Optimality criteria in inverse problems for tongue-jaw interaction

A.S. Leonov

Moscow Physical Engineering Institute, Moscow, Russia
leonov@illposed.msk.ru

V.N. Sorokin

Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow, Russia
vns@iitp.ru

10.21437/Eurospeech.2003-649

Abstract

We consider the system of articulators "jaw – tip of the tongue" in order to investigate instant and integral optimality criteria in the variational approach to the solution of speech inverse problem "from total displacement of articulators to their controls". The required experimental data i.e., coordinates of the tip of the tongue and lower incisor have been measured by the use of the X-ray microbeam system together with EMGs of *masseter, longitudinalis superior* and *longitudinalis inferior*. These data have been registered for sequences of syllable /*ta*/ with different articulation rates, as well as for an elevation and lowering of the tongue tip in non-speech mode. We analyze instant and integral criteria of the work, kinetic energy, elastic and inertial forces for the system. In speech mode, total displacements of the tongue tip and the jaw are simulated perfectly by the use of any instant and integral criterion, mentioned above. At the same time, the own displacements of the tongue tip and the jaw are reproduced well by means of integral criteria only. On the contrary, the own displacements in non-speech mode are reproduced satisfactory only by the use of any instant optimality criterion.

1. Introduction

The variational method for the solution of inverse problems consists in searching conditional extrema of an optimality criterion in a space of variables for inverting mathematical model. In order to the obtained solution of inverse problem for the vocal tract be adequate to actual processes of speech production, it is necessary that optimality criteria in inverse problems are in accordance with ones in the control system of the articulation. It is known that the instant criteria of the work and the kinetic energy provide satisfactory solutions of static inverse problems for fricatives and vowels [1], [2] as well as for dynamic inverse problems for diphthongs and connected speech [3], [4], [5]. The same criteria reproduce the compensation effects for the bite-block and the reorganization effects of controls score under changes of the articulation rate [6]. The corresponding results have been obtained with the use of experimental data on the shape of the vocal tract. However, the striated muscles have a specific property: even to maintain a fixed (non-neutral) articulator state, the motor units of these muscles should contract continuously consuming the energy in the result.

Thus, we can see in speech inverse problems the importance of integral optimality criteria i.e., quantities taking into account all the values of instant criteria for some time interval. The system of bounded articulators "jaw – tip of the tongue" is useful for detailed study of optimality criteria at the level of the articulation another than the shape of the vocal tract. Here we can analyze these criteria at the level of controls for articulator displacements. A specific experimental data open up these possibilities.

2. Experimental data

We use for our mathematical modeling the data obtained by one of the authors at the University of Tokyo in 1980. In these experiments, horizontal and vertical coordinates of the lower incisor and tip of the tongue have been registered by the X-ray microbeam system together with EMGs for the muscle *masseter* or the muscles *longitudinalis superior* and *longitudinalis inferior* for one speaker. The test material consists of three kinds of tasks: 1) a cyclic repeating for sequences of syllable /*ta*/; the duration of one syllable is about 135 ms, 2) the articulation of syllable /*ta*/ with intervals varying from 700 ms to 1200 ms; 3) slow elevating and lowering the tongue tip with the task "do not touch hard palate". The muscle *masseter* elevates the jaw and shifts it forwards. In the experiments, two hook-wire electrodes, placed on 1 cm, were injected into the left *masseter* and their output signals have been summed. The contraction of the *longitudinalis superior* (the upper longitudinal muscle of the tongue) bents the tongue tip upwards while the contraction of the *longitudinalis inferior* (the lower longitudinal muscle of the tongue) bents the tongue tip downwards. Two hook-wire electrodes were injected into the left and right parts of the the *longitudinalis superior*, and one electrode recorded EMG from the *longitudinalis inferior*. The total displacement of the tongue tip, measured in the experiments, is the sum of its own displacement and the displacement of the jaw. This enables us to find the own displacements of the tongue tip as a difference of its total displacement and the displacement of the jaw. We take into account the vertical displacements only.

