



HAL
open science

A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing

Ralf Winkler, Liang Ma, Pascal Perrier

► **To cite this version:**

Ralf Winkler, Liang Ma, Pascal Perrier. A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing. ISSP 2011 - 9th International Seminar on Speech Production, Jun 2011, Montréal, Canada. pp.235-236. hal-00605796

HAL Id: hal-00605796

<https://hal.science/hal-00605796v1>

Submitted on 4 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A model of optimal speech production planning integrating dynamical constraints to achieve appropriate articulatory timing

Ralf Winkler^{1,2,3}, Liang Ma^{1,4}, Pascal Perrier¹

¹GIPSA-lab, CNRS UMR 5216, Grenoble University, France

²ZAS/Phonetik, Berlin, Germany

³Technische Universität, Berlin, Germany

⁴Zhejiang University, China

1. INTRODUCTION

It has been often suggested that similarly to skilled arm or limb movements in humans, the production of speech gestures could be based on an optimal planning in the central nervous system. This planning would use internal representations of the speech production apparatus [1, 2] to determine the motor command patterns allowing to achieve the desired communicative goals with a minimum of effort. In this context major issues are related to the nature of the speech communication goals (targets or spatiotemporal trajectories) and to the definition of minimum of effort (minimum of changes in motor commands, minimum velocity peak, trajectory smoothness, or minimum of jerk...). GEPPETO¹, the speech production model presented in this paper, has been designed within this general theoretical framework [3]. It is based on a motor control model involving optimal planning to shape a biomechanical model of the vocal tract coupled with a harmonic acoustical model of speech production. It will be shown how force constraints can be taken into account in order to achieve appropriate articulatory timing.

2. THE GEPPETO MODEL

In GEPPETO, speech goals are linked to phonemes and they are specified as 3D ellipsoids in the acoustic space of the first three formants. These ellipsoids are determined by their centres, and by standard deviations along the directions F1, F2, F3. These standard deviations are assumed to account for the acoustic variability that is tolerated around each canonical realisation without any consequence on the auditory perception. The motor control model includes a muscle force generation mechanism based on the Equilibrium-Point hypothesis (Feldman, 1986). The motor control variable is the muscle length threshold above which active muscle force is produced. Movements are generated by shifting the motor control variables at a constant rate of shift between motor targets. In GEPPETO, the motor targets are associated with the speech goals in the acoustic domain. These motor commands are sent to the seven muscles of a 2D biomechanical model of the tongue [4] that is embedded in a 2D description of the vocal tract boundaries. The tongue model deforms and moves as the result of the combination of the influence of the target motor commands, of their timing and of the dynamical properties of the model (muscle forces, tissues elasticity, friction and contacts with external structures). For a given speech sequence, the target motor commands are selected thanks to an optimal planning in which all the targets commands are kept within the closest possible neighbourhood compatible with the perceptual goals and with constraints in terms of a global level of force. More specifically, in GEPPETO, optimal planning is constrained by the double necessity to select motor commands appropriate to the achievement of target spectral patterns inside the target ellipsoids and to ensure that the global level of force remains within a given range during the whole movement. The choice of this range is guided by the intended speaking rate and by perceptual accuracy requirements: for slow speaking rates or low accuracy requirements, a low level of force can be used, but for fast speaking rates and great accuracy a strong level of force is required. To include these constraints in the optimal

¹ GEPPETO holds for "GEstures shaped by the Physics and by a PErceptually oriented Targets Optimization"

planning process, two internal forward models have been learned. The first one called “*static forward model*”, accounts for the relation between motor commands and formants. The second one called “*dynamical forward model*” associates the motor commands with the corresponding global muscle force level. Optimisation is achieved by sequential quadratic programming (SQP), which allows closely mimicking Newton's method (Quasi-Newton) for constrained optimization. In the current state of the model three ranges of global force levels have been defined: low, normal and high. The corresponding force values were derived from a database of 8293 tongue movements simulated with the 2D biomechanical model from the rest position to a randomly chosen target. The value of total muscle force was computed for all targets. Then, the minimum and maximum forces were extracted together with an intermediate value, and they were used as levels for the three force ranges.

3. RESULTS

This optimal planning process has been tested so far on VCV sequences, starting from the tongue rest position. Optimal motor command patterns were found for each segment of the sequences under the three force constraints. Then, the 2D biomechanical tongue model was used to simulate tongue movements for these 3 force levels and for three timings of the motor commands: slow, normal and fast. Simulations are currently in progress. The first results show that the global muscle force can have a significant impact on the articulatory trajectories, in terms of curvature and in terms of positions actually reached at targets. At a slow speaking rate, changing the global level of force tends to have less consequence on the articulatory positions reached at targets than at high speaking rates, but it seems to have an impact on the trajectory shape.

4. CONCLUSION

The results support the original hypothesis that applying the appropriate dynamical constraints in the optimal planning process helps dealing with articulatory timing and perceptual accuracy expressed in terms of articulatory positions at targets. In this perspective, contrary to the intrinsic timing theories, time control could be seen as a combination of centrally specified and physically constrained characteristics.

REFERENCES

- [1] Bailly, G. (1997). Learning to speak. Sensori-motor control of speech movements. *Speech Communication* 22(2-3), 251-267.
- [2] Guenther, F.H., Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, pp. 611–633.
- [3] Perrier, P. & Ma., L. (2008). Speech planning for v_1Cv_2 sequences: Influence of the planned sequence. *Proceedings of the 8th International Seminar on Speech Production (ISSP 2008)* (pp. 69-72). Université de Strasbourg, France.
- [4] Perrier, P., Payan, Y., Zandipour, M. & Perkell, J. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *JASA*, 114(3), pp. 1582-1599.