Control of a Speech Robot via an Optimum Neural-Network-Based Internal Model With Constraints

Iaroslav V. Blagouchine and Eric Moreau, *Senior Member, IEEE*

***Abstract*—An optimum internal model with constraints is proposed and discussed for the control of a speech robot, which is based on the human-like behavior. The main idea of the study is that the robot movements are carried out in such a way that the length of the path traveled in the internal space, under external acoustical and mechanical constraints, is minimized. This optimum strategy defines the designed internal model, which is responsible for the robot task planning. First, an exact analytical way to deal with the problem is proposed. Next, by using some empirical findings, an approximate solution for the designed internal model is developed. Finally, the implementation of this solution, which is applied to the control of a speech robot, yields interesting results in the field of task-planning strategies, task anticipation (namely, speech coarticulation), and the influence of force on the accuracy of executed tasks.**

***Index Terms*—Artificial neural networks (ANNs), constrained optimization, Lagrange's multipliers method, mathematical and computational issues in robotics control, mathematical physics, models and theories of speech production, λ-model [equilibrium-point hypothesis (EPH)], optimum control, optimum task planning, path and trajectory planning, robotics of speech production, robot-motion planning, variational calculus.**

I. Introduction

ROBOTICS of speech production is quite a challenging subject in modern design and engineering. Since it is well-known that the tongue is one of the principal elements, which is responsible for speech production, many speech robots are based on the modeling of the movements of the tongue. To ensure the quality of these movements and, consequently, that of the produced speech (both have to be as close as possible to the real ones), the control of such robots is extremely important. To this end, the principles of control of such artificial-intelligence devices are often borrowed from human beings, and many robots are based on human-like behavior and are modeled in close conjunction with the motor-control theories [1]–[5]. Currently, there is no single unique theory in the field of motor control and task planning. Over the past 80 years, many different approaches were developed and are currently in competition; these include the electromyographic approach [1]–[3], the information-channel approach [3], [6]–[8], the global economy of the diverse mechanical factors approaches [9]–[15], the equilibrium-point hypothesis (EPH, which is also known as the λ-model) [1], [5], [16]–[27], the internal models' approaches [28]–[30], etc.

Among these approaches, the EPH, economy's approaches, and internal models' ones have received some special interest in speech robotics.

The EPH is a development of the classic linear damped spring model of muscle [31], which is completed by central nervous system (CNS) influence. According to the EPH, the muscle can be modeled as a nonlinear spring, which is controlled by a special motor command λ, which descends from CNS. The force $\mathcal{F}$, which is generated by such a muscle, depends on the difference between its actual length l and the CNS motor command λ, as well as on several other physical parameters associated with muscle. In other words, $\mathcal{F}(l,\lambda) = f(s)H(s)$, where $f(\cdot)$ is the transfer function of muscle, $H(\cdot)$ is the Heaviside step function, and the parameter s, which is called *activation*, is defined as $s = l - \lambda$ for the static case and as $s = l - \lambda + \varkappa\nu$ for the dynamic case, where parameters $\varkappa$ and ν are, respectively, the damping factor and the speed of muscle lengthening. As to the transfer function of muscle $f(\cdot)$, it has been noted that it is a nonlinear function, and it is quite well-approximated by an exponential function. Thus

$$\mathcal{F}(\lambda) = \rho\,(e^{cs} - 1)\,H(s) \tag{1}$$

where c is the form parameter, and ρ is a parameter related to the force-generating capability of muscle [19], [32], [33]. By expanding the latter expression in the Maclaurin series for $s > 0$, it is straightforward to see that for $0 < s \ll 1$, the muscle, in first approximation, behaves as a classic linear spring

$$\mathcal{F}(\lambda) = \rho \sum_{n=1}^{\infty} \frac{(cs)^n}{n!} = \rho\, c\, s + O(s^2) \tag{2}$$

which is another argument in favor of this model. From the point of view of robotics and cybernetics, the λ-model is especially attractive, because it provides a simple mathematical mean to conceive artificial-intelligence devices based on the human-like behavior, without going into details about the underlying principles of motor control. The EPH has also become quite popular in the articulatory speech-production field, for which correct

Manuscript received April 14, 2009; revised July 7, 2009 and September 17, 2009. First published November 13, 2009; current version published February 9, 2010. This paper was recommended for publication by Associate Editor T. Kanda and Editor J.-P. Laumond upon evaluation of the reviewers' comments.

I. V. Blagouchine is with the Department of Telecommunication, Institute of Engineering Sciences of Toulon-Var-School of Engineering, University of Toulon, Toulon F-83162, France, and also with the Department of Mobile Communication, Eurécom, F-06904 Sophia-Antipolis, France (e-mail: iaroslav.blagouchine@univ-tln.fr).

E. Moreau is with the Department of Telecommunication, Institute of Engineering Sciences of Toulon-Var-School of Engineering, University of Toulon, Toulon F-83162, France (e-mail: moreau@univ-tln.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2009.2033331

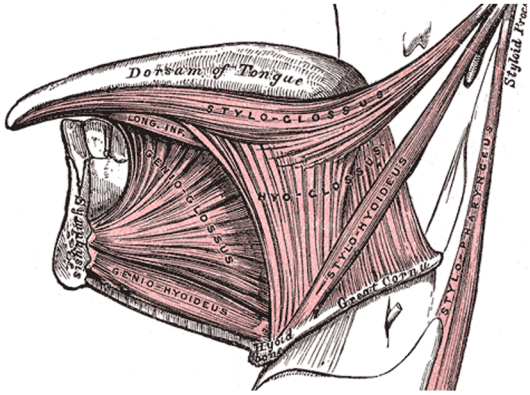


Fig. 1. Simplified anatomical structure of tongue (from [43]).

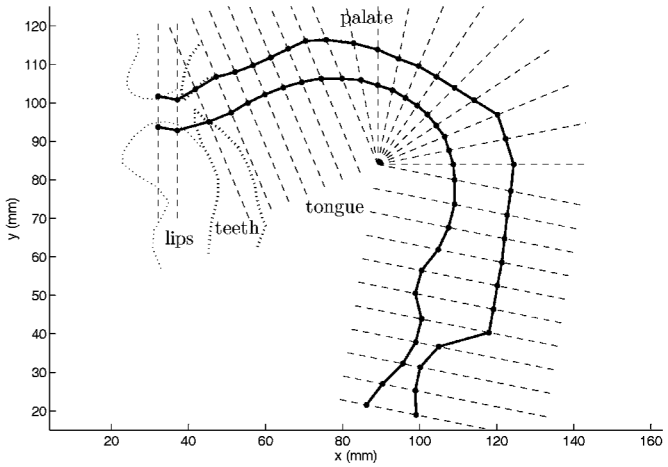


Fig. 2. Modeled vocal tract and its further cutting by an acoustical-tube model for the computation of formants $\boldsymbol{F}$. The upper contour is the palate, and the lower one is the tongue *dorsum*. The lips' area is variable, which ranges from 0.5 to 3.0 cm^2, depending on the vowel.

modeling of the tongue and jaw movements is of great importance, because these are physically responsible for speech production. Due to the increasing researcher's interest and to the constantly growing computational capacities, many of such works have arisen over the past 15 years (e.g., see, [33]–[40]).

One of these examples is the articulatory-based speech robot that we used in our study (see, e.g., [33], [41] and [42]). This robot represents an artificial tongue, which is modeled by six main muscles that are responsible to shape and move the tongue in the sagittal plane: *posterior* and *anterior* parts of *genioglossus*, *styloglossus*, *hyoglossus*, *inferior* and *superior longitudinalis*, and *verticalis* (see Fig. 1) [43]. Each of these muscles is controlled by its own motor command λ_i, $i = 1, \ldots, n$, $n = 6$, according to the EPH.[1] Their forces are generated according to (1), with different ρ for each muscle, and with constant $c = 1$ cm^{-1} [19], [32]. Initial vocal-tract geometry is reconstructed from anatomical cineradiographic data. By means of the finite-element method [44], the tongue is divided into small volumes connected by 221 nodes, each of which anatomically belongs to the defined muscle(s). The motion of each node is then described by a second-order ordinary differential equation (ODE) with damping and external terms, due to viscosity, gravity, and contact reaction forces. The stiffness matrix, which determins the distribution of the internal forces within the finite-element structure, is calculated by the finite-element algorithm. Such a complex system of the ODEs is solved numerically by means of the Runge–Kutta method using MATLAB software, which finally gives the trajectory of motion of each node and, by further interpolation, the motion of the tongue body. In order to achieve the vocal-tract reconstruction, lips, palate, and pharynx are also added to model mechanical contacts with tongue (see Fig. 2). The jaw is represented by static rigid structures to which the tongue is attached. Note, finally, that there are also other articulatory-based speech robots that might be interesting [45]–[59], especially because the optimum internal model that we will introduce is meant as a general model and can be used with many other speech robots using similar principles of control.

The diverse-economy approaches consider that the movements are defined by some economy principle, that is to say, the movements are always carried out in such a way that some criterion is optimized. These approaches are basically inspired from analytical mechanics, namely, from *principle of least action* [60]–[69], which is one of the most universal principles of physics (many fundamental equations of physics can be deduced from it). This principle states that the motions are always carried out in such a way that the *action*[2] is minimum. However, because of high complexity of biosystems, direct applications of this principle in motor-control theories are quite limited. Under these circumstances, the exact mathematical description being almost impossible, and one of the possible solutions might be to describe them more globally, i.e., by supposing that some global criterion is optimized during the movement. This criterion, which is often known as *a cost*, may be defined in many different ways, e.g., time cost, energy cost, force cost, impulse cost, accuracy cost, etc. In this field, the concept that appears to be the most frequent and interesting is that of the minimum of the *jerk* cost[3] [9]–[11]. The idea of the economy principles also affected robotics and, in particular, the speech-motor-control community, and several studies using economy principles appeared. Basically, these works propose to search for the shortest trajectories between the steady-state positions in the command internal space [32], [36], [39], [70]–[73]. In some of these works, the shortest distance principle is explicitly stated, and thus, authors suggest using straight lines as solutions (the straight line is the shortest trajectory between two points if there are no constraints). In others, it is implicitly formulated by constant-rate transitions between steady-state positions, i.e., again by straight-line transitions. These works reported that by shifting motor-control commands λ at constant speed, realistic articulatory movements and speech signals may be produced. It is interesting to note that the minimization of the trajectory

[1]For simplicity, we will write these motor commands as components of vector $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_n)$.

[2]The action is the definite integral over time interval of the Lagrangian, the latter being the difference between kinetic and potential energies.

[3]Jerk, also known as *jolt*, is the rate of change of acceleration, i.e., the third derivative of displacement with respect to time. The jerk cost is, therefore, defined as definite integral over time interval of the square of jerk.

length (3), as well as that of the jerk cost, are quite similar, at least, mathematically, to that of the action. In fact, in all three cases, one seeks to optimize a trajectory-depending functional, which is given as definite integral over time interval. Finally, the exploration of optimum principles may also be interesting in the field of inverse problems. By optimizing a cost functional (or just function) under constraints on outputs, one can find the corresponding inputs (as we will show later). This idea is not novel in the speech field. For more details, see [74]–[76].[4]

Finally, there is another approach that should be mentioned in the human-like robotics context: the internal models [28]–[30]. This approach supposes that any living creature has an internal representation of all the external tasks he/she can do.[5] Thus, typically, the learning of a new task implies, *inter alia*, the determination of the corresponding place in the internal space. It is also important to note that there is no bijection between internal and external spaces, since the same task can be achieved differently. The notion of tasks is closely related to that of *targets*, and it is very important for the planning theories. In speech motor control, there are different interpretations of targets, which is also known as *reference frames*. For example, these may be vocal-tract configurations (e.g., tongue shape, constriction position, and lips area) and, consequently, the output acoustic patterns, which may be expressed in terms of formants. Since there is great variability of the formants for the same vowel, the auditory system normalizes them, in order to recognize the vowel, which is the so-called *target-normalization theory*. The other theory is more complex and takes into account not only the static acousticophonetic parameters but also the dynamic ones, such as transitions and their character[6] (e.g., linear, different nonlinear forms, etc.). These transitions are often associated with formant transitions [79], [80], because it has been empirically established that this dynamic acoustic information also contributes for vowel identification [81] (see also various coarticulation references given in Section III-A2), which is the so-called *dynamic-target-specification theory* [82], [83]. However, these two basic interpretations of targets are far from being exhaustive, and the reader might particularly appreciate the theoretical study in [78], where one can find *muscle-length targets*, *articulator targets*, *constriction-position targets*, *acoustic targets*, *auditory perceptual targets*, etc.

The aim of our study is to propose an optimum internal model for the control of the speech robots based on the EPH. The motion planning of the robot is performed in its internal space, whose coordinates are $\boldsymbol{\lambda}$-motor commands of the EPH (internal space $\boldsymbol{\lambda}$ is, therefore, n-dimensional space). Being inspired by the principle of least action, it is proposed that the robot task planning is based on the global optimum principle, which is related to the aforementioned internal space, with external constraints related to the execution of tasks (e.g., the quality of the executed tasks), or in motor-control words, to the targets. It is proposed, namely, from that *all the movements, including those of the tongue, which are mainly responsible for speech production and are controlled by the* $\boldsymbol{\lambda}$ *commands according to the EPH, are carried out in such a way that the length of path, which is traveled in the internal space* $\boldsymbol{\lambda}$, *is minimized, under external physical constraints, namely, acoustical and mechanical ones*. The robot's behavior is, therefore, completely determined by this optimum principle, which permits finding the corresponding optimum commands $\boldsymbol{\lambda}$, which are sent to the robot. Therefore, the originality of our work consists in two important differences with respect to previously referenced works in the EPH-based robotics field. First, previous works do not perform the minimization of length in order to find corresponding motor commands $\boldsymbol{\lambda}$. They just use the fact that the straight line is the shortest path between two points. However, the latter fact is true if there are no constraints,[7] and with constraints, their approaches cannot provide solutions. Second, the optimization that we carry out is, in addition, a constraint one. We first perform it under one constraint (the acoustical one) and then under two constraints (the acoustical and the mechanical ones).

II. Optimum Internal Model

A. Preliminaries

First of all, we specify what we exactly mean by external physical constraints. The acoustical constraints consist in the specification of the sound that we wish to produce. Its specification is made in terms of the spectrum, and since the optimum internal model is mainly designed for vowels, the latter can be roughly approximated via the first k formants of vowel, which are denoted by vector $\boldsymbol{F} \equiv (F_1, \ldots, F_k)$. In practice, the formants are obtained via the BTM, which is followed by an acoustical tube model (see Fig. 3). First, the BTM provides the vocal-tract geometry $(\boldsymbol{x}, \boldsymbol{y})$, and then, the acoustical tube model cuts the vocal tract in cross sections (see Fig. 2) and approximates it by a tube of variable cross section. This yields the area function of the vocal tract, by which, the formants are computed [84]–[86]. The mechanical constraint consists in the requirement to keep the prescribed mean force's level contained in the tongue during speech production. This level is calculated as the arithmetic (or sample) mean of the absolute values of the forces at each node of the BTM. Physically, this level may be interpreted as mean muscular tongue effort, or measure of global tongue stiffness, and phonetically, it helps to account for lax and tense vowels.

B. Mathematical Formalization of the Formants–Commands and Force–Commands Relationships and Learning of the Artificial Neural Networks

As we saw before, the BTM is not a fully analytical robot. In other words, there are no explicit analytical relationships

[4]This work seems to be misinterpreted in [77], where the cost function from [76] was called "length," while it is clearly called by its author as "variation," and in addition, the formula provided in [76] does not represent the length.

[5]It has been even suggested that as in human beings, the most probable site for the latter may be the cerebellar cortex [26].