3. Mathematical statement of problems

Let $x_1(t)$ be a vertical displacement of the jaw from the neutral position, $x_2(t)$ be an analogous displacement of the tongue tip and $u_1(t)$, $u_2(t)$ be commands of the control system for these articulators. We describe the relations of all these values by the system of differential equations: $\ddot{x}_k + 2g_k\dot{x}_k + \omega_k^2 x = u_k(t)/m_k$, $k = 1,2$, with zero initial conditions. Here ω_k and g_k are respectively the eigenfrequency and the damping coefficient for the characteristic oscillations and m_k is the mass of the k-th articulator. The function $u_k(t)$ can be interpreted as an effective muscle force, controlling the articulator. We investigate the relation of the total displacement of the tongue tip $y(t) = x_1(t) + x_2(t)$ and the dynamic vector of the control forces $u(t) = (u_1(t), u_2(t))$. This relation can be written in a form:

$$y(t) = \int_0^t q_1(t-\tau)u_1(\tau)d\tau + \int_0^t q_2(t-\tau)u_2(\tau)d\tau \overset{def}{=} A[u], \quad (1)$$

where

$$q_k(t) = \frac{1}{2p_k}\left[\exp(-(g_k - p_k)t) - \exp(-(g_k + p_k)t)\right],$$

$$p_k = \sqrt{g_k^2 - \omega_k^2},\ k = 1,2.$$

The equation (1) presents a model binding the controls and the total displacement of the articulators by means of the operator of the model $A[u]$. The inverse problem for the model (1) is to find controls for each articulator from given approximate total displacement $y_\delta(t)$ with a mean-square error δ. It is clear that such inverse problem can have non-unique solution on the class of integrable functions. The stated inverse problem will be solved by the approach developed in [7]. In this approach, we take as an approximation to the control forces $u(t)$ to be found the vector $u^\delta(t) = (u_1^\delta, u_2^\delta)$ such that

$$\Omega[u^\delta(t)] = \min\left\{\Omega[u(t)]:\ u(t) \in U,\ \|A[u] - y_\delta\| \le \delta\|y_\delta\|\right\}. \quad (2)$$

This means that we select among all acceptable controls $u(t) \in U$ a dynamic vector $u^\delta(t)$ simulating with accuracy δ the total displacement $y_\delta(t)$ and minimizing at the same time given optimality criterion $\Omega[u]$.

4. Optimality criteria

Several criteria to the solution of the problem (2) were investigated in order to get adequate simulation for the individual displacements of the jaw and tongue tip. The object is to choose a criterion that simulates measured displacements $x_1^\delta(t)$ and $x_2^\delta(t)$ in the best way. Mathematically, this means the choice of Ω such that to minimize the value:

$$\left\|\int_0^t q_1(t-\tau)u_1^\delta(\tau)d\tau - x_1^\delta(t)\right\| + \left\|\int_0^t q_2(t-\tau)u_2^\delta(\tau)d\tau - x_2^\delta(t)\right\| \to \min_\Omega \quad (3)$$

The following optimality criteria were considered.

Type 1. Integral criteria:

$$\Omega_W[u] = \frac{1}{2T}\sum_{k=1}^2 \int_t^{t+T} c_k x_k^2(\tau)d\tau,$$

$$\Omega_F[u] = \frac{1}{T}\sum_{k=1}^2 \int_t^{t+T} c_k |x_k(\tau)| d\tau,$$

$$\Omega_K[u] = \frac{1}{2T}\sum_{k=1}^2 \int_t^{t+T} m_k \dot{x}_k^2(\tau)d\tau,$$

$$\Omega_T[u] = \frac{1}{T}\sum_{k=1}^2 \int_t^{t+T} \left[m_k \ddot{x}_k(\tau)\right]^2 d\tau.$$

They can be interpreted in such a way: Ω_W is the mean total work of elastic forces that appear in the system " jaw – tongue tip" under action of controls in a time T; Ω_F is the mean total elastic force in the system; Ω_K is the mean kinetic energy of the articulators in a time T; Ω_T is the mean total inertial force in the system in a time T.

