Tongue motor control stability: integrating feedback, dynamical internal representation and optimal planning

Pierre Baraduc, Pascal Perrier

Tongue motor control stability: integrating feedback, dynamical internal representation and optimal planning.

Pierre Baraduc & Pascal Perrier

Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

Pierre.Baraduc@grenoble-inp.fr

Introduction

Speech motor control is known to be highly resistant to perturbations of various origins. In particular several studies have demonstrated the capacity of speech motor control system to deal with steady state perturbations such as bite-block [1], lip-tube [2], or false-palate [3], using the excess degrees of freedom of the relations between motor commands and articulatory positions or between articulatory positions and spectral characteristics of speech sounds. In some cases, a period of training was required for the subjects to achieve satisfactory adjustment of their motor control. Efficient adaptations to long lasting time-varying perturbations of the auditory-feedback [4], of the jaw movements [5, 6, 7], or of vocal tract morphology [8] have also been shown.

Differences between the expected and actual sensory feedback can follow from a structural change in the environment, as in the aforementioned studies, but they can also arise in an abrupt manner due either to unexpected mechanical perturbations, such as those we experience when we speak while running, jumping or eating, or to unexpected sensory stimuli like a sudden noise in the auditory channel. To our knowledge this question has been barely addressed in the speech motor control literature, and only the FACTS model [9] offers a modeling framework to deal at the level of the Central Nervous System with transient perturbations. The FACTS model relies for that on the concept of internal models and state feedback control [10] and provides a realistic account of the response of human subjects to a mechanical perturbation suddenly applied to the jaw during the production of bi-labial stops.

In this study we propose a slightly different modeling approach to account for both the planning of articulatory gestures and the online use of sensory feedback during speech production. Our approach relies on the concept of optimal feedback control [11], which posits that the system makes use of (a) the sensory feedbacks which are delayed and noisy, and (b) a prediction of the auditory and somatosensory feedbacks using an internal model of the dynamics of the speech production system, in order to plan articulatory movements that minimize a cost (effort, imprecision…). This ensures judicious choice of motor commands, task stability, and provides optimal correction of motor commands in case of an unexpected external perturbation. An evaluation of the model is provided (a) by assessing the impact of cost minimization on trajectory formation and coarticulation patterns in VV, VVV and VCV sequences and (b) by exploring the impact of sensory and motor noise on trajectory formation and variability.

Method

In a first step, 20,000 simulations of a finite element model of the tongue [12] have been run in order to describe the effect of combinations of muscle activations ramps across six muscles of the tongue (3 intrinsic, 3 extrinsic). Model identification techniques were then applied to

formulate a computationally tractable (polynomial) dynamical model of the tongue, which could be seen as a simplified dynamical internal model of the speech production system. Collision of the tongue with the palate or pharynx was modeled as a simple nonlinear velocity gating. From each instantaneous tongue position, the first three formants produced were calculated with a harmonic model of the vocal tract [13]. Sensory feedback of the system comprised formant frequencies, as well as tongue position, velocity and muscle activation (EMG). Feedback updated the state of the system through an extended Kalman filter. We were first interested in the generation of trajectories from a rest (e.g. schwa-like) tongue position to one of the [i], [a] and [ɔ] vowels in presence of random noise added to the motor commands and to the sensory signals. Having set the control problem as a goal-constrained neuromotor effort minimization, we applied standard numerical techniques to generate trajectories (muscle activation and tongue kinematics) from the initial posture to the final endpoints defined either in postural space or in acoustic (F1-F2-F3) space.
In parallel, we significantly simplified the dynamical model to explore multigoal sequences (VVV or VCV) while adding to the model tactile feedback for consonants (contact location and force). Consonants were modeled either as positional or tactile goals or a combination of both. We computed the optimal trajectories for the VXV sequences using various cost formulations.

Results

Optimization predicted complete muscular activation and kinematic patterns in time. Movements of flesh points were roughly sigmoidal, in accordance with reported articulatory recordings [14]. Muscle activation was consistent with the findings of tongue EMG studies [15]; most movements involved a long initial bell-shaped activation of agonists followed by a final short braking antagonist burst. Trajectories in acoustic space were interestingly sometimes strongly curved (esp. in F1-F2 space), which warrants further examination. For the same acoustic goal, muscular, mechanical and mechanic-acoustic redundancy, coupled to effort minimization, led to different final tongue postures depending on initial posture. Variability in final acoustic production emerged as a consequence of sensory and motor noise, for a fixed goal and starting posture, providing an estimate of the sensorimotor contribution to phonemic variability.
Concerning tri-goal (VXV) sequences, the simplified model offered interesting insights in to long-term coarticulation effects. For example, lengthening the initial vowel within the same time frame could lead to a faster initial movement, and a modification of the XV trajectory further on, due to neuromuscular low-pass filtering and mechanical inertia.

Conclusion

Ongoing work on VVV and VCV sequences using the full model will allow further comparison to existing experimental data, and allow assessing the validity of our model that generates articulatory trajectories with a reduced number of free parameters.

References

[1] Lubker et al. (1979), J. *Phonetics*, *7*(2), 147-161.
[2] Savariaux et al. (1995), *JASA*, 9*8*(5), 2428-2442.
[3] Brunner et al. (2011), *CLP*, *25*(8), 705-724.
[4] Houde & Jordan (1998), *Science*, *279*, 1213-6.
[5] Tremblay et al. (2003), *Nature*, *423*(6942), 866.
[6] Feng et al. (2011), *J. Neurophysiol.*, *106*, 667-79.
[7] Lametti et al. (2012), J. *Neurosci.*, *32*, 9351-9358
[8] Honda et al. (2002), *J. Phonetics*, *30*(3), 281-302.
[9] Parrell et al. (2019), *PLoS Comp. Biol* 15: e1007321.
[10] Houde & Nagarajan (2011), *Front. Human Neurosci.*, *5*, 82.
[11] Todorov & Jordan (2002), *Nat. Neurosci.*, *5*, 1226.
[12] Perrier et al. (2003), *JASA*, *114*(3), 1582-1599.
[13] Badin & Fant (1989*), STL/QPSR, 2*(3), 53-108.
[14] Saltzman & Munhall (1989), *Ecol. Psych, 1*; 333-82
[15] Baer et al. (1988), *Ann. Bull. RILP*, *22*, 7-19.