[6]The question where the transitions are planned and how they are controlled is also a subject of controversial discussion; some works suggested that it may be in spatial reference frames, while others reported that it may be more closely related to physical levels (e.g., joints and muscles) [78].

[7]Or they are trivial, e.g. straight line or plane passing through endpoints.

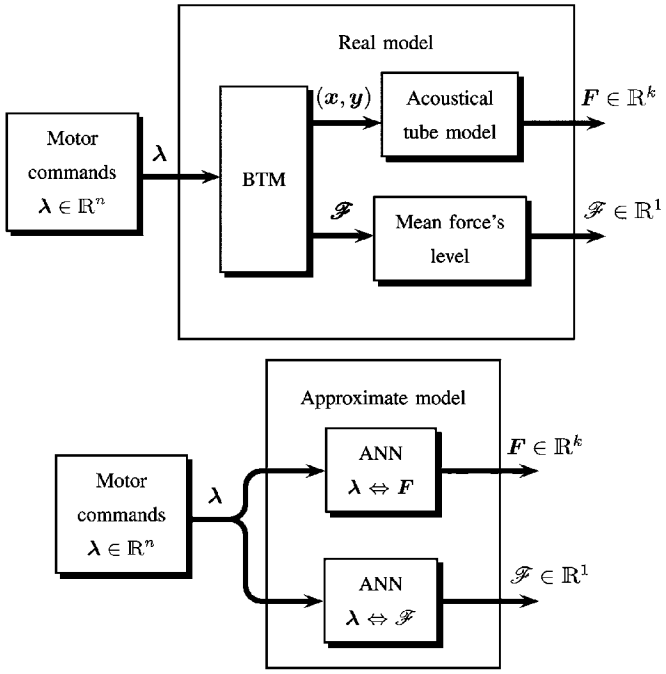


Fig. 3. (Top) Direct or real model and its replacement by the (bottom) approximate one.

between its inputs and outputs. However, in order to mathematically implement the minimization algorithm of the optimum internal model, we need the analytical relationships between the formants $\boldsymbol{F}$ and the motor commands $\boldsymbol{\lambda}$, which are denoted by vector field $\boldsymbol{F}(\boldsymbol{\lambda})$, and between the global mean force's level $\mathscr{F}$ and $\boldsymbol{\lambda}$, which are denoted by scalar field $\mathscr{F}(\boldsymbol{\lambda})$, as well as their derivatives. For this reason, we approximated the BTM, followed by an acoustical tube model, by two artificial neural networks (ANNs) [5], [87]–[91] (see Fig. 3). The choice of the ANNs for similar problems was already suggested by several authors [92]–[97]; moreover, the ANNs are precisely known for their good properties for multidimensional approximations. Besides, the replacement of a particular BTM by ANNs has potentially another application: the generalization of the proposed internal model for its use with other speech robots based on the EPH or using similar principles of control, whose input–output relationships may be approximated by the ANNs.

The learned ANNs (for details, see the Appendix) reveal the general nonlinear character of the dependencies $\boldsymbol{F}(\boldsymbol{\lambda})$ and $\mathscr{F}(\boldsymbol{\lambda})$, as shown in Fig. 4. Nevertheless, one can note that the dependencies $\boldsymbol{F}(\boldsymbol{\lambda})$ and, especially, $\mathscr{F}(\boldsymbol{\lambda})$ are not highly nonlinear; it suggests that it would be also reasonable to try the use of the multidimensional polynomials instead of the ANNs, since the former are "lighter" for calculations from the computational point of view.

C. Model Itself

The optimum internal model, which is designed according to the principle of the shortest path in the internal space under constraints, logically leads to the calculus of variations [98]–[105]. In fact, the problem to find a curve, whose length is least under constraints, is one of the typical problems of

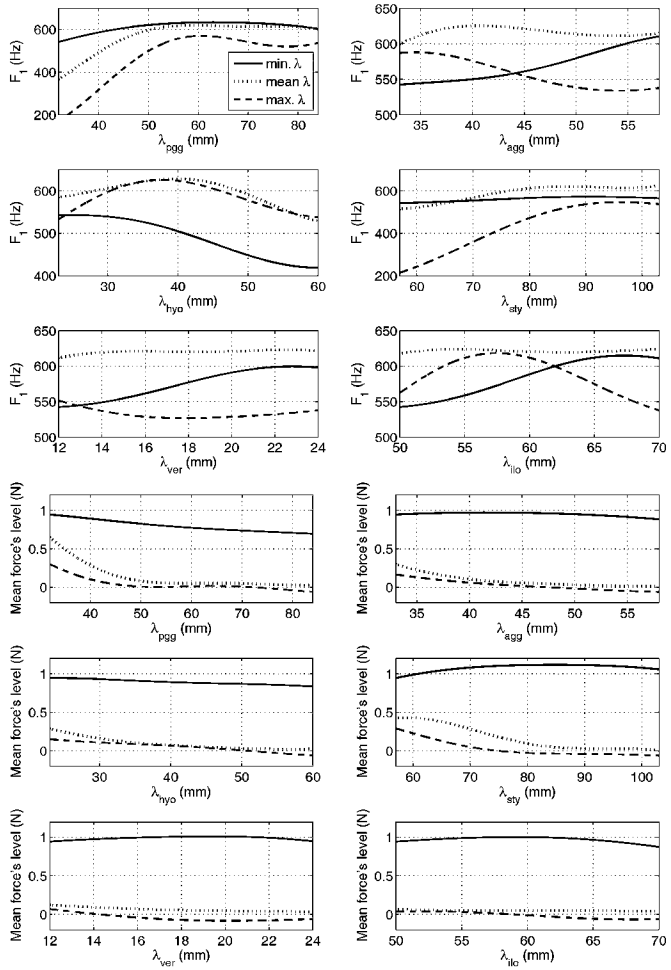


Fig. 4. Dependencies (upper six subfigures) $F_1(\boldsymbol{\lambda})$ and (lower six subfigures) $\mathscr{F}(\boldsymbol{\lambda})$ approximated by the corresponding ANN. Since both the functions $F_1(\boldsymbol{\lambda})$ and $\mathscr{F}(\boldsymbol{\lambda})$ depend on six variables ($n = 6$), they are shown in the following way: We fix five variables out of six and show the dependency solely on the remaining sixth variable. Six panels for F_1 and for $\mathscr{F}$ show these dependencies, where the sixth variable switches from λ_1 to λ_6, respectively. Three different cases are presented in each small panel; they are obtained by setting five fixed variables to their minimal, mean, and maximal values, respectively.

variational calculus, which is known as *geodesic problem*. The length of a curve, which is given in parametric form $\lambda_i \equiv \lambda_i(t)$, in the n-dimensional space $\boldsymbol{\lambda}$, can be written as [99], [100], [102]–[105]

$$\mathscr{L}[\boldsymbol{\lambda}(t)] = \int_{t_1}^{t_2} \sqrt{\dot{\lambda}_1^{\,2} + \cdots + \dot{\lambda}_n^{\,2}}\; dt = \int_{t_1}^{t_2} \left\| \dot{\boldsymbol{\lambda}}(t) \right\| dt \quad (3)$$

where t_1 and t_2 are, respectively, the initial and final times of movement $\boldsymbol{\lambda}(t_1)$ and $\boldsymbol{\lambda}(t_2)$—their corresponding positions in the internal space $\boldsymbol{\lambda}$. We will now seek the vector-valued function $\boldsymbol{\lambda}(t)$, i.e., the set of functions $\{\lambda_i(t)\}_{i=1}^{n}$, which minimizes this integral under two constraints: acoustical and mechanical.

The acoustical constraint consists in the specification of the initial and final phonetic targets, which correspond, respectively, to the initial t_1, and final t_2 moments. These targets are the zones in the formant space $\boldsymbol{F}$. Formally, the constraint is defined as the appertaining of the first k formants of each produced vowel to its own specific formant zone, which is defined by a k-dimensional *ellipsoidorectangle* in the formant space $\boldsymbol{F}$. In other words,

mathematically, the formants of the jth produced vowel, which are denoted by $\boldsymbol{F}_j \equiv (F_{1,j}, \ldots, F_{k,j})$, must satisfy $H(G_{\mathrm{a},j}) = 0$, where

$$G_{\mathrm{a},j}(\boldsymbol{F}_j) = \sum_{l=1}^{k} \frac{(F_{l,j}(\boldsymbol{\lambda}_j) - \overset{\circ}{F}_{l,j})^{2\eta_j}}{\epsilon_{l,j}^{2\eta_j}} - 1 \tag{4}$$

where $\overset{\circ}{F}_{l,j}$ are the prescribed formants, $F_{l,j}$ are the produced ones, parameters $\epsilon_{l,j}$ define the axes of the formant ellipsoidorectangle, η_j defines its shape (rounded or rectangular), $\boldsymbol{\lambda}_j \equiv \boldsymbol{\lambda}(t_j)$ is the vector motor command that is responsible for the production of jth vowel, and $j = 1$ and 2, since we have only two targets (vowels): the initial one and the final one. By an *ellipsoidorectangle*, we actually mean a voluminous k-dimensional figure, which is obtained from the previous equation by setting $G_{\mathrm{a},j} = 0$. For $\eta_j = 1$, it is a k-dimensional ellipse; then, by increasing the parameter η_j, it becomes more and more rectangular, and finally, for large η_j, it becomes definitively a hyperrectangle. The equality, which is given by $H(G_{\mathrm{a},j}) = 0$, means that we wish the formants of the produced sounds to be inside their ellipsoidorectangles; this can be viewed as a weak constraint, because we do not specify where exactly we want the formants to be; they must be just somewhere inside the ellipsoidorectangles. We could pose $G_{\mathrm{a},j} = -1$, i.e., the strict belonging to the given set of formants of the produced vowels; however, this constraint is too rigid and, in practice, it does not seem very real, since the slight fluctuation of the phonetic targets is always present (actually, parameters $\epsilon_{l,j}$ were precisely introduced for this purpose, i.e., in order to define the size of the formant zones),[8] or in addition, $G_{\mathrm{a},j} = 0$, i.e., the strict appertaining to the surface of the formant ellipsoidorectangle (which is another weak constraint, because we do not specify where exactly on the ellipsoidorectangle's surface the formants must be, but it is stronger than $H(G_{\mathrm{a},j}) = 0$). Furthermore, the constraints of the type $H(G_{\mathrm{a},j}) = 0$ will be called *the constraints of the first kind*; those of the type $G_{\mathrm{a},j} = 0$ will be called *the constraints of the second kind*.

The mechanical constraint simply consists in the equality of the global mean force's level $\mathscr{F}$ to the prescribed value $\overset{\circ}{\mathscr{F}}(t)$ (time-dependent in general), which must be kept during the whole transition between t_1 and t_2:

$$G_{\mathrm{m}}(\boldsymbol{\lambda}, t) = \mathscr{F}(\boldsymbol{\lambda}) - \overset{\circ}{\mathscr{F}}(t) = 0 \tag{5}$$

i.e., this constraint has to be satisfied every time and everywhere and not only in t_1 and t_2, as it is for the acoustical one. Note that the introduction of the constraints in the model aims precisely to mathematically formalize the targets (see Section I). The acoustical constraints represent actually a sort of static targets, which are in accordance with target-normalization theory (especially that of the first kind). In contrast, the mechanical constraint, which is a dynamic one, corresponds to the dynamic-target-specification theory (since it must be satisfied during the transition and not only at the static endpoints belonging to some zone) and aims to better represent the reality of the system.

As we may recall from variational calculus, the function $\boldsymbol{\lambda}(t)$ that minimizes the functional (3) is the solution of the corresponding system of the Euler–Lagrange differential equations. For the ordinary variational problem, which requires the stationarity of the functional

$$Y[\boldsymbol{\lambda}(t)] = \int_{t_1}^{t_2} f(\boldsymbol{\lambda}, \dot{\boldsymbol{\lambda}}, t)\, dt \tag{6}$$

with given fixed boundary conditions $\boldsymbol{\lambda}(t_1) = \boldsymbol{\lambda}_1$, and $\boldsymbol{\lambda}(t_2) = \boldsymbol{\lambda}_2$, under the m constraints $G_j(\boldsymbol{\lambda}, t) = 0$, for $j = 1, \ldots, m$, the solution can be found from the following system of n Euler–Lagrange equations:

$$\left\{ \frac{\partial}{\partial \lambda_i} \left(f + \sum_{j=1}^{m} \mu_j G_j \right) - \frac{d}{dt} \frac{\partial}{\partial \dot{\lambda}_i} \left(f + \sum_{j=1}^{m} \mu_j G_j \right) = 0 \right. \tag{7}$$

where $\mu_j \equiv \mu_j(t)$ are the Lagrange's undetermined multipliers. The latter equation may be reduced to the following one, which is represented in vector form as

$$\frac{\partial f}{\partial \boldsymbol{\lambda}} + \sum_{j=1}^{m} \mu_j \frac{\partial G_j}{\partial \boldsymbol{\lambda}} - \frac{d}{dt} \frac{\partial f}{\partial \dot{\boldsymbol{\lambda}}} = \mathbf{0} \tag{8}$$

where $\partial / \partial \boldsymbol{\lambda}$ is the operator of partial differentiation with respect to each component of the vector $\boldsymbol{\lambda}$. Note that since we have n differential partial equations and m equations of constraint, we can find all n components of $\boldsymbol{\lambda}(t)$ and m Lagrange's multipliers; the remaining $2n$ unknowns, due to n differential equations of second order, can be found from the $2n$ boundary or initial conditions. Note also that the constraints G_j may be static or dynamic, that does not change the previously mentioned differential equation, since they do not contain $\dot{\boldsymbol{\lambda}}$ (for more information, see [100] and [104], where we can also find the cases of the constraints given as ODEs). Obviously, similar reasoning also applies to the mechanical constraint (5).