Type 2. Instant criteria are calculated for every moment of time t and can be obtained mathematically from integral criteria by passage to the limit as $T \to 0$:

$$\Omega_{IW}[u(t)] = \frac{1}{2}\sum_{k=1}^2 c_k x_k^2(t),$$

$$\Omega_{IF}[u(t)] = \sum_{k=1}^2 c_k |x_k(t)|,$$

$$\Omega_{IK}[u(t)] = \frac{1}{2}\sum_{k=1}^2 m_k \dot{x}_k^2(t),$$

$$\Omega_{IT}[u(t)] = \sum_{k=1}^2 \left[m_k \ddot{x}_k(t)\right]^2.$$

The problem (2) has different form depending on what type of optimality criterion is used i.e., integral or instant one. For the integral criteria, optimal control forces $u^\delta(t) = (u_1^\delta(t), u_2^\delta(t))$ solve in time intervals of a sort $[t, t+T] = \{[0,T], [T, 2T], \ldots, [(N-1)T, NT]\}$ the following variant of (2):

$$\Omega[u^\delta] = \min\{\Omega[u]:\ \int_t^{t+T}\left(A[u] - y_\delta(\tau)\right)^2 d\tau \le \delta\int_t^{t+T} y_\delta^2(\tau)d\tau,\ |u_1| \le F_1, |u_2| \le F_2\} \quad (4)$$

Here N is the number of the minimization intervals at hand and the quantities $F_{1,2}$ represent maximal values of control forces admissible for the jaw and tip of the tongue respectively. For the instant criteria Ω_I, the found control forces $u^\delta(t)$ for each time moment t provide solution to another version of problem (2), namely:

$$\Omega_I[u^\delta] = \min\{\Omega_I[u]:\ |A[u] - y_\delta(t)| \le \delta|y_\delta(t)|,\ |u_1(t)| \le F_1,\ |u_2(t)| \le F_2\}. \quad (5)$$

The solution to the extremal problem (4), as well as (5), is unique and stable with respect to small variations of input data $y_\delta(t)$ for all optimality criteria listed above.

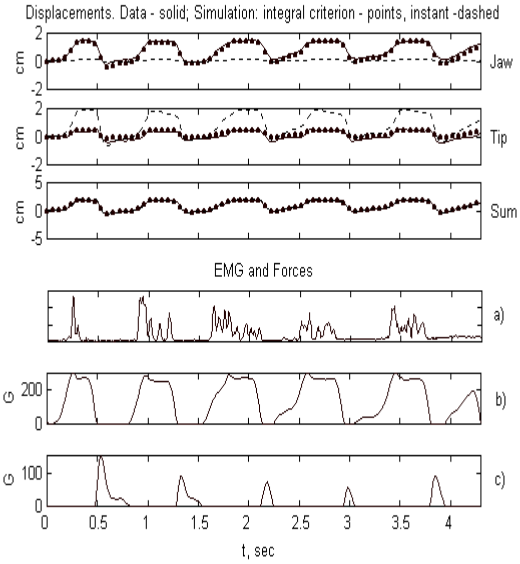

Figure 1: Sequence of syllables /*ta*/. EMG of the *masseter* (a), jaw elevating force (b), jaw lowering force (c).

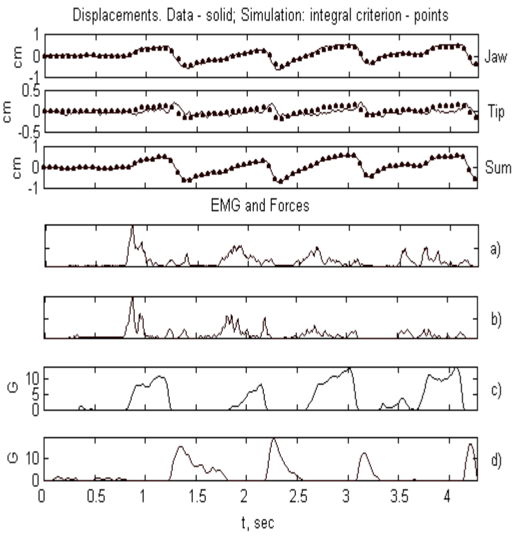

Figure 2: Sequence of syllables /*ta*/. EMG of the *longitudinalis superior* (a), EMG of the *longitudinalis inferior (b),* tongue tip elevating force (c), tongue tip lowering force (d).

