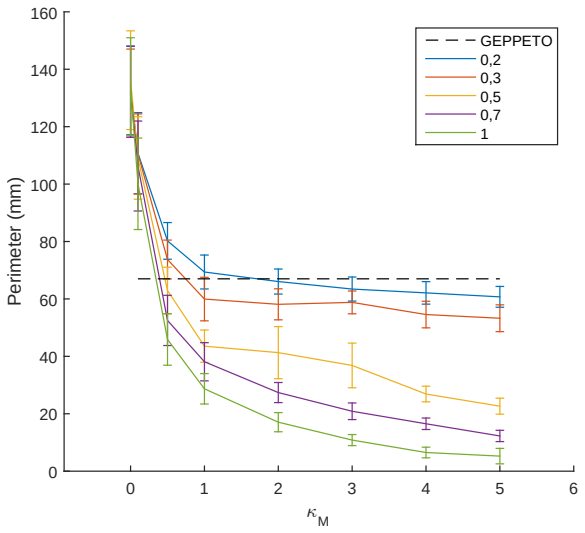

Fig. 11 Average distances obtained with the Bayesian three-phoneme model as a function of the parameter κ_M. Results for different values of κ_S are plotted (listed in the insert on the right). Error bars indicate variability obtained over 100 random samplings. The black horizontal dashed line represents the value obtained with GEPPETO.

ing framework. The results of simulations indicate that, as for its optimal control version, the Bayesian three-phoneme model correctly infers motor control variables that perform the desired motor task satisfying the specified perceptual and motor constraints. Furthermore, for specific values of the parameters characterizing the strengths of the constraints in the Bayesian three-phoneme model, simulations suggest the equivalence of results obtained by both models. This equivalence is evaluated on the basis of the comparison of average values obtained with the Bayesian three-phoneme model with the optimal solution obtained with GEPPETO. Nevertheless, it can be shown that the optimal control model can be obtained as a particular case of the Bayesian three-phoneme model if one looks for the configuration of control variables that maximize the posterior probability given by $P(M^{1:3} \mid \Phi^{1:3})$. The derivation of this result is provided as a supplementary material and rests on the property that the negative logarithm of $P(M^{1:3} \mid \Phi^{1:3})$ turns out to be equivalent to the cost function of GEPPETO. Therefore, maximizing the probability $P(M^{1:3} \mid \Phi^{1:3})$ is identical to minimizing the equivalent cost function of GEPPETO, showing that the Bayesian three-phoneme model can be simplified to GEPPETO in this specific implementation scheme. Note that there are mathematical theorems showing that a Bayesian scheme exists for any set of cost functions and optimal behaviour. These are known as complete class theorems (Brown, 1981; Robert, 2007). Knowing this theoretical context, stating that the Bayesian reformulation of GEPPETO is able to account for its optimal control scheme is not surprising. However, the theorems state the existence of the Bayesian reformulation; our contribution goes further, by defining the structure taken by this reformulation in our case. This is discussed in more detail in the sections below. Note also that parameters κ_S and κ_M are absent from the GEPPETO model. The inference perspective on motor control equips models with parameters that encode confidence or precision. In other contexts, these parameters could reflect important sources of inter-subject variability; and, possibly, an explanation for neurological and psychiatric symptoms (e.g. Parkinson's disease). In addition, such parameters could have a key role in modulating the gain of policy selection or motor execution and may play a pivotal role in phenomena like sensory attenuation and action observation (Friston et al, 2011).