For the functional (3), the Euler–Lagrange equations are particularly simple, because the integrand contains only the derivative of $\boldsymbol{\lambda}(t)$, which is a particularity of all geodesic problems, i.e., we have the following system:

$$\left\{ \frac{d}{dt} \frac{\dot{\lambda}_i}{\sqrt{\dot{\lambda}_1^2 + \cdots + \dot{\lambda}_n^2}} - \mu \frac{\partial \mathscr{F}}{\partial \lambda_i} = 0 \qquad \forall\, i = 1, \ldots, n \right. \tag{9}$$

with the additional equation (5) to find $\mu(t)$. The latter expression can also be written as

$$\frac{d}{dt} \frac{\dot{\boldsymbol{\lambda}}}{\|\dot{\boldsymbol{\lambda}}(t)\|} - \mu \frac{\partial \mathscr{F}}{\partial \boldsymbol{\lambda}} = \mathbf{0}\,. \tag{10}$$

After the total differentiation with respect to time, it becomes

$$\frac{\ddot{\boldsymbol{\lambda}} \langle \dot{\boldsymbol{\lambda}}, \dot{\boldsymbol{\lambda}} \rangle - \dot{\boldsymbol{\lambda}} \langle \dot{\boldsymbol{\lambda}}, \ddot{\boldsymbol{\lambda}} \rangle}{\|\dot{\boldsymbol{\lambda}}\|^3} - \mu \frac{\partial \mathscr{F}}{\partial \boldsymbol{\lambda}} = \mathbf{0} \tag{11}$$

[8] For instance, it is well known that the distribution of formants about its mean $\overline{\boldsymbol{F}}_j$ is near-normal. Thus, for the particular case $\eta_j = 1$, the parameters $\boldsymbol{\epsilon}_j$ define the formant zone of the constant probability level a^{-1}, $a > 1$, with respect to the maximum level at $\overline{\boldsymbol{F}}_j$, if we pose the referents $\overset{\circ}{\boldsymbol{F}}_j = \overline{\boldsymbol{F}}_j$ and $\boldsymbol{\epsilon}_j$ equal to $\sqrt{2 \ln a}$ × standard deviation of the aforementioned normal distribution; in other words, $\boldsymbol{\epsilon}_j$ define the formant equiprobability's ellipses.

where $\langle\cdot\rangle$ denotes the scalar product. Moreover, for the particular case $n = 3$, by using the well-known rule of linear algebra transforming the scalar products into the vector ones, we may simplify (11) as follows:

$$\frac{\dot{\boldsymbol{\lambda}} \times (\ddot{\boldsymbol{\lambda}} \times \dot{\boldsymbol{\lambda}})}{\|\dot{\boldsymbol{\lambda}}\|^3} - \mu \frac{\partial \mathscr{F}}{\partial \boldsymbol{\lambda}} = \mathbf{0}. \tag{12}$$

We can even generalize (9)–(12), for the cases when there are more than one constraint that must be fulfilled along optimum solution $\boldsymbol{\lambda}(t)$, i.e., in each point of $\boldsymbol{\lambda}(t)$. By using (7) or (8), we can generalize (11), as follows:

$$\frac{\ddot{\boldsymbol{\lambda}} \langle \dot{\boldsymbol{\lambda}}, \dot{\boldsymbol{\lambda}} \rangle - \dot{\boldsymbol{\lambda}} \langle \dot{\boldsymbol{\lambda}}, \ddot{\boldsymbol{\lambda}} \rangle}{\|\dot{\boldsymbol{\lambda}}\|^3} - \sum_{j=1}^{m} \mu_j \frac{\partial G_j}{\partial \boldsymbol{\lambda}} = \mathbf{0}. \tag{13}$$

Note that the acoustical constraints [see (4)] are not present in any of these equations. This is because they are related only to the boundary conditions but not to the whole path $\boldsymbol{\lambda}(t)$, which is precisely the complexity of our case. In fact, (9)–(13) and (5) are only the necessary conditions to which the optimum solution $\boldsymbol{\lambda}(t)$ must satisfy, and they are not sufficient for its complete determination. Generally, the latter is carried out with the help of the boundary conditions. However, in our case, these conditions are not given explicitly but implicitly via the acoustical constraints (4), i.e., t_1 is related to $G_{\mathrm{a},1}$, t_2, and to $G_{\mathrm{a},2}$, as follows:

$$t_j \Rightarrow \boldsymbol{\lambda}_j \Rightarrow \boldsymbol{F}_j \Rightarrow G_{\mathrm{a},j}, \qquad j = 1, 2. \tag{14}$$

In this case, which is often called in literature the *undetermined endpoints case* [100], the optimum function $\boldsymbol{\lambda}(t)$ must also satisfy a supplementary system of the differential equations,[9] involving the derivatives of the acoustical constraints, and this condition is sufficient to completely determine the optimum solution $\boldsymbol{\lambda}(t)$, thereby giving the trajectory of motion in the internal space.

The solution of such a system of partial differential equations represents a quite complicated problem of mathematical physics (we recall that the dependencies $\boldsymbol{F}(\boldsymbol{\lambda})$ and $\mathscr{F}(\boldsymbol{\lambda})$ are given by two nonlinear ANNs). Thus, first, we would like to discuss some mathematical issues related to the (9)–(13), raising serious questions about validity of some variants of the EPH, and then propose a solution for the problem.

D. Discussion of the Drawbacks of the Linearized λ-Model

1) Brief Description of the Linearized λ-Model: Classic EPH does not imply any dynamic description of the applied motor commands $\boldsymbol{\lambda}$. In order to bring some dynamics to the system, the so-called *linearized λ-model* was proposed. This variant of the EPH is basically the classic λ-model, supplementary supposing the time transitions between static positions in the internal space $\boldsymbol{\lambda}$ may be effected only at a constant rate. In other words, during the change of posture, from one to another, the commands $\boldsymbol{\lambda}$ are modified linearly with time, i.e., $\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} t + \boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the constant coefficients (vectors). We have already mentioned this model in Section I; it is a simplest implementation of the shortest distance principle in the internal motor-command space. However, we will show that this model works only in trivial cases, and it does not support dynamic systems well.

2) Contradiction With the Principle of the Shortest Path: We will show now that if the constraints on the form of the path, i.e., $G_j(\boldsymbol{\lambda}, t)$, are nonlinear[10] in $\boldsymbol{\lambda}$, the principle of the shortest path and the linearized λ-model are in mathematical contradiction.

Reductio ad absurdum: The unique solutions that the linearized λ-model allows are the linear ones: $\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} t + \boldsymbol{\beta}$. If we now substitute these linear solutions into (13), we obtain that the left part of this equation is always equal to zero, which also means that the remaining part is always zero; however, the dependencies $G_j(\boldsymbol{\lambda}, t)$, $j = 1, \ldots, m$ are, in general, nonlinear in $\boldsymbol{\lambda}$, and therefore, the sum of their derivatives cannot be null in all cases. Furthermore, if at least one of these dependencies is nonlinear in $\boldsymbol{\lambda}$, the sum of their derivatives cannot be null, which means that the first term in the left part of (9)–(13) cannot be always null, and thus, the optimum solutions cannot be linear functions $\boldsymbol{\lambda}(t) \neq \boldsymbol{\alpha} t + \boldsymbol{\beta}$. It is simple to show that the unique case when this first term vanishes is that when $\boldsymbol{\lambda}(t)$ becomes a linear function. An extremely simple proof, which is based on geometry, may be done for the particular 3-D internal space case. From (12), we have $\dot{\boldsymbol{\lambda}} \times (\ddot{\boldsymbol{\lambda}} \times \dot{\boldsymbol{\lambda}}) = \mathbf{0}$, which means either $\dot{\boldsymbol{\lambda}}$ is parallel to $(\ddot{\boldsymbol{\lambda}} \times \dot{\boldsymbol{\lambda}})$, or one of them is zero. The parallelism is impossible because $(\ddot{\boldsymbol{\lambda}} \times \dot{\boldsymbol{\lambda}})$ is orthogonal to both of its arguments, and one of them is precisely $\dot{\boldsymbol{\lambda}}$. Thus, one of these vectors is null. In the most general case, $\ddot{\boldsymbol{\lambda}} = \mathbf{0}$, and therefore, $\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} t + \boldsymbol{\beta}$.

It is important to note that the impossibility of the linear solutions is not due to any particular formulation of the constraints but to the nature of the geodesic problem itself, for which the solutions are, in fact, always determined by the constraints (e.g., for the original geodesic problem, the Earth's surface determines the corresponding solutions). If the constraints are nonlinear, the optimum solutions $\boldsymbol{\lambda}(t)$ cannot be linear. Thus, only the trivial constraints related to the whole traveled path $\boldsymbol{\lambda}(t)$ [e.g., those described in footnote 7] can be compatible with the linearized λ-model. These findings cast doubts on the linearized λ-model for the task planning and its general use in motor-control theories.

3) Contradiction With the Finite-Energy and Finite-Power Principles: Another contradiction is that with the finite energy and power principles. Generally, the processes having finite energy and, especially, finite power, are said to be physically stable (note that some processes can have the infinite energy, but finite power, for instance, classic small nondamped harmonic oscillations). It is not complicated to show that the Feldman's linearized λ-model (both static and dynamic variants; see Section I) may lead to the mechanical process of infinitely growing energy and power. We decided to compare four spring models: the classic linear damped spring model, the exponential damped spring model, the exponential damped spring model controlled by linear λ command (i.e., static Feldman's model

[9] The so-called left-hand and right-hand endpoint requirements [100].

[10] Or even linear, but in other contexts than described in footnote 7.

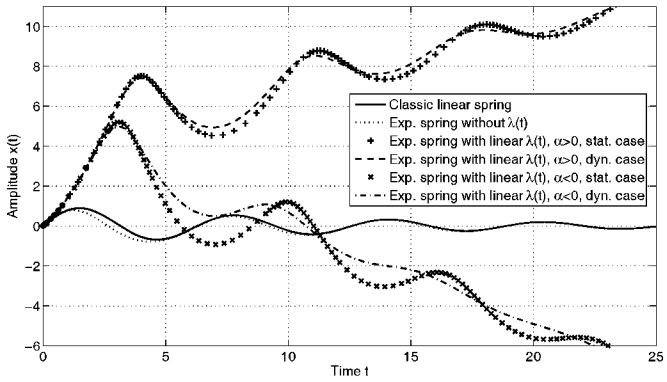


Fig. 5. The behaviors of the classic linear spring, the exponential spring, and the exponential springs of the Feldman's linearized λ-model (static and dynamic variants) for positive and negative α.

with $\lambda(t) = \alpha t + \beta$), and the exponential damped spring model controlled by linear λ command with the adjustment of feedback related to the current velocity[11] (i.e., dynamic Feldman's model with $\lambda(t) = \alpha t + \beta$). The first model is classically described by a second-order ODE of motion: $\ddot{x} + \gamma\dot{x} + \rho x = 0$. The second one can be described by

$$\ddot{x} + \gamma\dot{x} + \rho(e^{\eta x} - 1) = 0 \tag{15}$$

the third model by

$$\ddot{x} + \gamma\dot{x} + \rho(e^{\eta x - \alpha t - \beta} - 1) = 0 \tag{16}$$

and the fourth one by

$$\ddot{x} + \gamma\dot{x} + \rho(e^{\eta x + \kappa\dot{x} - \alpha t - \beta} - 1) = 0 \tag{17}$$

the displacement being a function of time, i.e., $x \equiv x(t)$, and γ, ρ, η, κ, α, and β being constant parameters. These parameters are (for all models) given by $\gamma = 0.16$, $\rho = 1$, $\eta = 1$, $\kappa = 0.05$, $\alpha = +0.25$ (case $\alpha > 0$), $\alpha = -0.50$ (case $\alpha < 0$), and $\beta = 5$. The initial conditions that are fixed for all differential equations are given by $x(0) = 0$, and $\dot{x}(0) = 1$. Unfortunately, only the first equation has an exact analytical solution. We, therefore, had to resort to numerical methods in order to obtain the corresponding solutions (namely, we used the `ode45` MATLAB's function, which is based on the Runge–Kutta method). The results are shown in Fig. 5. From this figure, we can ascertain that both linear and exponential noncontrolled spring models have a stable exponentially damped oscillating solutions. On the contrary, for the exponential spring models, which are controlled by linear λ command, the solutions are infinitely increasing tending to $+\infty$ for $\alpha > 0$ and infinitely decreasing tending to $-\infty$ for $\alpha < 0$; therefore, for both latter cases, they are not in the $\mathbb{L}^1$, for $t \in [0, \infty)$, and neither in $\mathbb{L}^2$, for $t \in [0, \infty)$, and the energy and power of such a process tend to infinity.

4) Potentially Limited System Dynamics: Finally, third drawback of linearized λ-models lie in the fact that the prescription of linear variations of $\boldsymbol{\lambda}(t)$ may potentially limit the dynamics of the system. Taking into account that the dependency $\mathscr{F}(\boldsymbol{\lambda})$ is associated with the physics of the system (its biomechanics) and cannot be changed without relearning, the prescription of linear $\boldsymbol{\lambda}(t)$ impacts to the dynamics of the system. In this case, it is clear that if $\mathscr{F}(\boldsymbol{\lambda})$ is linear in $\boldsymbol{\lambda}$, so does $\mathscr{F}(t) \equiv \mathscr{F}(\boldsymbol{\lambda}(t))$ in t; if $\mathscr{F}(\boldsymbol{\lambda})$ is quadratic in $\boldsymbol{\lambda}$, so does $\mathscr{F}(t)$ in t, etc. In this context, it is difficult to imagine how, with linearized λ-model, one can truly implement dynamic-target-specification theory, thus allowing, for example, some variability of transitions (see Section I), if the system dynamics are fixed by its biomechanics and the CNS influence is taken into account only in a sketchy linear form.

[11] It is sometimes called the proprioceptive feedback.

E. Solution in First Approximation

The exact analytical solution of the optimum internal model problem, which is described in Section II-C, is not simple. The solution of a system of partial differential equations, which imply neural networks, and whose boundary conditions are not given explicitly, is a quite complicated problem of mathematical physics. On the other hand, one can easily see from Fig. 4 that the dependencies $\mathscr{F}(\boldsymbol{\lambda})$ are not strongly nonlinear, and therefore, they may be replaced by the straight lines in first approximation. As to the dependencies $F(\boldsymbol{\lambda})$, they are strongly nonlinear; however, as it follows from Section II-C, these constraints do not affect the form of the solution but only its endpoints. Thus, basing on Section II-D2, the optimum solutions become the straight lines $\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} t + \boldsymbol{\beta}$, and the undetermined coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the variables by which the mechanical and acoustical constraints must be satisfied. In other words, in our case, where the constraints on the form of the path are not far from the linear ones, the linearized λ-model can be actually viewed as the first approximation to the global problem of finding the optimal path. We can no longer "play" with the form of this optimum path $\boldsymbol{\lambda}(t)$ (that was precisely the main role of variational calculus) but only with the limits of the integral (3), which are written in implicit form, i.e., $\boldsymbol{\lambda}_1 \equiv \boldsymbol{\lambda}(t_1)$ and $\boldsymbol{\lambda}_2 \equiv \boldsymbol{\lambda}(t_2)$. Since the dependency $\mathscr{F}(\boldsymbol{\lambda})$ is now the linear one and the optimum solution $\boldsymbol{\lambda}(t)$ is the straight line, it is sufficient that the constraint (5) was satisfied only in two points (e.g., at the ends $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$) in order to be satisfied in every point of the optimum solution $\boldsymbol{\lambda}(t)$. Mathematically, this means that the mechanical constraint (5), which was initially on the form of the path, now becomes that on the boundary conditions. As to the acoustical constraints (4), it remains as before on the boundary conditions $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

By substituting linear solutions $\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} t + \boldsymbol{\beta}$ into the functional (3), we obtain the following well-known formula of the straight line's length in n-dimensional space:

$$L(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \int_{t_1}^{t_2} \left\|\boldsymbol{\alpha}\right\| dt = \left\|\boldsymbol{\alpha}\right\|(t_2 - t_1) = \left\|\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1\right\|.$$

Moreover, we extend this approach to more than two vowels between which the commands are linear, say, p vowels. In this

case, the total length becomes

$$L(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p) = \sum_{j=1}^{p-1}\int_{t_j}^{t_{j+1}} \|\boldsymbol{\alpha}\|\,dt = \sum_{j=1}^{p-1}\|\boldsymbol{\lambda}_{j+1}-\boldsymbol{\lambda}_j\| = \sum_{j=1}^{p-1}\sqrt{\sum_{i=1}^{n}(\lambda_{i,j+1}-\lambda_{i,j})^2}. \quad (18)$$

Obviously, in this case, the boundary conditions have to be fulfilled in the points $\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\ldots,\boldsymbol{\lambda}_p$, instead of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

We now state the exact mathematical formulation of the problem: We seek to extremize the function $L(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p)$, with respect to the variables $(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p)$, under p acoustical[12] and p mechanical constraints [see also (4) and (5)]

$$\begin{cases} G_{\mathrm{a},j}(\boldsymbol{\lambda}_j)=0 & \forall\, j=1,\ldots,p \\ G_{\mathrm{m},j}(\boldsymbol{\lambda}_j)=0 & \forall\, j=1,\ldots,p. \end{cases} \quad (19)$$

Thus, instead of the optimization with respect to the form of the traveled path and its ends (defined implicitly via the boundary conditions), we now carry out the optimization only with respect to the ends $(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p)$. In other words, the initial problem of the constrained optimization of functional becomes that of the constrained optimization of function, which is usually more simple to solve.