5. Results of modeling

The value in the left-hand side of (3) was used as a measure of simulation quality for displacements of the jaw and the tongue tip. It was found that the appreciable difference between instant and integral criteria appears for the individual displacements of the articulators when $T \geq$ ~100 ms, while the total displacement is simulated always with good accuracy almost independent on the used criterion (see plots "Sum" in Fig.1 - 4). For the data from the cyclic and separate syllables, the individual displacements were reproduced well by the use of the integral criteria only (see Fig.1-3). All the results in Fig.1-3 correspond to Ω_T .

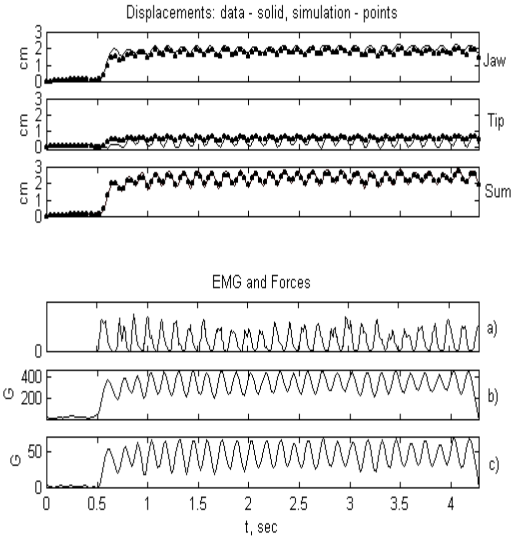

Figure 3: Cyclic articulation of syllables /*ta*/. EMG of the *masseter* (a), jaw elevating force (b), tongue tip elevating force (c).

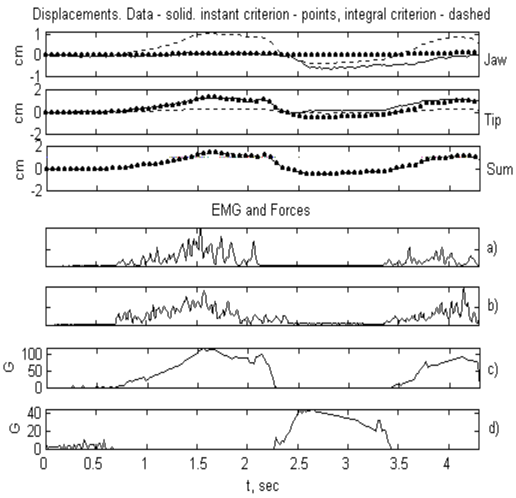

Figure 4: Non-speech movements of the tongue tip. EMG of the *longitudinalis superior* (a), EMG of the *longitudinalis inferior (b),* tongue tip elevating force (c), tongue tip lowering force (d).

The accuracy of reproduction for the other integral criteria differs to only a small extent. The values of found control forces $u_1^\delta(t)$ and $u_2^\delta(t)$, attributed to the muscles of the jaw and tongue tip respectively, are in physiologically plausible ranges (not more than 400 G and 50 G) (Fig.1,3, b,c, Fig.2, c,d). As opposed to that, the instant criteria produce significant errors in simulation of individual displacements, while a good accuracy of simulation for the total displacement is achieved due to compensatory properties of the system "jaw – tongue tip ". Unlike the experimental data with syllable articulation, the non-speech movements of the tongue tip are simulated well for voluntary elevating and lowering only by the use of the instant criteria (see Fig.4, the criterion Ω_{IT}). Again, the accuracy of simulation is practically independent on the concrete instant criterion.

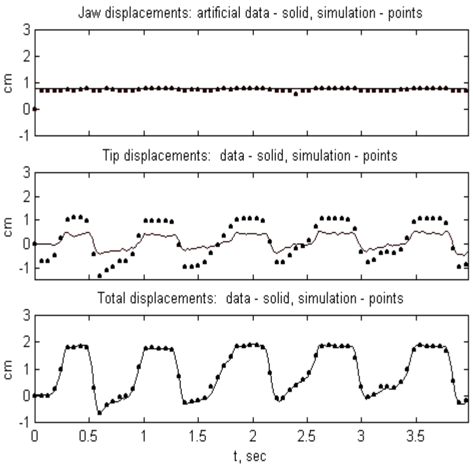

Figure.5. Bite-block simulation.