4.2 Addressing redundancy and variability in formal terms

We were interested in the problem of how a feedforward model of motor planning can solve the indeterminacy characterizing the specification of motor control variables for achieving a desired motor task, without resulting in a stereotyped behavior. The essence of the dilemma was rooted in the fact that on the one hand indeterminacy arises from redundancy, i.e. from the multiplicity of solutions to the problem, and on the other hand solving redundancy, i.e. eliminating all possible solutions but one, inevitably results in stereotypy. We suggested that variability could be recovered at this point by assuming that even if the planning problem is driven by an optimality assumption, the actual solution might not be a stereotyped one. The absence of stereotypy may be first due to inherent computational limitations of the search for optimal solutions. In GEPPETO the optimization algorithm relies on a gradient descent scheme. Crucially, due to non-linearities relating variables in the model, the cost function may feature multiple local minima and the solutions obtained by gradient descent techniques may be highly dependent on the initial values of the optimization algorithm. Initializing the gradient descent algorithm in GEPPETO with different starting positions does indeed drive convergence to different locally "optimal" solutions. In particular this explains why the solution obtained with GEPPETO, as shown in Figure 9, appears to have a greater perimeter value than the optimal solution found with the Bayesian three-phoneme model. The result for GEPPETO was actually chosen as the best one out

of 100 different initializations of the descent algorithm. The fact that the gradient descent algorithm has failed to converge in all of these 100 initializations indicates the degree of complexity of the optimization process.

In this context, it could be argued that variability in speech production arises from the existence of these multiple local solutions into which the optimization process may differently converge depending on its initial configuration. However, the variability introduced this way cannot be formally justified as actually arising from the model itself, since it is just an indirect consequence of the failure of its implementation for finding the true optimal solution that the model actually predicts. Moreover, this *ad hoc* implementation can only account for variability in a qualitative way and does not have any theoretical or cognitive foundation.

In contrast, formulating the feedforward planning process within a Bayesian modeling framework has allowed us addressing the indeterminacy of the problem in addition of dealing with variability in formal terms. This is made possible by the fact that the Bayesian approach does not solve indeterminacy by suppressing all solutions but one. Instead, the Bayesian framework characterizes every possible configuration by its probability to achieve the task. Redundancy is then solved by randomly selecting motor control variables under the corresponding probability distribution. The optimal achievement of the task is still ensured in average, since the most probable motor control variables inferred under this process correspond to the more relevant ones for the motor task. Variability becomes therefore an inherent consequence of the formalism. Furthermore, the variability generated with this approach has a specific structure that could be compared with experimental data. For instance the model predicts that the relative frequency of selected motor control variables is given by the probability $P(M^{1:3} \mid \Phi^{1:3})$.

Therefore, the advantage of the Bayesian modeling approach is to suggest that a probabilistic description of the planning process is able to deal with the selection of solutions to an ill-posed problem without destroying variability (Colas et al, 2010). This allows to treat variability in formal terms and not as the result of an *ad hoc* implementation of the model.

The pertinence of an approach that designs models integrating multiple local solutions in formal terms is illustrated by the work of Ganesh et al (2010). Their work indicates that motor memory plays a crucial role influencing the outcome of the planning process, in addition to the optimization of cost related to error and effort. Thus, motor memory would be responsible for setting variable initial states of the motor system, which would influence the convergence of the search for optimal solutions toward local optima. Even if the Bayesian three-phoneme model that we have presented does not account for the role of motor memory in the planning process, the Bayesian modeling approach offers a framework in which motor memory could be modeled via a set of local approximations to the complete probability distribution, as it would be performed by local Laplace approximation or by standard variational inference methods. This raises the question of how agents would encode the knowledge described by the probability distributions involved in the presented scheme. While a complete representation of a complex knowledge would involve an important amount of resources, it would be natural to select a simpler approximation to this knowledge as it would be advantageous for the agent and often sufficient for practical purpose. Indeed, there is a fairly established literature on active inference using variational Bayes in the context of Bayesian filtering (also known as predictive coding). In brief, by equipping predictive coding with reflexes, active inference simulates action trajectories, action observation and indeed communication (e.g., the bird song examples in Friston and Frith (2015)). As in the current Bayesian formulation, active inference dispenses with cost functions and replaces them with prior beliefs about the way the motor plant should behave.