The latter optimization problem is classically solved by means of the Lagrange's undetermined multipliers method for functions. This method consists in introducing a composite function $U(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p,\boldsymbol{\mu}_\mathrm{a},\boldsymbol{\mu}_\mathrm{m})$ of $n\times p+2p$ variables, which is a sum of function $L(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p)$ of $n\times p$ variables and of $2p$ constraints (19), which are weighted by the corresponding Lagrange's multipliers $\boldsymbol{\mu}_\mathrm{a}\equiv(\mu_{\mathrm{a},1},\ldots,\mu_{\mathrm{a},p})$, and $\boldsymbol{\mu}_\mathrm{m}\equiv(\mu_{\mathrm{m},1},\ldots,\mu_{\mathrm{m},p})$

$$U = L + \sum_{h=1}^{p}\mu_{\mathrm{a},h}G_{\mathrm{a},h}(\boldsymbol{\lambda}_h) + \sum_{h=1}^{p}\mu_{\mathrm{m},h}G_{\mathrm{m},h}(\boldsymbol{\lambda}_h). \quad (20)$$

Here, index j was replaced by h in order not to get confused with the further derivatives. The optimization itself consists in finding the optimum set $(\widetilde{\boldsymbol{\lambda}}_1,\ldots,\widetilde{\boldsymbol{\lambda}}_p,\widetilde{\boldsymbol{\mu}}_\mathrm{a},\widetilde{\boldsymbol{\mu}}_\mathrm{m})$ such that

$$\left.\frac{\partial U}{\partial\boldsymbol{\lambda}_j}\right|_{\boldsymbol{\lambda}_j=\widetilde{\boldsymbol{\lambda}}_j}=\mathbf{0},\qquad \left.\frac{\partial U}{\partial\boldsymbol{\mu}_\mathrm{a}}\right|_{\boldsymbol{\mu}_\mathrm{a}=\widetilde{\boldsymbol{\mu}}_\mathrm{a}}=\mathbf{0},\qquad \left.\frac{\partial U}{\partial\boldsymbol{\mu}_\mathrm{m}}\right|_{\boldsymbol{\mu}_\mathrm{m}=\widetilde{\boldsymbol{\mu}}_\mathrm{m}}=\mathbf{0} \quad j=1,\ldots,p \quad (21)$$

or, in short, $\mathrm{grad}\,U=\mathbf{0}$. We recall that the equality of the derivatives of U to zero with respect to the Lagrange's multipliers gives actually the constraints given in (19), which is precisely the interest of the Lagrange's method: The constrained optimization of the function L of $n\times p$ variables reduces to the unconstrained one of the function U of $n\times p+2p$ variables, where the last $2p$ variables are the Lagrange's undetermined multipliers.

[12]Note that use of the acoustical constraints of the second kind instead of the first one cannot be considered too restrictive for our problem from an acousticophonetic point of view. On the one hand, the position of the borders is variable and can be adjusted via the parameters $\epsilon_{l,j}$ and $\overset{\circ}{F}_{l,j}$. On the other hand, as we shall see later, almost always, the solution for the constraints of the second kind is also that for the constraints of the first kind.

The procedure of differentiation is quite particular for the length's function $L(\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_p)$. The derivative with respect to $\boldsymbol{\lambda}_j$ is not always calculated in the same way; this is because, the first $\boldsymbol{\lambda}_1$ and the last $\boldsymbol{\lambda}_p$ terms are present only once in the sum (18), while all the intermediate terms $\boldsymbol{\lambda}_2,\ldots,\boldsymbol{\lambda}_{p-1}$ are present twice. Thus, by differentiating with respect to each component of $\boldsymbol{\lambda}_1$, we obtain the following system:

$$\begin{cases} \dfrac{\partial L}{\partial\lambda_{i,1}} = \dfrac{-(\lambda_{i,2}-\lambda_{i,1})}{\sqrt{\sum_{i=1}^{n}(\lambda_{i,2}-\lambda_{i,1})^2}}, & i=1,\ldots,n \end{cases} \quad (22)$$

or in the following vector form:

$$\frac{\partial L}{\partial\boldsymbol{\lambda}_1} = -\frac{\boldsymbol{\lambda}_2-\boldsymbol{\lambda}_1}{\|\boldsymbol{\lambda}_2-\boldsymbol{\lambda}_1\|}. \quad (23)$$

With respect to $\boldsymbol{\lambda}_2,\ldots,\boldsymbol{\lambda}_{p-1}$, we obtain

$$\frac{\partial L}{\partial\boldsymbol{\lambda}_j} = \frac{\boldsymbol{\lambda}_j-\boldsymbol{\lambda}_{j-1}}{\|\boldsymbol{\lambda}_j-\boldsymbol{\lambda}_{j-1}\|} - \frac{\boldsymbol{\lambda}_{j+1}-\boldsymbol{\lambda}_j}{\|\boldsymbol{\lambda}_{j+1}-\boldsymbol{\lambda}_j\|} \quad (24)$$

where $j=2,\ldots,p-1$. Finally, for $j=p$, it yields

$$\frac{\partial L}{\partial\boldsymbol{\lambda}_p} = \frac{\boldsymbol{\lambda}_p-\boldsymbol{\lambda}_{p-1}}{\|\boldsymbol{\lambda}_p-\boldsymbol{\lambda}_{p-1}\|}. \quad (25)$$

The differentiation of the constraints is less sophisticated. For the acoustical one, it becomes

$$\frac{\partial G_{\mathrm{a},h}(\boldsymbol{\lambda}_h)}{\partial\boldsymbol{\lambda}_j} = \begin{cases} 2\eta_j\displaystyle\sum_{l=1}^{k}\frac{\left(F_{l,j}(\boldsymbol{\lambda}_j)-\overset{\circ}{F}_{l,j}\right)^{2\eta_j-1}}{\epsilon_{l,j}^{2\eta_j}}\cdot\frac{\partial F_{l,j}}{\partial\boldsymbol{\lambda}_j}, & h=j \\ 0, & h\neq j \end{cases}$$

for $j=1,\ldots,p$, where the derivatives of the lth formant of jth vowel with respect to the motor commands of this vowel $\boldsymbol{\lambda}_j$ are calculated according to (36), i.e.,

$$\frac{\partial F_{l,j}}{\partial\boldsymbol{\lambda}_j} = 2\sum_{s=1}^{S_1}b_1^2 v_{s,l}(\boldsymbol{w}_s-\boldsymbol{\lambda}_j)\,e^{-\left(\|\boldsymbol{w}_s-\boldsymbol{\lambda}_j\|\,b_1\right)^2} \quad (26)$$

for $l=1,\ldots,k$, $j=1,\ldots,p$, and $\dim\boldsymbol{\lambda}_j=n$ (i.e., in all $k\times p\times n$ derivatives). It may be noted that since the constraints on the boundary conditions of the hth vowel are independent from the jth vowel, these derivative are all null. The derivatives of the mechanical constraint are simply given by

$$\frac{\partial G_{\mathrm{m},h}(\boldsymbol{\lambda}_h)}{\partial\boldsymbol{\lambda}_j} = \begin{cases} \dfrac{\partial\mathscr{F}(\boldsymbol{\lambda}_j)}{\partial\boldsymbol{\lambda}_j}, & h=j \\ 0, & h\neq j \end{cases} \quad (27)$$

where the last derivatives are calculated similarly to the formant ones (without index l and with the ANN weights corresponding to the $\boldsymbol{\lambda}\Leftrightarrow\mathscr{F}$ network).

The system (21) cannot be solved analytically. The optimization of U was, therefore, performed numerically by means of the gradient descent method (which is also known as the method of steepest descent), implemented in the MATLAB programming language.

TABLE I
FREQUENCIES (IN HERTZ) WE USED FOR THE DEFINITION OF THE PHONETIC TARGETS FOR THE ACOUSTICAL CONSTRAINT

Vowel (IPA/MATLAB)	$\overset{\circ}{F}_{1,j}$ ($\epsilon_{1,j}$)	$\overset{\circ}{F}_{2,j}$ ($\epsilon_{2,j}$)
i/i	290 (90)	2181 (90)
i/i *	270 (90)	2290 (90)
e/e	375 (45)	2097 (105)
ɛ/E	505 (60)	1710 (120)
a/a	628 (60)	1284 (45)
œ/oe	497 (38)	1488 (120)
ɔ/c	550 (45)	975 (90)
ɔ/c *	570 (56)	840 (80)
y/y	321 (45)	1685 (300)

Note, that in some cases, formant ellipsoidorectangles are defined artificially in order to better observe the desired effects. As to the formant frequencies, they are taken from [106], except those with * taken from [86] and [107] (the latter are used only in Section III-B).

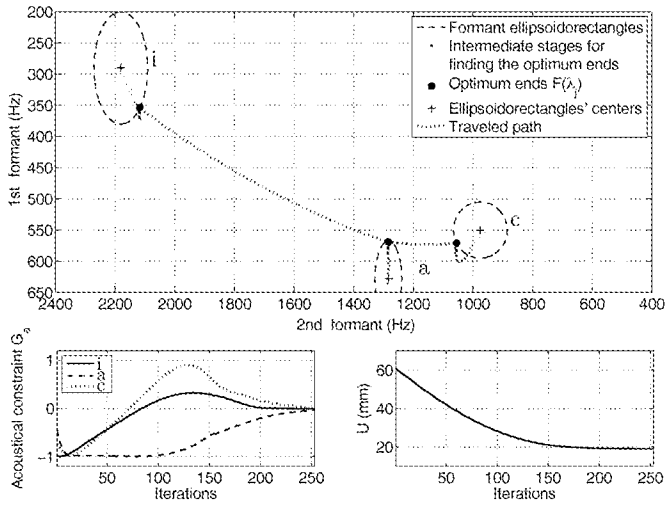


Fig. 6. Optimization of the sequence [i a ɔ]. Formant zones are defined by the corresponding formant ellipses. Notation: "Optimum ends $F(\lambda_j)$" $\equiv (\widetilde{\boldsymbol{F}}_1, \ldots, \widetilde{\boldsymbol{F}}_p)$.

III. RESULTS

A. Only Acoustical Constraints Are Applied

1) Simulations With the Model and Effect of Different Task-Planning Strategies: We first considered a slightly simpler case, which is one without the third term in (20), i.e., without mechanical constraints (5) at all. First of all, we fixed static phonetic targets (i.e., prescribed formants) according to Table I. Besides, the formant frequencies $\overset{\circ}{\boldsymbol{F}}_j$ are also used for the initialization of optimization algorithm: From the initial database used for the ANN learning, we search for the couples of data $\boldsymbol{\lambda} \Leftrightarrow \boldsymbol{F}$, having $\boldsymbol{F}$ as close as possible to the formant zone's centers $\overset{\circ}{\boldsymbol{F}}_j$ (these frequencies are generally found in the range ± 10 Hz for the first formant, and ± 30 Hz for the second one). The corresponding motor commands $\boldsymbol{\lambda}$ are taken as initial points for the gradient descent method. Note also that since there is no bijection between spaces $\boldsymbol{\lambda}$ and $\boldsymbol{F}$, there is a finite $(n-k)$-dimensional volume in $\boldsymbol{\lambda}$, any point of which maps to the same point in $\boldsymbol{F}$, e.g., to the projection of the initial point $\overset{\circ}{\boldsymbol{F}}_j$. Thus, the initial point in $\boldsymbol{\lambda}$ cannot be uniquely determined. Indeed, taking into account that the numerical optimization method is the gradient descent one, the optimum solution may be potentially influenced by the choice of this initial point $\boldsymbol{\lambda}$. However, the projection of the initial point (i.e., ellipsoidorectangle's center), regardless of the initial point itself in $\boldsymbol{\lambda}$, is quite close to the solution (i.e., to the ellipsoidorectangle's border), and taking into account that the function $L(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$, which is given by (18), is "sufficiently" convex, and the constraints (19) can be considered locally monotonic[13] (see Fig. 4), and therefore, the local minima problem does not really affect the solution. On the other hand, we also tested this potential dependency empirically, and the optimization algorithm always returned the same final solution (with ANN accuracy), regardless of the initial point.

The optimization results given by our algorithm for the sequence of three vowels [i a ɔ] are shown in Fig. 6. As we can observe from this figure, the formant zones were defined as ellipses, i.e., all $\eta_j = 1$. The set of the Lagrange's multipliers found by our optimization algorithm is $\widetilde{\boldsymbol{\mu}}_a = (0.0670, 0.1700, 0.0880)$ mm, and that of the optimal motor commands $\widetilde{\boldsymbol{\lambda}}_1, \ldots, \widetilde{\boldsymbol{\lambda}}_p$ (in millimeters) is given by

$$\begin{aligned}
\widetilde{\boldsymbol{\lambda}}_1 &= (33.76, 48.12, 46.84, 74.92, 18.38, 63.78), && \text{for [i]}\\
\widetilde{\boldsymbol{\lambda}}_2 &= (46.96, 48.02, 41.78, 74.04, 18.70, 63.90), && \text{for [a]}\\
\widetilde{\boldsymbol{\lambda}}_3 &= (51.04, 49.44, 41.21, 72.02, 18.74, 64.02), && \text{for [ɔ]}
\end{aligned}$$

where the gradient step was set to 0.0125. The set of the corresponding formants $(\widetilde{\boldsymbol{F}}_1, \ldots, \widetilde{\boldsymbol{F}}_p) \equiv (\boldsymbol{F}(\widetilde{\boldsymbol{\lambda}}_1), \ldots, \boldsymbol{F}(\widetilde{\boldsymbol{\lambda}}_p))$, which is the projection of the found optimal commands to the formant space, is given by (in hertz, first and second formants)

$$\begin{aligned}
\widetilde{\boldsymbol{F}}_1 &= (353.3, 2117.2), && \text{for [i]}\\
\widetilde{\boldsymbol{F}}_2 &= (569.1, 1284.4), && \text{for [a]}\\
\widetilde{\boldsymbol{F}}_3 &= (570.9, 1054.5), && \text{for [ɔ].} \qquad (28)
\end{aligned}$$

We can clearly observe that when the optimization is finished, all three acoustical constraints vanish, and the function U reaches its minimum, which is designated by $\widetilde{U}$, and is equal to 18.962 mm. A total of 253 iterations of the method of steepest descent for the Lagrange's multipliers method were necessary to find this optimal solution. Note that the path traveled in the formant space $\boldsymbol{F}$ is shown in dots, in order to emphasize the fact that the real path is actually traveled in the internal space $\boldsymbol{\lambda}$, and the path shown in the formant space is only its projection to $\boldsymbol{F}$. Analogously, the optimum endpoints are found in the internal space $\boldsymbol{\lambda}$, and we showed their projections to the formant space. Note that they are exactly on ellipse's borders, that means the fulfillment of the acoustical constraints from (19) (the same can be actually directly observed from the lower left panel of Fig. 6).

We present now a different case of optimization: All initial parameters are the same, except the definition of the formant zone

[13] One may also note that the projections of the traveled paths onto the formant space $\boldsymbol{F}$ (shown in dots in the figures) are just slightly curved in all experiments.

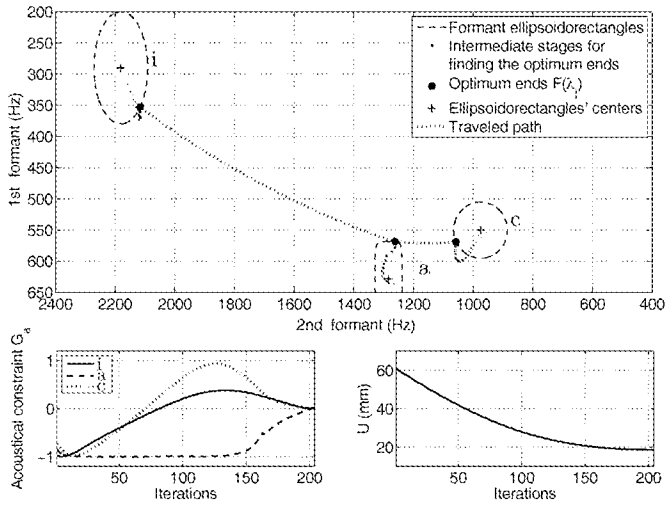


Fig. 7. Optimization of the sequence [i a ɔ]. Formant zones of [i] and [ɔ] are defined by formant ellipses (i.e., $\eta_1 = \eta_3 = 1$) and that of [a] by an ellipsoidorectangle with parameter $\eta_2 = 3$.