Previously, the ability of the instant criteria to reproduce the bite-block effect at the level of vocal tract shape was found in [6]. In the current study, the bite-block effect was simulated for the system "jaw - tongue tip". In the simulation, the jaw was “fixed” in a selected position i.e., the experimental displacements of the jaw were replaced by an artificial constant displacement, while the total displacements of the tongue tip remained to be as measured. Solving the extremal problem (4) with such data for all listed integral criteria, we found that the experimental movements of the tongue tip can be simulated in this situation with good accuracy in spite of the presence of “bite-block”. The plateau in the total displacement (lower part of Fig.5) indicates the restriction of the tongue tip movement due to its encounter with the hard palate, that is the alveolar closure. It is seen that the simulated alveolar closure is achieved thus compensating the bite-block.

6. Discussion and conclusions

The variational method (2) for solving inverse problem “from total displacement of the articulators to individual controls” ensures physiologically plausible results under right choice of the optimality criterion. For example, Fig. 4 shows the qualitative resemblance of calculated control function and EMG of the *longitudinalis superior*. Note, that the experiments have been carried out to guarantee the action only of this muscle for voluntary elevating and lowering the tongue tip. One cannot expect a quantitative closeness of EMG and calculated controls since we do not include in the model any assumptions about relations between electrical activity of the muscles and muscle forces. The analogous qualitative resemblance of calculated forces and EMG of activity for the *levator palatini* (the only muscle elevating the velum) has been found in our previous numerical experiments [5]. A difference in dynamical form of calculated forces and EMG in Fig. 1 - 2 shows that not only the muscles *masseter, longitudinalis superior* and *longitudinalis inferior*, are active, but one should expect the activity of other muscles-synergists. The difference between the optimality criteria providing the best approximation of movements in the speech and non-speech mode is in agreement with a concept of two distinct controlling centers for these modes. It is well known that, for certain kinds of cerebral cortex damage (for example, for a brain stroke), the articulation ability is lost while the control of non-speech movements is intact.

Analyzing the experimental data for fast cyclic articulation in comparison with the articulation of syllables separated by pauses, one can see a reorganization in motions of the jaw and the tip of the tongue (Fig. 3). Namely, the amplitude of motions for the tongue tip is substantially below than that for the jaw, and constant components of displacements appear. These reorganization effects are also reproduced in solving inverse problem with the integral optimality criteria.

The displacements of the tongue tip jointly with the jaw are equally well simulated by both the instant and integral criteria. Meanwhile, the individual displacements of the articulators are reproduced well by the integral criteria only. Instant criteria proved to be effective also in the simulation of vocal tract shape [5]. Apparently, the difference between the instant and integral criteria could not be revealed at this level because of the compensatory abilities of the articulatory system.

Thus, it is found that the integral optimality criteria are in operation at the level of articulatory displacements control, whereas the integral and instant criteria are not distinguishable in action at the level of the vocal tract shape control.

7. References

[1] Sorokin V.N. “Inverse problem for fricatives”, *Speech Communication*, *Vol. 14, N3, 249 – 262, 1994*.

[2] Sorokin V.N., Leonov A.S. and Trushkin A.V. “Estimation of stability and accuracy of inverse problem solution for the vocal tract”, *Speech Communication*, *Vol. 30, N1, 55-74, 2000*.

[3] Shirai K. and Kobayashi T. “Estimating articulatory motion from speech wave”, *Speech Communication*, *Vol. 5, 159-170, 1986*.

[4] Schoentgen J. and Ciocea S. “Kinematic formant-to-area mapping”, S*peech Communication*, *Vol. 21, 227-244, 1997*.

[5] Leonov A.S. and Sorokin V.N. “Inverse problem for the vocal tract: Identification of control forces from articulatory movements”, *Pattern Recognition and Image Analysis*, *Vol. 10, 110-126, 2000*.

[6] Leonov A.S. and Sorokin V.N. “Controls in the internal model: Score reorganization and compensation” (*in press*).

[7] Tikhonov, A.N., Leonov, A.S. and Yagola, A.G., *Nonlinear ill-posed problems*. *Vol.1,2*. Chapman and Hall, London, New-York , 1998.