5 Conclusion

We propose to conclude by widening the discussion of our contribution. We first note that, of course, marrying optimal control or optimal planning theories and probabilistic modeling has already a long history. Previous approaches abound; we provide a few, mostly classical entry points in the vast literature of decision-theoretic planning in robotics and AI (Kaelbling et al, 1998; Boutilier et al, 1999; Attias, 2003; Murphy, 2002; Toussaint, 2009; Kappen et al, 2012), and in the Bayesian Decision Theory in cognitive modeling (Wolpert, 2007; Daunizeau et al, 2010; Ma, 2010, 2012)

However, our approach differs in that, instead of marrying probabilities and cost functions, we proposed to do away completely with the notion of cost functions , something that has already been proposed within the active inference scheme (Friston, 2011; Friston et al, 2009, 2012). We have reformulated the cost function of an existing, optimality based model, as probability distributions, motivated by a desire to obtain trial-to-trial variability in a principled manner. A consequence that we have already exposed is that the optimality based model could be seen as a special case of the proposed Bayesian three-phoneme model.

However, this nesting of models does not imply that the Bayesian modeling framework would be inherently more powerful than optimal based modeling. Indeed, one could strive to expand optimal based models to recover trial-to-trial variability, with other mathematical methods that we do not imagine at the moment. Therefore, this warrants caution concerning the interpretation of our contribution. Since a single model can be expressed equivalently in two different mathematical formalisms, none of these formalisms can be claimed, at face value, to be e.g. more biologically plausible than the other.

This is a singular epistemological stand, in the current debate about the theoretical contribution of Bayesian modeling to cognitive sciences (Jones and Love, 2011; Bowers and Davis, 2012; Hahn, 2014). A single Bayesian model does not bring much evidence that the brain would encode and manipulate probabilities (the so-called Bayesian Brain Theory), if only because of mathematical equivalent model found in other formalisms. Indeed, the Bayesian formalism itself appears under-constrained. In other words, writing *a* Bayesian model of a given cognitive function is always feasible. As a side note, this does not preclude a reifiability based argument in favor of the Bayesian Brain Theory; if there are many Bayesian models of many cognitive functions, then one can find probabilities more likely to be "used" by natural cognitive systems. Our current contribution is a step in this research program. This is also complementary to, and orthogonal to, studies of the same question at the microscopic biological level (e.g., neural or population of neuron based accounts of cognitive processes; (Pouget et al, 2013)).

Advantages, then, are to be found on other grounds. We propose to highlight the interest of Bayesian modeling as a mathematical modeling tool. As we have seen, to express knowledge, a single mathematical construct, that is, probability distributions, is required. Such a unified formalism is interesting in several aspects. For instance, it makes comparison and composition of pieces of knowledge easier. In our case, this was illustrated by the composition of speech constraints of varied nature, one concerning motor economy, the other concerning perceptual discriminability. Furthermore, additional constraints are easy to combine in order to enrich the model (e.g., we added to the model, but did not describe here for brevity, a parallel branch to acoustic targets that controls the force output, in order to modulate speech rate while conserving intelligibility).

Interpretability of the model also benefits from this unified formalism. In the present Bayesian model, the perceptual step-shaped constraint of GEPPETO was derived from the inversion of an internal representation of phonemes in terms of simple distributions of the spectral properties of the produced sounds. The perceptual constraint is therefore interpreted in the Bayesian model as an internal perceptual categorization of the produced acoustic sound. This is interesting as it shows explicitly how perception is assumed to be involved in the control process. It further illustrates an additional advantage of the Bayesian framework as being well suited for treating perception and action in a unified framework (Toussaint, 2009; Friston, 2010).