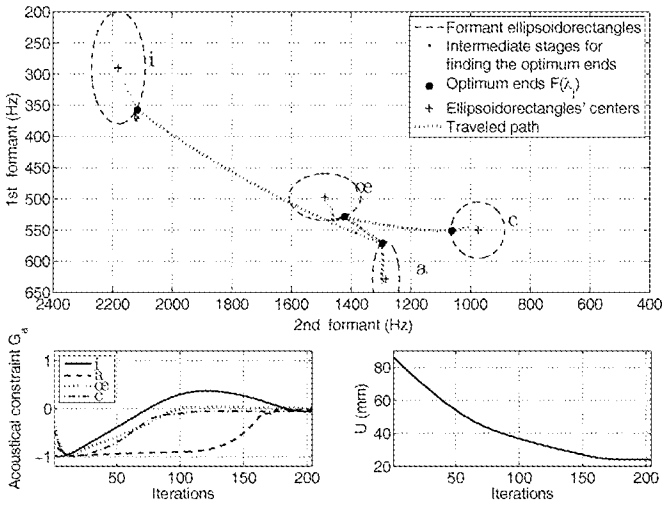


Fig. 8. Optimization of the sequence [i a œ ɔ]. Formant zones are defined by the corresponding formant ellipses.

for the vowel [a], which is defined by an ellipsoidorectangle with parameter $\eta_2 = 3$ (see Fig. 7). For this case, the optimization algorithm found: $\widetilde{\boldsymbol{\mu}}_a = (0.0645, 0.0580, 0.0850)$ mm, and $\widetilde{U} = 18.693$ mm. The set of the formants $(\widetilde{\boldsymbol{F}}_1, \ldots, \widetilde{\boldsymbol{F}}_p)$, which corresponds to the found optimal commands, is given by (in hertz, first and second formants)

$$\begin{aligned} \widetilde{\boldsymbol{F}}_1 &= (352.7, 2115.4), \qquad \text{for [i]} \\ \widetilde{\boldsymbol{F}}_2 &= (568.1, 1262.5), \qquad \text{for [a]} \\ \widetilde{\boldsymbol{F}}_3 &= (568.9, 1057.3), \qquad \text{for [ɔ].} \end{aligned} \tag{29}$$

By comparing two latter sets (28) and (29), or Figs. 6 and 7, we note that the main difference is the formants of [a], especially the second formant; in the first case, it is 21.9 Hz (1.7%) greater than that in the second one. This actually means that by accenting differently the same vowel in the same sequence, we can obtain its different acoustical variants. Note that the parameter $\eta = 3$ means, on the one hand, a different geometrical form of the formant zone, and on the other hand, when the solution reaches its border and starts to leave the formant zone, the acoustical constraint increases much more strongly than that for the normal ellipse given by $\eta = 1$. Thus, the acoustical constraint defines not only the formant zone but the degree of accentuation of this zone as well (note that we employ the word "accentuation" especially in this sense). Therefore, even the small simple changes of strategy of task planning can have an impact on the formant space $\boldsymbol{F}$. Note also that the character of the impact is mostly local, i.e., other vowels of the sequence, whose strategy remained unchanged, showed a relatively small modifications of their positions in the formant space $\boldsymbol{F}$.

2) Phenomenon of Task Anticipation: We found out that our model is in accordance with the phenomenon of *task anticipation* or, more precisely, with its acoustic variant observed for a long time by phoneticians and which is known in phonetics as *coarticulation* or *effect of phonetic environment* [80]–[82], [108]–[112]. In the sequence of several vowels, this phenomenon represents the influence of the following vowel(s) on the previous one(s); especially, it concerns two vowels following one after another, i.e., in the sequence of p vowels, the jth vowel is especially influenced by the $(j+1)$th vowel $\forall\, j = 1, \ldots, p-1$. Acoustically, this phenomenon can be observed in terms of formants. Usually, this phenomenon, in different degrees (depending on the context and other conditions, e.g., accentuation), is present in real speech, which is why we wanted to find out if the proposed model was able to reproduce it.

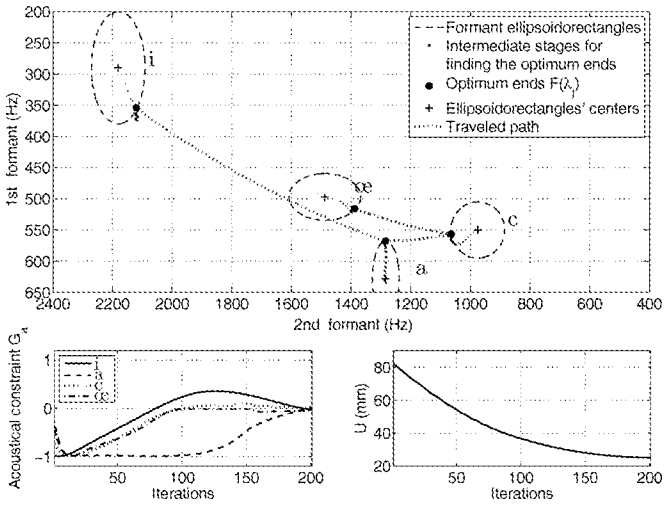


Fig. 9. Optimization of the sequence [i a ɔ œ]. Formant zones are defined by the corresponding formant ellipses.

For demonstration, we choose an example where the triple phenomenon of the task anticipation is produced: sequences [i a œ ɔ] versus [i a ɔ œ]. Thus, the acoustical anticipation will be studied on vowels [a], [ɔ], and [œ]. In addition, two different planning strategies will be compared. The optimizations of the sequences with elliptic planning strategies related to the constraints are performed in Figs. 8 and 9. For the former, the optimization algorithm returns $\widetilde{\boldsymbol{\mu}}_a = (0.0673, 0.1770, 0.1200, 0.0780)$ mm, $\widetilde{U} = 23.989$ mm, and the

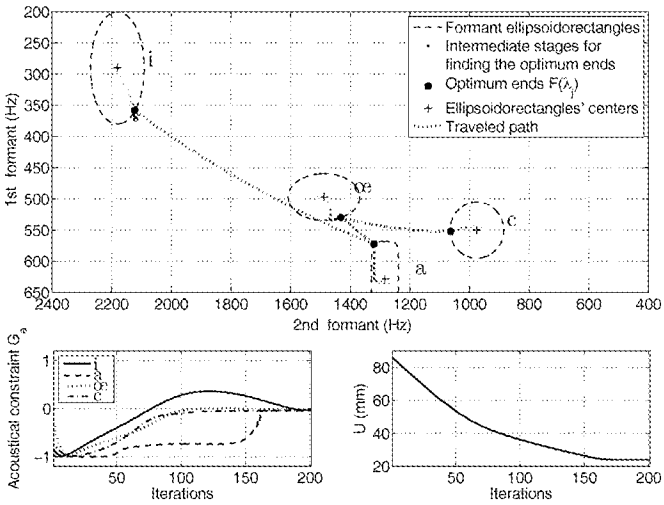


Fig. 10. Optimization of the sequence [i a œ ɔ]. Formant zone of [a] is defined by an ellipsoidorectangle of $\eta_2 = 3$.

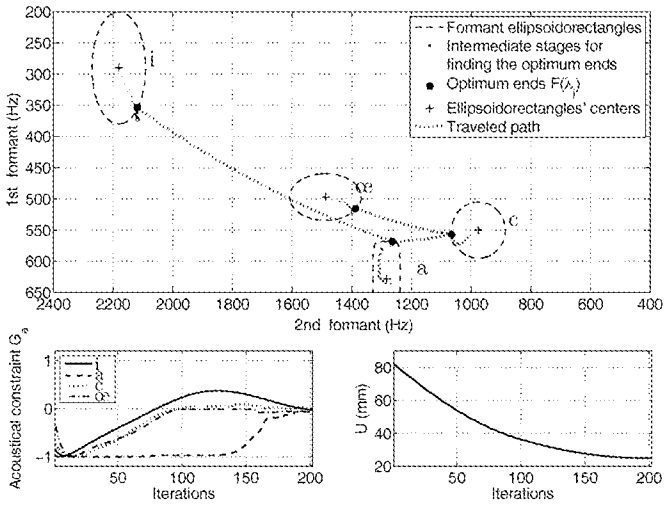


Fig. 11. Optimization of the sequence [i a ɔ œ]. Formant zone of [a] is defined by an ellipsoidorectangle of $\eta_2 = 3$.

optimal formant set (in hertz)

$$\begin{aligned}
\widetilde{\boldsymbol{F}}_1 &= (357.4, 2117.2), \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (571.1, 1296.0), \qquad \text{for [a]} \\
\widetilde{\boldsymbol{F}}_3 &= (528.6, 1422.1), \qquad \text{for [œ]} \\
\widetilde{\boldsymbol{F}}_4 &= (551.0, 1062.8), \qquad \text{for [ɔ].}
\end{aligned} \tag{30}$$

For the latter, the optimization algorithm returns $\widetilde{\boldsymbol{\mu}}_{\mathrm{a}} = (0.0658, 0.1980, 0.1419, 0.061)$ mm, $\widetilde{U} = 25.010$ mm, and the optimal formant set (in hertz)

$$\begin{aligned}
\widetilde{\boldsymbol{F}}_1 &= (354.0, 2119.9), \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (567.6, 1285.1), \qquad \text{for [a]} \\
\widetilde{\boldsymbol{F}}_3 &= (556.9, 1066.5), \qquad \text{for [ɔ]} \\
\widetilde{\boldsymbol{F}}_4 &= (515.8, 1388.0), \qquad \text{for [œ].}
\end{aligned} \tag{31}$$

Thus, we can note a light anticipation on each vowel; however, by taking into account the precision, we mainly observe the anticipation on the vowel [œ]; the formant difference between the two cases is $\Delta \boldsymbol{F} = (12.76, 34.12)$ Hz, which represents 2.4% on each formant, or 3.4% of total difference.[14]

We will now show that according to our model, the chosen strategy can be the reason for greater or smaller anticipation. Once again, we will change the strategy for the vowel [a], by defining its formant zone by an ellipsoidorectangle of $\eta_2 = 3$, and we will compare two previous sequences under these conditions. The results are presented in Figs. 10 and 11. For the sequence [i a œ ɔ], our optimization algorithm returned $\widetilde{\boldsymbol{\mu}}_{\mathrm{a}} = (0.0674, 0.052, 0.123, 0.078)$ mm, $\widetilde{U} = 23.885$ mm, and the optimal formant set (in hertz)

$$\begin{aligned}
\widetilde{\boldsymbol{F}}_1 &= (358.0, 2122.9), \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (572.0, 1320.3), \qquad \text{for [a]} \\
\widetilde{\boldsymbol{F}}_3 &= (529.7, 1431.6), \qquad \text{for [œ]} \\
\widetilde{\boldsymbol{F}}_4 &= (551.8, 1063.5), \qquad \text{for [ɔ].}
\end{aligned} \tag{32}$$

For the sequence [i a ɔ œ], it gave $\widetilde{\boldsymbol{\mu}}_{\mathrm{a}} = (0.065, 0.080, 0.1421, 0.061)$ mm, $\widetilde{U} = 25.028$ mm, and the optimal formant set (in hertz)

$$\begin{aligned}
\widetilde{\boldsymbol{F}}_1 &= (353.3, 2118.9), \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (568.3, 1263.6), \qquad \text{for [a]} \\
\widetilde{\boldsymbol{F}}_3 &= (557.0, 1065.2), \qquad \text{for [ɔ]} \\
\widetilde{\boldsymbol{F}}_4 &= (515.4, 1387.4), \qquad \text{for [œ].}
\end{aligned} \tag{33}$$

We obtain the anticipation on [a] for its second formant to be 56.7 Hz, which represents 4.3% of its initial value. Note that the previous vowel [i] also resulted to be concerned by this modification of strategy for [a]; it decreased its first formant by 4.7 Hz (1.3%). Thus, our model confirms that the degree of anticipation may depend on the chosen strategy. Moreover, since the anticipation on [a] was very small (practically null) in the case of the elliptic formant strategy, one can suppose that actually, the anticipation itself may be one of the consequences of the chosen strategy. Thus, we think that the acoustical anticipation, or coarticulation, is due not to the muscular mechanics and body dynamics, as was claimed in several previous studies [32], [72], but due probably to the centrally planned mechanisms.[15]

Finally, it would be also interesting to compare the obtained results to the real ones obtained by phoneticians. Unfortunately, quantitative analysis of such a kind seems quite difficult and inconsistent for several reasons. First, the proposed solution is only a first-approximation solution. Second, the BTM is not ideal, and neither is the ANN. Third, it is very difficult to compare our results with those found in phonetic literature, because

[14]The latter is $\%_{\mathrm{td}} = 100\sqrt{(\Delta F_1/F_1)^2 + \cdots + (\Delta F_k/F_k)^2}$.

[15]Moreover, these works reported that not only context-sensitivity arises from biomechanics, but it also need not be represented in control CNS commands, while we obtained this effect precisely from CNS-planning mechanism.

the measurement techniques are very different, and even basic acousticophonetic data strongly differ (e.g., compare [107] with [112], although both studies are on American English vowels). Moreover, in most of the coarticulation works, the anticipation extent is measured not by frequency deviation (as we did) but by human-perception error rate, i.e., by means of the so-called discriminant analysis (percentage of correct versus incorrect identification by listeners) [82], [107]. This analysis in based on the fact that nearly all errors involved confusions between adjacent vowels [107], [112]. However, such an error rate is difficult to interpret in absolute formant values that we reported. Another problem is that the discriminant analysis is quite subjective (it is produced by some speakers, and perceived by some listeners), and therefore, it is individual-dependent. Fourth, in our model, there are some individual-dependent parameters, η_j and $\epsilon_{i,j}$, which are meant to represent concrete speaker; in other words, the set of these parameters is meant to represent age, sex, accent, effort, concentration, weariness, etc. However, of course, their concrete values are difficult to estimate numerically. This is why we think that it is better to compare the obtained results qualitatively, rather than quantitatively. If, nevertheless, we accept to compare our results quantitatively with those that were provided by a small amount of works where the coarticulation frequency deviations were reported, e.g., [80] and [108], we find out that they are of the same order (for example, anticipation on [ɑ] is up to 10 Hz for F_1 and up to 50 Hz for F_2, depending on the context). However, more interesting is the overall analysis of these results. It, in particular, shows that the anticipation extent depends on the size of formant zones, as well as on the position of the vowel inside formant space. Generally, vowels having larger formant zones and greater opening angles to the neighboring vowels possess greater anticipation, e.g., cases of [æ], [ɛ], [œ], [ɪ], [ʊ], [u], but not [i] having small opening angle or [ɑ] and [a] having small opening angles and zones (see e.g., [80], [107], and [112]). We obtained similar results: quite small anticipation on [i], [a] (case of small formant zone, $\eta = 1$) or [ɔ] (which was limited to only two close neighbors), and greater anticipation on [œ] and on [a] (case of larger formant zone, $\eta = 3$).

3) Few Words About Two Kinds of the Acoustical Constraints From Practical Point of View and Their Roles in the Previous Optimizations: One can also note that, although we gave the solution for the constrained problem with the acoustical constraints of the second kind (see Section II-C), in fact, it will be also that for the problem with weaker constraints, those of the first kind, if we suppose the local monotonicity inside the ellipsoidorectangles[16] of the dependencies $\boldsymbol{F}(\boldsymbol{\lambda})$. The latter can be generally done, because as we previously saw (see Fig. 4; Section II-B), these relationships, depending on the interval, can be locally considered monotonic. Furthermore, not only it gives the solution for the problem with constraints of the first kind, but also, it is better to use the constraints of the second kind to find a solution to the problem with the constraints of the first kind. In fact, as we may recall from previous figures, during

[16]See also little discussion devoted to the initial point choice in Section III-A.1.

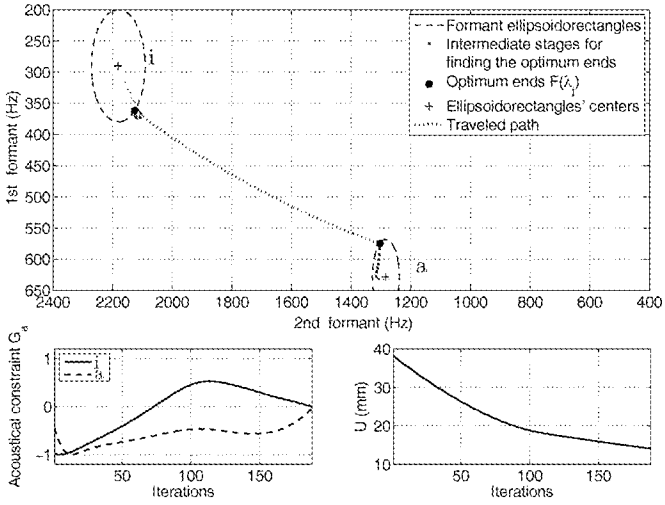


Fig. 12. Correct optimization of the sequence [i a] by using the constraints of the second kind; those of the first kind are, therefore, also satisfied and the minimum of U is reached.

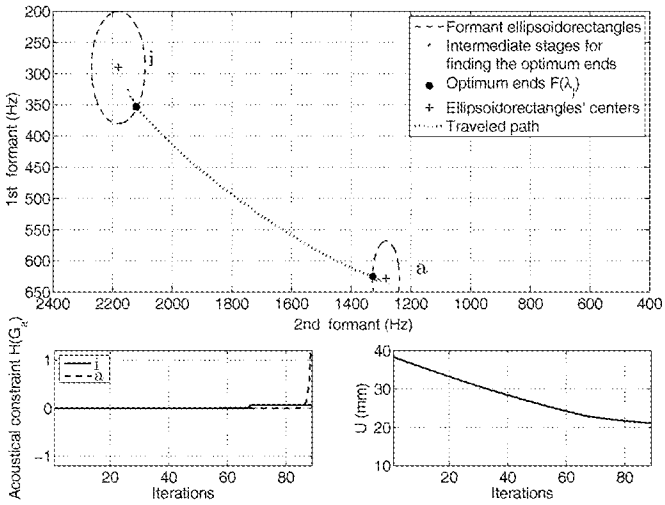


Fig. 13. Incorrect optimization of the sequence [i a] by using the constraints of the first kind; those of the second kind are also satisfied, but the minimum of U is not reached.

the optimization process (different iterations), the solution may temporarily leave the authorized formant zone. This is not so important for the constraints of the second kind but is very important for those of the first kind, because the Heaviside function will not permit leaving even temporarily, the authorized zone, by wrongfully stopping the gradient algorithm because of its sharp increase (this mostly remains true, even if we approximate the Heaviside function by its continuous variants; see the formula given later). We present a comparison of such cases in Figs. 12 and 13. Note that the acoustical constraints of the first and second kinds are fulfilled in both cases (the formant projections of the optimal solutions are on ellipse's borders); however, the correct solution is given only by the model with the constraints of the second kind, i.e., we have $\widetilde{U} = 14.084$ mm, in the case of the constraints of the second kind, while in the case of the constraints of the first kind, $\widetilde{U} = 21.105$ mm, which is the wrong solution, because the length is not minimum. Thus, the

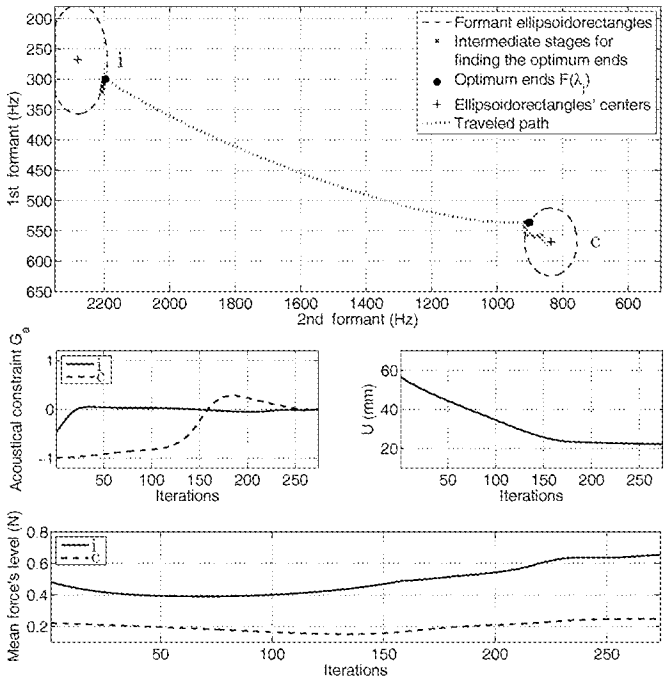


Fig. 14. Optimization of the sequence [i ɔ] with corresponding mean force's level. Formant zones are defined by the corresponding formant ellipses. Only acoustical constraints are applied.

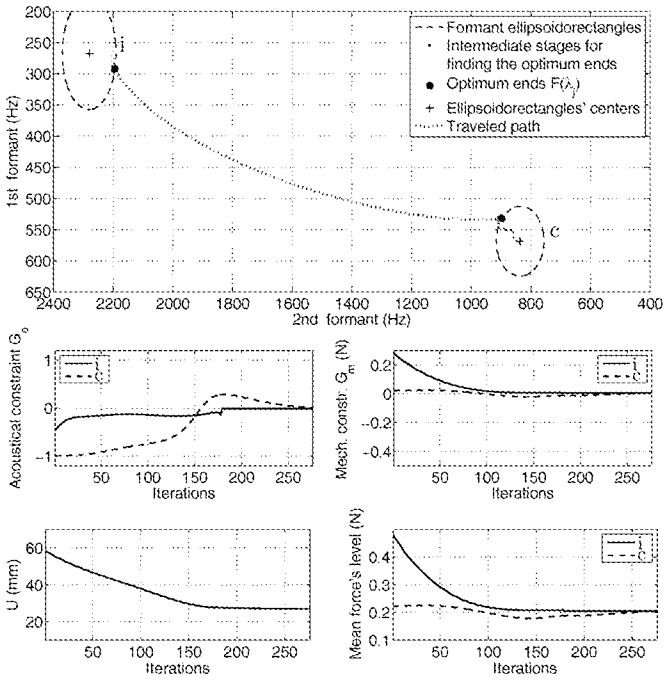


Fig. 15. Optimization of the sequence [i ɔ] with mechanical and acoustical constraints. All formant zones are defined by ellipses. Case of equal mean force's levels (low levels).

gradient algorithm was unable to find the correct solution by using directly the constraints of the first kind; therefore, it is better to circumvent these difficulties by using the constraints of the second kind to find the optimal solution for the constraints of the first kind. Finally, in order to improve the quality of convergence of the optimization algorithm, the Heaviside function was replaced by its continuous approximation $H(z) \approx (1 + \text{th}\, az)/2$, with parameter a taken in the range 60–1000.

B. *Both Acoustical and Mechanical Constraints Are Applied*

The main drawback of our model without mechanical constraints is that the optimal solution gives the formants that are always on the borders of their ellipsoidorectangles, for both kinds of the acoustical constraints. Moreover, for many vowels, only some parts of these borders can be covered (e.g., for the vowel [a], only the upper part of its formant zone can be covered), and the covered zone depends essentially on the position of vowel inside the vocalic triangle. This is actually the consequence of the chosen strategy of planning (i.e., the minimization of the traveled path under constraints) and of the fact that the dependency $\boldsymbol{F}(\boldsymbol{\lambda})$ is often near-monotonic. Therefore, one can suppose that the addition of the mechanical constraints may permit reaching some uncovered zones.

First of all, in all previous simulations unconstrained mechanically, the mean force's level, which corresponds to the optimal ends of the traveled path, was different. Now, we run another simulation unconstrained mechanically, i.e., sequence [i ɔ].[17] The optimization results are shown in Fig. 14; the numerical ones are given as $\widetilde{U} = 22.23$ mm

$$\begin{aligned}
\widetilde{\boldsymbol{\mu}}_{\text{a}} &= (0.0647, 0.1124)\,\text{mm} \\
\widetilde{\boldsymbol{F}}_1 &= (299.7, 2196.3)\,\text{Hz}, \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (535.9,\ \ 901.0)\,\text{Hz}, \qquad \text{for [ɔ]} \\
\widetilde{\mathscr{F}}_1 &= 0.65\,\text{N}, \qquad \text{for [i]} \\
\widetilde{\mathscr{F}}_2 &= 0.25\,\text{N}, \qquad \text{for [ɔ]}
\end{aligned} \tag{34}$$

where $\widetilde{\mathscr{F}}_j \equiv \mathscr{F}(\widetilde{\boldsymbol{\lambda}}_j)$, $j = 1, \ldots, p$; in other words, j corresponds to the sound in the sequence. We will now try to prescribe the mean force's level with both kinds of the acoustical constraints and to prescribe the equal global mean force's level to each vowel in the sequence.

In the first example, we try to force both vowels to decrease their mean force's level by reaching the value 0.20 N. The results of this optimization are given as $\widetilde{U} = 26.90$ mm

$$\begin{aligned}
\widetilde{\boldsymbol{\mu}}_{\text{a}} &= (0.36, 0.13)\,\text{mm} \\
\widetilde{\boldsymbol{\mu}}_{\text{m}} &= (6.8, -1.0)\,\text{mm/N} \\
\widetilde{\boldsymbol{F}}_1 &= (292.5, 2194.4)\,\text{Hz}, \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (531.8,\ \ 897.0)\,\text{Hz}, \qquad \text{for [ɔ]} \\
\widetilde{\mathscr{F}}_1 &= 0.20\,\text{N}, \qquad \text{for [i]} \\
\widetilde{\mathscr{F}}_2 &= 0.20\,\text{N}, \qquad \text{for [ɔ].}
\end{aligned}$$

Fig. 15 shows the optimization. Mainly, we note that [ɔ] did not significantly change its position on the formant zone's border, while [i] was slightly influenced. It can be understood: the

[17] We recall that in this section, [i] and [ɔ] are those with ∗ from Table I.

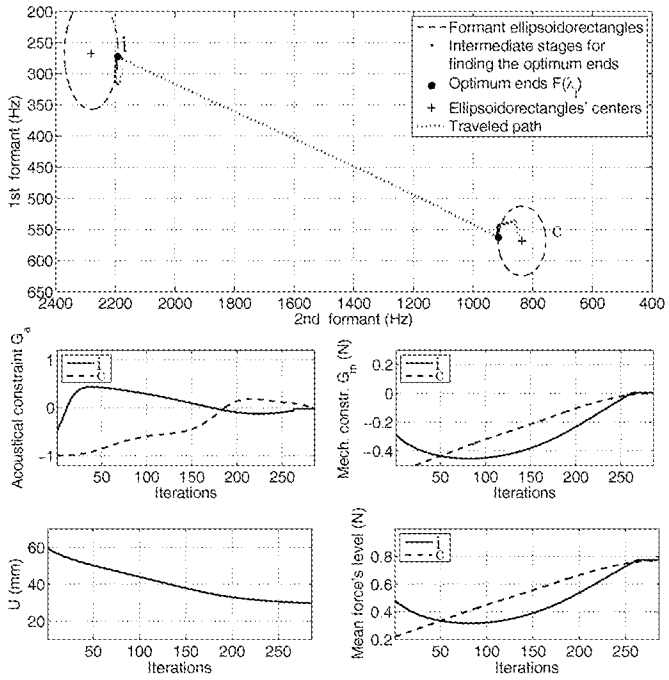


Fig. 16. Optimization of the sequence [i ɔ] with mechanical and acoustical constraints. All formant zones are defined by ellipses. Case of equal mean force's levels (high levels).

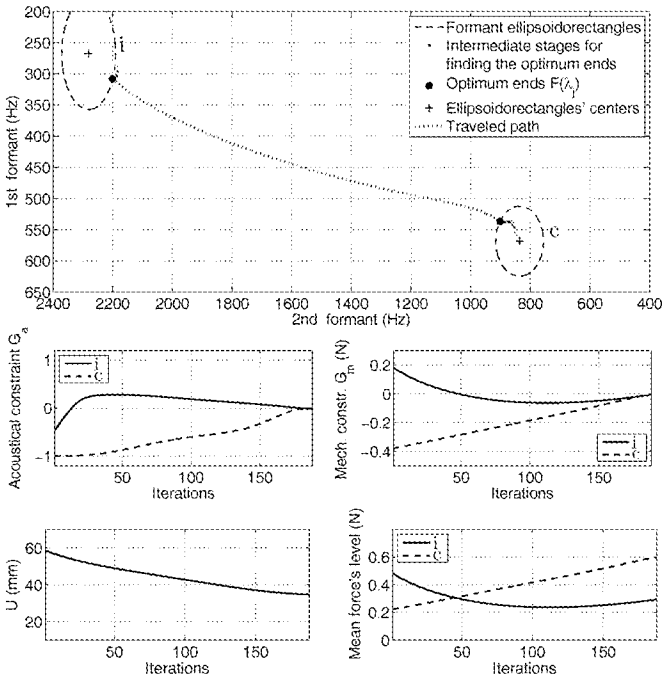


Fig. 17. Optimization of the sequence [i ɔ] with mechanical and acoustical constraints. All formant zones are defined by ellipses. Case of different mean force's levels (low and high levels).

prescribed mean force's level is much closer to [ɔ], which is why it impacts especially on [i], whose unconstrained level was high. On the other hand, the modification of its position was not very great, notwithstanding the quite different muscular commands. This is normal because, as we previously said, there is no bijection between the internal and external spaces; thus, a whole domain in the internal space may correspond to one point in the external space. In other words, the most significant changes were produced in this domain of the internal space; more precisely, for [i], the total difference in the internal space $\boldsymbol{\lambda}$ is $\%_{\mathrm{td}} = 32.8\%$, while that in the external space $\boldsymbol{F}$ is only $\%_{\mathrm{td}} = 2.4\%$! This is also related to the fact that the length of path traveled in the internal space increased: 26.90 mm versus 22.23 mm.

In the second example, we will try to obtain a high level of the prescribed mean force's level, say $\mathring{\mathscr{F}} = 0.77$ N, for all vowels in the same sequence. The optimization returns the following: $\widetilde{U} = 29.65$ mm

$$\begin{aligned}
\widetilde{\boldsymbol{\mu}}_{\mathrm{a}} &= (0.077, 0.191)\,\mathrm{mm} \\
\widetilde{\boldsymbol{\mu}}_{\mathrm{m}} &= (0.89, -5.50)\,\mathrm{mm/N} \\
\widetilde{\boldsymbol{F}}_1 &= (272.6, 2191.4)\,\mathrm{Hz}, \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (562.5,\ \ 915.4)\,\mathrm{Hz}, \qquad \text{for [ɔ]} \\
\widetilde{\mathscr{F}}_1 &= 0.77\,\mathrm{N}, \qquad \text{for [i]} \\
\widetilde{\mathscr{F}}_2 &= 0.77\,\mathrm{N}, \qquad \text{for [ɔ].}
\end{aligned}$$

Fig. 16 shows the optimization. In this case, both vowels were influenced: formants of [i] by $\%_{\mathrm{td}} = 9.1\%$ and those of [ɔ] by $\%_{\mathrm{td}} = 5.2\%$; the corresponding values in the internal space $\boldsymbol{\lambda}$ are much greater: $\%_{\mathrm{td}} = 19.9\%$ and $\%_{\mathrm{td}} = 43.7\%$, respectively.