Moreover, to manipulate knowledge, the few rules of probability calculus are sufficient. Inference directly and automatically derives from the choice of the model; in more technically precise words, defining the joint probability distribution and the probabilistic questions asked to this joint probability distribution completely constrains the resulting inference processes. In that sense, our method differs from most other approaches, by placing the focus of modeling on knowledge expression, instead of the inference process. The cognitive model we propose is therefore resolutely representational; it lies at the algorithmic level of Marr's classical hierarchy (Marr, 1982). We expect this original perspective to expand in cognitive science modeling, in the wake of the current explosion of probabilistic programming languages (Goodman et al, 2008; Gordon et al, 2014).

Acknowledgements Authors wish to thank Pierre Bessière and Jean-Luc Schwartz for guidance and inspiring conversations.

References

Attias H (2003) Planning by probabilistic inference. In: Bishop CM, Frey BJ (eds) Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL

Bessière P, Laugier C, Siegwart R (eds) (2008) Probabilistic Reasoning and Decision Making in Sensory-Motor Systems, Springer Tracts in Advanced Robotics, vol 46. Springer-Verlag, Berlin

Bessière P, Mazer E, Ahuactzin JM, Mekhnacha K (2013) Bayesian Programming. CRC Press, Boca Raton, Florida

Boutilier C, Dean T, Hanks S (1999) Decision theoretic planning: Structural assumptions and computational leverage. Journal of Artificial Intelligence Research 10:1–94

Bowers JS, Davis CJ (2012) Bayesian just-so stories in psychology and neuroscience. Psychological Bulletin 138(3):389–414

Brown LD (1981) A complete class theorem for statistical problems with finite sample spaces. The Annals of Statistics 9(6): 1289–1300

Calliope (1984) La parole et son traitement automatique. Masson

Colas F, Diard J, Bessière P (2010) Common bayesian models for common cognitive issues. Acta Biotheoretica 58(2-3):191–216

Daunizeau J, den Ouden HEM, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ (2010) Observing the observer (I): Meta-bayesian models of learning and decision-making. PLoS one 5(12):e15,554

Feldman AG (1986) Once more on the equilibrium-point hypothesis (λ model) for motor control. Journal of motor behavior 18(1):17–54

Friston K (2010) The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11(2):127–138

Friston K (2011) What is optimal about motor control? Neuron 72(3):488–98

Friston K, Mattout J, Kilner J (2011) Action understanding and active inference. Biological cybernetics 104(1-2):137–160

Friston K, Samothrakis S, Montague R (2012) Active inference and agency: optimal control without cost functions. Biological cybernetics 106(8-9):523–541

Friston KJ, Frith CD (2015) Active inference, communication and hermeneutics. Cortex 68:129–143

Friston KJ, Daunizeau J, Kiebel SJ (2009) Reinforcement learning or active inference? PLoS ONE 4(7):e6421

Ganesh G, Haruno M, Kawato M, Burdet E (2010) Motor memory and local minimization of error and effort, not global optimization, determine motor behavior. Journal of neurophysiology 104(1):382–390

Goodman ND, Mansinghka VK, Roy DM, Bonawitz K, Tenenbaum JB (2008) Church: a language for generative models. In: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, vol 22, p 23

Gordon AD, Henzinger TA, Nori AV, Rajamani SK (2014) Probabilistic programming. In: Proceedings of the 36th International Conference on Software Engineering (ICSE 2014, Future of Software Engineering track), ACM, New York, NY, USA, pp 167–181

Guenther FH (1995) Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological review 102(3):594–621

Guenther FH, Hampson M, Johnson D (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychological review 105(4):611–633

Hahn U (2014) The Bayesian boom: good thing or bad? Frontiers in Psychology 5:765

Honda K (1996) Organization of tongue articulation for vowels. Journal of Phonetics 24:39–52

Jones M, Love B (2011) Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. Behavioral and Brain Sciences 34:169–231

Jordan MI (1996) Computational motor control. In: Gazzaniga MS (ed) The Cognitive Neurosciences, MIT Press, Cambridge, MA, pp 597–609