Two previous simulations showed that the model allows the prescription of equal mean force's levels to each vowel in sequence. However, this case is not very realistic, and the prescription of different mean force's levels to the vowels would better represent the reality. In the next example, for the sequence [i ɔ], we will try to prescribe law mean force's level to the first vowel (which was high in the simulation unconstrained mechanically), which is given by $\mathring{\mathscr{F}}_{\mathrm{i}} = 0.3$ N and, conversely, high level to the second one, which is given by $\mathring{\mathscr{F}}_{\mathrm{ɔ}} = 0.6$ N. This simulation, which is shown in Fig. 17, returned the following optimization results: $\widetilde{U} = 34.70$ mm

$$\begin{aligned}
\widetilde{\boldsymbol{\mu}}_{\mathrm{a}} &= (0.139, 0.211)\,\mathrm{mm} \\
\widetilde{\boldsymbol{\mu}}_{\mathrm{m}} &= (2.78, -4.17)\,\mathrm{mm/N} \\
\widetilde{\boldsymbol{F}}_1 &= (308.1, 2200.6)\,\mathrm{Hz}, \qquad \text{for [i]} \\
\widetilde{\boldsymbol{F}}_2 &= (536.8,\ \ 901.7)\,\mathrm{Hz}, \qquad \text{for [ɔ]} \\
\widetilde{\mathscr{F}}_1 &= 0.30\,\mathrm{N}, \qquad \text{for [i]} \\
\widetilde{\mathscr{F}}_2 &= 0.60\,\mathrm{N}, \qquad \text{for [ɔ].}
\end{aligned}$$

Both vowels were influenced, as well as the character of the projection of the traveled path. As to the formant deviation, those of [i] changed by $\%_{\mathrm{td}} = 2.8\%$, and those of [ɔ] changed by $\%_{\mathrm{td}} = 1.9\%$; the corresponding values in the internal space $\boldsymbol{\lambda}$ are much larger: $\%_{\mathrm{td}} = 24.7\%$, and $\%_{\mathrm{td}} = 32.4\%$ respectively, with respect to the simulation unconstrained mechanically (see (34) and Fig. 14).

These results show that regardless of the mean force's level, it is possible to keep the vowels in their formant zones (on the border), and thus, there was no impact on the fulfillment of the acoustical constraints; mathematically, the problem could always be solved. The prescribed mean force's level has variable impact on the external formant space $\boldsymbol{F}$ by letting sometimes to accede to the formant zones uncovered before. On the contrary, the prescribed mean force's level strongly impacts on the internal command space $\boldsymbol{\lambda}$, which is always influenced in great degree; in addition, the distance between the optimum ends always increases, which suggests that change of the external tasks may be often compensated by the CNS.

IV. Conclusion

We presented an optimum neural-network-based internal model for the control of the speech robots, controlled by EPH. The internal model was designed according to the principle of the shortest path in the internal command space, under acoustical and mechanical constraints, which were meant to represent static acoustic and dynamic mechanic targets.

We first dealt with the obtaining of the exact analytical solution, which was based on the calculus of variations. We proved that this solution cannot be, in general, a linear function that casts doubts on the so-called linearized λ-model, which has received some interest in artificial-intelligence device modeling and supposes that the time transitions in the internal space may be only linear. Moreover, it was also shown that the linearized λ-model may have some potential instability issues because of infinitely growing energy and power and that it may not be fully compatible with dynamic-specification target theory. We, therefore, suggest to reconsider the linearized λ-model.

Then, by using some empirical findings, we developed a first approximation solution for the proposed optimum internal model. Experimental tests showed that this model was in accordance with the phenomenon of the acousticophonetic anticipation (i.e., coarticulation), and it also showed that the degree of the latter is closely related to the chosen task-planning strategy. Moreover, it was also suggested that the anticipation itself may be due to the internal CNS strategy and not to the dynamics or biomechanics of physical system (such as the human body), as was previously suggested in several works [32], [72].

As to the influence of the mean muscular tongue effort, which helps to account for lax and tense vowels, its strict prescription via mechanical constraints did not lead to the problem overposed mathematically, and we always succeeded to obtain optimum solutions. Thus, the model did not permit to answer to questions related to the relationships between the forces and the hypo/hyper-speech (see, e.g., [113]), namely, we could not affirm that a greater or smaller mean force's level leads to the restriction or enlargement of the formant zones. We could reach the small levels, as well as the great ones. The strict prescription of the mean force's level has a variable impact on the external formant space. In contrast, it strongly impacts on the internal command space; the length of traveled path increased, and the motor commands become quite different, even if the impact on the formant space was small. In other words, the robot, which was controlled by the proposed internal model, was able to compensate the change of one of the external tasks, without significantly changing the quality of the another one, by the corresponding shift in the internal space, which was determined by the proposed optimum algorithm.

Appendix

ANN: Construction, Learning, and Validation

To construct the approximate model, we employed the two-layer ANN using radial basis neurons for the first layer (called also hidden or radial layer) and linear neurons for the second one (which is also called output layer) [5], [87]–[91]. The chosen radial basis transfer function is the Gaussian curve $e^{-(\cdot)^2}$. The choice of the radial basis networks is motivated by the fact that they are usually considered as one of the best for most nonlinear approximations, thus giving a good tradeoff between the complexity and the precision of the network.

By basing on the architecture of such a network [91], the formalization of the input–output relationships is not complicated. The output of the sth radial layer neuron a_s is given by

$$a_s = e^{-(\|\boldsymbol{w}_s - \boldsymbol{\lambda}\|b_1)^2}, \qquad s = 1, \ldots, S_1 \tag{35}$$

where $\boldsymbol{w}_s \equiv (w_{1,s}, \ldots, w_{n,s})$ is the input weight vector of the sth radial neuron, $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_n)$ is the input training vector, b_1 is the narrowness' parameter for the Gaussian function of the sth hidden neuron, S_1 is the number of the hidden layer neurons, and $\|\cdot\|$ denotes the Euclidian distance. The lth output of the linear layer α_l can be written as

$$\alpha_l = \boldsymbol{v}_l \boldsymbol{a} + b_{2,l} = \sum_{s=1}^{S_1} v_{s,l} e^{-(\|\boldsymbol{w}_s - \boldsymbol{\lambda}\|b_1)^2} + b_{2,l} \tag{36}$$

for $l = 1, \ldots, S_2$, where $\boldsymbol{a}$ is the column vector $(a_1, \ldots, a_{S_1})^t$, $\boldsymbol{v}_l$ is the output weight line vector $(v_{1,l}, \ldots, v_{S_1,l})$, S_2 is the number of the output linear neurons, and $b_{2,l}$ is the bias of the lth linear neuron (where "^{t}" stands for transpose). The learning procedure consists of determining the weight vectors $\boldsymbol{w}_s$, $\boldsymbol{v}_l$, and the bias vector $\boldsymbol{b}_2 \equiv (b_{2,1}, \ldots, b_{2,S_2})$ by minimizing the sum of output squared errors (SSEs) on the set of input–output data, where parameters b_1 are often left for the users.

Practical implementation of the ANNs was carried out with neural network toolbox of MATLAB [91], by means of `newrb` function. To learn the network, we first generated 17 000 random motor commands $\boldsymbol{\lambda}$ distributed uniformly, occupying a six-orthotope in the internal space $\boldsymbol{\lambda}$. Then, these motor commands are applied to the BTM one after another (see Fig. 3). This produces 17 000 output formant vectors $\boldsymbol{F}$ (see Fig. 18) and output global mean force's levels $\mathscr{F}$. For $\boldsymbol{F}$, only first two formants are taken into account, i.e., $k = 2$ (first two formants are, in general, sufficient to distinguish the vowel [86], [107]).[18] Then, 17 000 data were split in two parts: 6000 data were given to the ANNs in order to choose 340 optimal ones for the construction of the latter (hidden layer of the ANN is, therefore,

[18] Despite that the values of n and k were fixed in experiments, all the formulas will be written for arbitrary values of n and k, in order to keep the generality of the model.

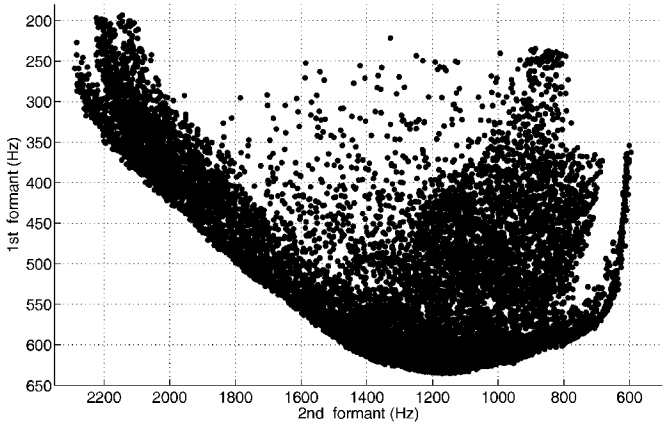


Fig. 18. Distribution of the 17 000 output vectors $\boldsymbol{F}$, i.e., first two formants F_1 and F_2.

composed of 340 radial layer neurons, i.e., $S_1 = 340$; for more details, see [91]), and the remaining 11 000 were used to test the accuracy of the obtained ANNs, which is about 1% on each output.

REFERENCES

[1] M. L. Latash, *Neurophysiological Basis of Movement*. Champaign, IL: Hum. Kinet., 1998.

[2] M. L. Latash and F. Lestienne, *Motor Control and Learning*. New York: Springer Sci.–Bus. Media, 2006.

[3] R. A. Schmidt and T. D. Lee, *Motor Control and Learning*, 4th ed. Champaign, IL: Hum. Kinet., 2005.

[4] P. Cordo and S. R. Harnad, *Movement Control*. Cambridge, U.K: Cambridge Univ., 1994.

[5] M. A. Arbib, *The Handbook of Brain Theory and Neural Networks*, 2nd ed. Cambridge, MA: MIT, 2003.

[6] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.

[7] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses," *J. Exp. Psychol.*, vol. 67, no. 2, pp. 103–112, 1964.

[8] P. M. Fitts and B. K. Radford, "Information capacity of discrete motor responses under different cognitive sets," *J. Exp. Psychol.*, vol. 71, pp. 475–482, 1966.

[9] N. Hogan, "An organizing principle for a class of voluntary movements," *J. Neurosci.*, vol. 4, no. 11, pp. 2745–2764, 1984.

[10] N. Hogan, "Moving gracefully: Quantitative theories of motor coordination," *Trends Neurosci.*, vol. 10, pp. 170–174, 1985.

[11] W. L. Nelson, "Physical principles for economies of skilled movements," *Biol. Cybern.*, vol. 46, pp. 135–147, 1983.

[12] M. Dornay, M. K. Y. Uno, and R. Suzuki, "Minimum muscle–tension change trajectories predicted by using a 17-muscle model of the monkey's arm," *J. Motor Behav.*, vol. 28, no. 2, pp. 83–100, 1996.

[13] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement," *Biol. Cybern*, vol. 61, no. 2, pp. 89–101, 1989.

[14] H. Hatze and J. D. Buys, "Energy-optimal controls in the mammalian neuromuscular system," *Biol. Cybern.*, vol. 27, pp. 9–20, 1977.

[15] P. Zukofsky, "Arm movements in skilled violin playing," presented at the 22nd Annu. Meeting Psychon. Soc., Philadelphia, PA, 1981.

[16] D. G. Asatryan and A. G. Feldman, "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. I: Mechanographic analysis of the work of the joint or execution of a postural task," *Biophys. J.*, vol. 10, pp. 925–935, 1965.

[17] A. G. Feldman, "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. II: Controllable parameters of the muscles," *Biophys. J.*, vol. 11, pp. 565–578, 1966.

[18] A. G. Feldman, "Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III: Mechanographic analysis of execution by man of the simplest of motor tasks," *Biophys. J.*, vol. 11, pp. 766–775, 1966.

[19] A. G. Feldman and G. N. Orlovsky, "The influence of different descending systems on the tonic stretch reflex in the cat," *Exp. Neurol.*, vol. 37, pp. 481–494, 1972.

[20] A. G. Feldman, "Once more on the equilibrium point hypothesis (λ model) for motor control," *J. Motor Behav.*, vol. 18, pp. 17–54, 1986.

[21] A. G. Feldman, "Change of muscle length as a consequence of a shift in an equilibrium of muscle load system," *Biophys.*, vol. 19, pp. 544–548, 1974.

[22] A. G. Feldman, "Control of the length of a muscle," *Biophys.*, vol. 19, pp. 766–771, 1974.

[23] E. Bizzi, N. Hogan, F. A. Mussa-Ivaldi, and S. F. Giszter, "Does the nervous system use equilibrium-point control to guide single and multiple joint movements?" *Behav. Brain Sci.*, vol. 15, pp. 603–613, 1992.

[24] X. Gu and D. H. Ballard, "An equilibrium point based model unifying movement control in humanoids," in *Proc. Robot. Sci. Syst.*, 2006, pp. 1–7.

[25] E. Bizzi, N. Hogan, F. A. Mussa-Ivaldi, and S. F. Giszter, "The equilibrium-point framework: A point of departure," *Behav. Brain Sci.*, vol. 15, pp. 808–815, 1992.

[26] X. Gu and D. H. Ballard, "Robot movement planning and control based on equilibrium point hypothesis," in *Proc. IEEE Conf. Robot., Autom. Mechatron.*, 2006, pp. 1–6.

[27] S. A. Migliore and S. DeWeerth, "Control of robotic joints using principles from the equilibrium point hypothesis of animal motor control," M.S. thesis, Georgia Inst. Technol., Atlanta, GA, 2004.

[28] M. Kawato, "Internal models for motor control and trajectory planning," *Curr. Opin. Neurobiol.*, vol. 9, pp. 718–727, 1999.

[29] J. R. Flanagan and A. M. Wing, "The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads," *J. Neurosci.*, vol. 17, pp. 1519–1528, 1997.

[30] M. Kawato, M. Isobe, Y. Maeda, and R. Suzuki, "Coordinates transformation and learning control for visually-guided voluntary movement with iteration: A Newton-like method in a function space," *Biol. Cybern.*, vol. 59, no. 3, pp. 161–177, 1988.

[31] J. D. Cooke, "The organization of simple skilled movements," in *Advances in Psychology: Tutorials in Motor Behavior*, G. E. Stelmach and J. Requin, Eds. Amsterdam, The Netherlands: North-Holland, 1980.

[32] P. L. Gribble, D. J. Ostry, V. Sanguineti, and R. Laboissière, "Are complex control signals required for human arm movement?" *J. Neurophysiol.*, vol. 79, no. 3, pp. 1409–1424, 1998.

[33] V. Sanguineti, R. Laboissière, and D. J. Ostry, "A dynamic biomechanical model for neural control of speech production," *J. Acoust. Soc. Amer.*, vol. 103, no. 3, pp. 1615–1627, 1998.

[34] J. R. Flanagan, D. J. Ostry, and A. G. Feldman, "Control of human jaw and multi-joint arm movements," in *Cerebral Control of Speech and Limb Movements*, G. Hammond, Ed. New York: Springer-Verlag, 1990, pp. 29–58.

[35] D. J. Ostry, J. R. Flanagan, A. G. Feldman, and K. G. Munhall, "Human jaw movement kinematics and control," in *Tutorials in Motor Behavior II*, G. E. Stelmac and J. Requin, Eds. Amsterdam, The Netherlands: Elsevier, 1992, pp. 646–660.

[36] "Special issue on the equilibrium point hypothesis and its applications in speech," *Bull. Commun. Parlée*, vol. 4, pp. 5–110, 1998.

[37] R. Laboissiére, D. J. Ostry, and A. G. Feldman, "The control of multimuscle systems: Human jaw and hyoid movements," *Biol. Cybern.*, vol. 74, no. 3, pp. 373–384, 1996.

[38] R. Wilhelms-Tricarico, "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," *J. Acoust. Soc. Amer.*, vol. 97, no. 5, pp. 3085–3098, 1995.

[39] R. Wilhelms-Tricarico and C.-M. Wu, "A biomechanical model of the tongue," in *Proc. Bioeng. Conf.*, vol. 35, K. B. Chandran, R. Vanderby, Jr., and M. S. Hefzy, Eds. New York: ASME, 1997, pp. 69–70.

[40] P. Buscemi, M. Carlson, and R. W. Tricarico, "A computational approach to muscle modeling of the human tongue via the finite element method along with motion control correlations with MRI tracking data for simple speech patterns," *J. Med. Devices*, vol. 2, no. 2, p. 027548, 2008.

[41] H. Ranca, C. Servaisa, P.-F. Chauvyb, S. Debaudb, and S. Mischle, "Effect of surface structure on frictional behaviour of a tongue/palate tribological system," *Tribol. Int.*, vol. 39, no. 12, pp. 1518–1526, 2006.

[42] W. S. Levine, C. E. Torcaso, and M. Stone, "Controlling the shape of a muscular hydrostat: A tongue or tentacle," *Lect. Notes Control Inf. Sci.*, vol. 321, pp. 20–222, 2005.

[43] W. H. Lewis, Ed., *Gray's Anatomy of the Human Body, 20th U.S. ed.* Philadelphia, PA: Lea & Febiger, 1918.

[44] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method. Basic Formulation and Linear Problems*. New York: McGraw-Hill, 1989.

[45] P. E. Rubin, T. Baer, and P. Mermelstein, "An articulatory synthesizer for perceptual research," *J. Acoust. Soc. Amer.*, vol. 70, pp. 321–328, 1981.

[46] P. E. Rubin, E. Saltzman, L. Goldstein, R. McGowan, M. Tiede, and C. Browman, "CASY and extensions to the task-dynamic model," in *Proc. 1st ESCA Tutorial Res. Workshop Speech Producing Model., 4th Speech Prod. Semin.*, 1996, pp. 125–128.

[47] J. S. Perkell and K. N. Stevens, "A physiologically-oriented model of tongue activity in speech production," Ph.D. dissertation, Speech Commun. Group, Dept. Electr. Eng., Mass. Inst. Technol., Cambridge, MA, 1974.

[48] S. Kiritani, K. Miyawaki, O. Fujimura, and J. E. Miller, "Computational model of the tongue," *J. Acoust. Soc. Amer.*, vol. 57, no. S1, pp. S3–S3, 1975.

[49] Y. Kakita and O. Fujimura, "Computational model of the tongue: A revised version," *J. Acoust. Soc. Amer.*, vol. 62, no. S1, pp. S15–S16, 1977.

[50] M. M. Sondhi and J. Schroeter, "A nonlinear articulatory speech synthesizer using both time- and frequency-domain elements," in *Proc. ICASSP*, Tokyo, Japan, 1986, pp. 1999–2002.

[51] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-35, no. 7, pp. 955–967, Jul. 1986.

[52] S. Maeda, "An articulatory model of the tongue based on a statistical analysis," *J. Acoust. Soc. Amer.*, vol. 65, no. s1, p. S22, 1988.

[53] S. Maeda, "Improved articulatory model," *J. Acoust. Soc. Amer.*, vol. 84, no. s1, p. S146, 1988.

[54] F. Vogt, "Finite element modeling of the tongue," in *Proc. Int. Workshop Audio Vis. Speech Process.*, 2005, pp. 143–144.

[55] O. Engwall, "A 3D tongue model based on MRI data," in *Proc. 6th Int. Conf. Spoken Lang. Process.*, Beijing, China, 2000, pp. 901–904.

[56] K. van den Doel, F. Vogt, R. E. English, and S. Fels, "Towards articulatory speech synthesis with a dynamic 3D finite element tongue model," in *Proc. ISSP*, 2006, pp. 59–66.

[57] [Online]. Available: http://www.takanishi.mech.waseda.ac.jp/top/research/voice/index.htm

[58] K. Nishikawa, K. Asama, K. Hayashi, H. Takanobu, and A. Takanishi, "Development of a talking robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2000, pp. 1760–1765.

[59] K. Fukui, K. Nishikawa, S. Ikeo, E. Shintaku, K. Takada, H. Takanobu, M. Honda, and A. Takanishi, "Development of a talking robot with vocal cords and lips having human-like biological structures," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 2023–2028.

[60] P. L. M. de Maupertuis, "Accord de différentes lois de la nature qui avaient jusqu'ici paru incompatibles (Eng. trans.: "Accord between different laws of nature that seemed incompatible")," presented at the Mémoires l'Acad. Sci. Paris, Paris, France, 1744, p. 417.

[61] P. L. M. de Maupertuis, "Les lois du mouvement et du repos, déduites d'un principe de métaphysique (Eng. trans.: "Derivation of the laws of motion and equilibrium from a metaphysical principle")," presented at the Mémoires l'Acad. Sci. Paris, Berlin, Germany, 1746, p. 267.

[62] J.-L. Lagrange, *Euvres de Lagrange. Tome Onzième: Mécanique Analytique, Tome Premier, Quatrième éd.* Paris, France: Gauthier-Villars et fils, imprimeurs-libraires, 1888.

[63] W. R. Hamilton, "On a general method in dynamics," *Philos. Trans. R. Soc.*, 1834, pp. 247–308.

[64] W. R. Hamilton, "On a general method in dynamics," *Philos. Trans. R. Soc.*, 1835, pp. 95–144.

[65] H. Goldstein, *Classical Mechanics*. Reading, MA: Addison-Wesley, 1953.

[66] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. Berlin, Germany: Springer-Verlag, 1989.

[67] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics. Vol. I: Mechanics*, 3rd ed. Oxford, U.K.: Elsevier, 2003.

[68] L. N. Hand and J. D. Finch, *Analytical Mechanics*. Cambridge, U.K.: Cambridge Univ., 1998.

[69] C. Lanczos, *The Variational Principles of Mechanics*. Toronto, ON, Canada: Univ. Toronto, 1970.

[70] D. J. Ostry and K. G. Munhall, "Control of jaw orientation and position in mastication and speech," *J. Neurosci.*, vol. 71, pp. 1528–1545, 1994.

[71] P. L. Gribble, R. Laboissiére, and D. J. Ostry, "Control of human arm and jaw motion: issues related to musculo-skeletal geometry," in *Self-Organization, Computational Maps and Motor Control*, vol. 118. Amsterdam, The Netherlands: North-Holland/Elsevier, 1997, pp. 483–506.

[72] D. J. Ostry, P. L. Gribble, and V. L. Gracco, "Coarticulation of jaw movements in speech production: is context sensitivity in speech kinematics centrally planned?" *J. Neurosci.*, vol. 16, pp. 1570–1579, 1996.

[73] A. G. Feldman, S. V. Adamovich, D. J. Ostry, and J. R. Flanagan, "The origin of electromyograms—Explanations based on the equilibrium point hypothesis," in *Multiple Muscle Systems: Biomechanics and Movement Organization*, J. Winters and S. Woo, Eds. Berlin, Germany: Springer-Verlag, 1990, pp. 195–213.

[74] M. R. Schroeder, "Determination of the geometry of the human vocal tract by acoustic measurements," *J. Acoust. Soc. Amer.*, vol. 41, no. 4B, pp. 1002–1010, 1967.

[75] S. Hiroya and M. Honda, "Estimation of articulatory movements from speech acoustics using an hmm-based speech production model," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 2, pp. 175–185, Mar. 2004.

[76] R. Marret, "Apprentissage des relations entre commandes musculaires et géométrie de la langue," Master's thesis, Inst. Nat. Polytech. de Grenoble, Grenoble, France, 2002.

[77] P. Perrier, L. Ma, and Y. Payan, "Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue," in *Proc. 9th Eur. Conf. Speech Commun. Technol.*, 2005, pp. 1041–1044.

[78] F. H. Guenther, M. Hampson, and D. Johnson, "A theoretical investigation of reference frames for the planning of speech movements," *Psychol. Rev.*, vol. 105, pp. 611–633, 1998.

[79] D. Kewley-Port and S. S. Goodman, "Thresholds for second formant transitions in front vowels," *J. Acoust. Soc. Amer.*, vol. 118, no. 5, pp. 3252–3260, 2005.

[80] J. Hillenbrand, M. J. Clark, and T. M. Nearey, "Effects of consonant environment on vowel formant patterns," *J. Acoust. Soc. Amer.*, vol. 109, no. 2, pp. 748–763, 2001.

[81] T. L. Gottfried and W. Strange, "Identification of coarticulated vowels," *J. Acoust. Soc. Amer.*, vol. 68, no. 6, pp. 1626–1635, 1980.

[82] W. Strange, J. J. Jenkins, and T. L. Johnson, "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Amer.*, vol. 74, no. 3, pp. 695–705, 1983.

[83] J. E. Andruski and T. M. Nearey, "On the sufficiency of compound target specification of isolated vowels and vowels in /bvb/ syllables," *J. Acoust. Soc. Amer.*, vol. 91, pp. 390–410, 1992.

[84] H. Dudley and T. H. Tarnoczy, "The calculation of vowel resonances, and an electrical vocal tract," *J. Acoust. Soc. Amer.*, vol. 22, no. 6, pp. 740–753, 1950.

[85] G. Fant, *Acoustic Theory of Speech Production*. Hague, The Netherlands: Mouton, 1960.

[86] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[87] Y. H. Hu and J. N. Hwang, Eds., *Handbook of Neural Network Signal Processing*. Boca Raton, FL: CRC, 2002.

[88] L. Jain, *Recent Advances in Artificial Neural Networks. Design and Applications* (Int. Ser. Comput. Intell.), A. M. Fanelli, Ed. Boca Raton, FL: CRC, 2000.

[89] S. Chen, C. Cowan, and P. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[90] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.

[91] H. Demuth, M. Beale, and M. Hagan, *Neural Network Toolbox for Use With MATLAB (Version 5)*. Natick, MA: The MathWorks, 2006.

[92] M. Kawato, K. Furukawa, and R. Suzuki, "A hierarchical neural-network model for control and learning of voluntary movement," *Biol. Cybern.*, vol. 57, no. 3, pp. 169–185, 1987.

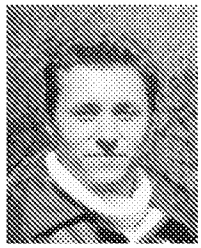
[93] F. H. Guenther, "A neural network model of speech acquisition and motor equivalent speech production," *Biol. Cybern.*, no. 72, pp. 43–53, 1994.

[94] T. D. Sanger, "Neural network learning control of robot manipulators using gradually increasing task difficulty," *IEEE Trans. Robot. Autom.*, vol. 10, no. 3, pp. 323–333, Jun. 1994.

[95] F. H. Guenther, "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," *Psychol. Rev.*, vol. 102, pp. 594–621, 1995.

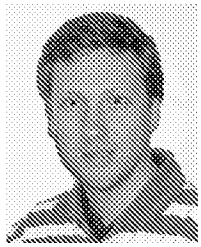
[96] F. H. Guenther and J. W. Bohland, "Learning sound categories: A neural model and supporting experiments. Acoustical science and technology," *Acoust. Sci. Technol.*, vol. 23, no. 4, pp. 213–220, 2002.

[97] M. G. Rahim, C. C. Goodyear, W. B. Kleijn, J. Schroeter, and M. M. Sondhi, "On the use of neural networks in articulatory speech synthesis," *J. Acoust. Soc. Amer.*, vol. 93, no. 2, pp. 1109–1121, 1993.
[98] L. Eulero. (1744). *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici latissimo sensu accepti*. Lausannæ/Genevæ, Switzerland: Apud Marcum-Michaelem Bousquet/Socios [Online]. Available: http://math.dartmouth.edu/~euler/pages/E065.html
[99] M. J. Forray, *Variational Calculus in Science and Engineering*. New York: McGraw-Hill, 1967.
[100] R. Weinstock, *Calculus of Variations With Applications to Physics and Engineering*. New York: McGraw-Hill, 1952.
[101] G. A. Bliss, *Lectures on the Calculus of Variations*. Chicago, IL: Univ. Chicago, 1947.
[102] V. I. Smirnov, *A Course of Higher Mathematics*, vol. I–V, Oxford, U.K.: Pergamon, 1964.
[103] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. I. New York: Interscience, 1966.
[104] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers. Definitions, Theorems, and Formulas for Reference and Review*, 2nd ed. enlarged and revised ed. New York: McGraw-Hill, 1968.
[105] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, 3rd ed. Berlin, Germany: Springer-Verlag, 1998.
[106] Calliope, *La Parole et Son Traitement Automatique*. Paris, France: Dunod, 1989.
[107] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Amer.*, vol. 24, no. 2, pp. 175–184, 1952.
[108] K. N. Stevens and A. S. House, "Perturbation of vowel articulations by consonantal context," *J. Speech Hear. Res.*, vol. 6, pp. 111–128, 1963.
[109] M. J. Macchi, "Identification of vowels spoken in isolation versus vowels spoken in consonantal context," *J. Acoust. Soc. Amer.*, vol. 68, no. 6, pp. 1636–1642, 1980.
[110] T. M. Nearey, "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Amer.*, vol. 85, no. 5, pp. 2088–2113, 1989.
[111] J. Talley, "Vowel perception in varied symmetric CVC contexts," *J. Acoust. Soc. Amer.*, vol. 108, no. 5, pp. 2601–2601, 2000.
[112] J. Hillenbrand, L. A. Getty, K. Wheeler, and M. J. Clark, "Acoustic characteristics of american english vowels," *J. Acoust. Soc. Amer.*, vol. 95, no. 5, pp. 2875–2875, 1995.
[113] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech Production and Speech Modelling*, W.J. Hardcastle and A. Marchal, Eds. Dordrecht, The Netherlands: Kluwer, 1990, pp. 403–439.



Iaroslav V. Blagouchine was born in St. Petersburg, Russia, on December 22, 1979. He received the B.S. degree in physics from the St. Petersburg State University in 2000, the M.S. degree in electronic engineering from the Grenoble Institute of Technology, and the Ph.D. degree in signal processing and applied mathematics from the École Centrale, France, in 2001 and 2009, respectively.

From 2001 to 2002, he was with the Department Señales, Sistemas y Radiocomunicaciones Universidad Politécnica de Madrid, Madrid, Spain. During 2003, he was a Research Engineer with the Grenoble Institute of Technology, Grenoble, France, where he was also a Teacher Assistant from 2004 to 2007. From 2007 to 2009, he was a Postdoctoral Researcher and a Teacher Assistant with the Telecommunication Department, University of Toulon, Toulon, France. Since September 2009, he has been a Research Engineer with the Department of Mobile Communication, Eurécom, Sophia Antipolis, France. His current research interests include biologically inspired robotics (especially equilibrium-point-hypothesis-based), speech robotics, constraint optimization techniques, variational calculus, and statistical signal processing.



Eric Moreau (M'96–SM'08) was born in Lille, France. He graduated from the Ecole Nationale Supérieure des Arts et Métiers, Paris, France, in 1989. He received the Agrégation de Physique degree from the Ecole Normale Supérieure de Cachan, Cachan Cedex, France, in 1990 and the DEA and Ph.D. degrees in signal processing from the Université Paris-Sud in 1991 and 1995, respectively.

From 1995 to 2001, he was an Assistant Professor with the Department of Telecommunications, Institute of Engineering Sciences of Toulon-Var-School of Engineering, University of Toulon, Toulon, France, where he is currently a Professor. His current research interests include constraint optimization, neural networks applications, and statistical signal processing.