Kaelbling L, Littman M, Cassandra A (1998) Planning and acting in partially observable stochastic domains. Artificial Intelligence 101(1-2):99–134

Kappen HJ, Gómez V, Opper M (2012) Optimal control as a graphical model inference problem. Machine learning 87(2):159–182

Kawato M (1999) Internal models for motor control and trajectory planning. Current opinion in neurobiology 9(6):718–727

Laboissière R, Ostry DJ, Feldman AG (1996) The control of multi-muscle systems: human jaw and hyoid movements. Biological cybernetics 74(4):373–384

Lebeltel O, Bessière P, Diard J, Mazer E (2004) Bayesian robot programming. Autonomous Robots 16(1):49–79

Ma L, Perrier P, Dang J (2006) Anticipatory coarticulation in vowel-consonant-vowel sequences: A crosslinguistic study of french and mandarin speakers. Proceedings of the 7th International Seminar on Speech Production (pp. 151-158), Ubatuba, Brazil.

Ma WJ (2010) Signal detection theory, uncertainty, and poisson-like population codes. Vision Research 50:2308–2319

Ma WJ (2012) Organizing probabilistic models of perception. Trends in Cognitive Sciences 16(10):511–518

Marr D (1982) Vision. A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, New York, USA

Ménard L (2002) Production et perception des voyelles au cours de la croissance du conduit vocal : variabilité, invariance et normalisation. Unpublished Ph.D. thesis, Université Stendhal de Grenoble

Murphy K (2002) Dynamic bayesian networks: Representation, inference and learning. Unpublished Ph. D. thesis, University of California, Berkeley, Berkeley, CA

Nelson W (1983) Physical principles for economies of skilled movements. Biological Cybernetics 46:135–147

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

Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, Guiod P (1997) Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. Speech communication 22(2):227–250

Perkell S J, Nelson L W (1985) Variability in production of the vowels /i/ and /a/. Journal of the Acoustical Society of America 77:1889–1895

Perrier P, Boë LJ, Sock R (1992) Vocal tract area function estimation from midsagittal dimensions with ct scans and a vocal tract castmodeling the transition with two sets of coefficients. Journal of Speech, Language, and Hearing Research 35(1):53–67

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

Perrier P, Ma L, Payan Y (2005) Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. In: Proceedings of Interspeech 2005, Lisbon, Portugal, pp 1041–1044

Poggio T, Girosi F (1989) A theory of networks for approximation and learning. Tech. rep., Artificial Intelligence Laboratory & Center for Biological Information Processing, MIT, Cambridge, MA, USA

Pouget A, Beck JM, Ma WJ, Latham PE (2013) Probabilistic brains: knowns and unknowns. Nature Neuroscience 16(9):1170–1178

Robert CP (2007) The Bayesian Choice – From Decision-Theoretic Foundations to Computational Implementation. Springer

Robert-Ribes J (1995) Modèles d'intégration audiovisuelle de signaux linguistiques : de la perception humaine a la reconnaissance automatique des voyelles. Unpublished Ph.D. thesis, Institut National Polytechnique de Grenoble

Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. Journal of Neurophysiology 79(6):3272–3278

Shim JK, Latash ML, Zatsiorsky VM (2003) Prehension synergies: trial-to-trial variability and hierarchical organization of stable performance. Experimental Brain Research 152(2):173–184

Todorov E (2004) Optimality principles in sensorimotor control. Nature neuroscience 7(9):907–915

Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nature neuroscience 5(11):1226–1235

Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. Neuroimage 39(3):1429–1443

Toussaint M (2009) Probabilistic inference as a model of planned behavior. Künstliche Intelligenz 3(9):23–29

Uno Y, Kawato M, Suzuki R (1989) Formation control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. Biological Cybernetics 61:89–101

Wolpert DM (2007) Probabilistic models in human sensorimotor control. Human Movement Science 26:511–524