Speech motor learning as a process between acoustic target achievement and movement optimisation

Jana Brunner[1 2 3] **, Phil Hoole**[4] **, Pascal Perrier**[2]
[1]Humboldt-Universität zu Berlin,
[2]GIPSA-lab, INPG, CNRS & Univ. Stendhal Grenoble
[3]Zentrum für Allgemeine Sprachwissenschaft, Berlin
[4]Institut für Phonetik und Sprachliche Kommunikation der Universität München

brunner@zas.gwz-berlin.de

Speech production is at the same time a semiotic and a motor task. As such it has to reach communicative objectives, while respecting the same constraints and rules of other skilled motor tasks carried out by humans.

Models of speech production and speech acquisition usually include an internal model. Jordan (1996), for example, proposes a model with two components, a forward model and an inverse model. The forward model gives a functional account of the relations between acoustic outputs and is therefore able to predict the output for a certain motor input. It is trained during speech acquisition from the generalisation of a large number of observations of the association between inputs and outputs (Kawato *et al.* (1990)).

The inverse model goes the other way and supplies a motor input for a desired acoustic output. Since there is a one-to-many relation between desired acoustic outputs and motor inputs constraints have to be applied to reduce the number of degrees of freedom and thus find a unique motor input. Classically, these constraints are associated with the concept of optimality; the inferred motor input is optimal according to a given criterion. One of the criteria often used in the literature is articulatory effort. During the training phase of the inverse model the articulatory effort is measured and the inputs of the model are adapted (classically with a gradient backpropagation technique) so that with more and more training motor commands are chosen for which the articulatory effort is low.

There is no unique definition of articulatory effort (see Nelson (1983), for an interesting analysis of this concept). For example, Guenther *et al.* (1998) or Perrier *et al.* (2005) solve the inverse problem via a minimisation of the Euclidean distance either in motor, articulatory or auditory-acoustic space.

When speech is perturbed speakers should be confronted with an internal model which no longer predicts the correct acoustic result for a certain motor input. The confrontation with new input-output pairs should lead to a retraining of the internal model. For a model of arm movements, Jordan (1989) has proposed that the training of the forward and the inverse model takes place at the same time and independently of one another. In the study presented here we are looking for evidence for an optimisation process during the adaptation to perturbed speech. Furthermore, if such a process exists, we are interested in whether it starts once the acoustic targets (associated with the semiotic objectives of speech production) have been reached or whether in speech, same as Jordan proposes for arm movements, the forward model, accounting for the new associations between motor commands and acoustic outputs, and the inverse model, allowing the gestural optimisation, are trained parallely and independently of each other.

Palatal prostheses which changed the palatal contour were made for seven speakers. Subjects wore the palates for 14 days and were recorded via electromagnetic articulography and acoustics at first without the prosthesis, then with the prosthesis regularly over the adaptation period, and at the end of the experiment once more without the prosthesis. The target sounds were several fricatives, stops and vowels embedded in nonsense words produced in a carrier phrase. The sentences were repeated

20 times per session in randomised order. The vocalic and consonantal gestures of the tongue tip were labeled on the tangential velocity signal using a 20% threshold criterion. Articulatory effort was assessed by means of peak tangential acceleration, tangential jerk and movement amplitude (Nelson (1983)).

The acoustic analysis included formant measurements and the measurements of six parameters characterising obstruent spectra. A discriminant analysis with these acoustic parameters was calculated with the preperturbed session as first group and the first perturbed session as second group. For the following sessions the probability of the producions to belong to the first group has been calculated. For the vowels it was found that speakers fully compensate rather soon, sometimes immediately, somtimes after a few minutes. For the stops it has been found that speakers differ in the time they need in order to compensate, but normally it takes a few days. The fricatives need longest for the acoustic adaptation.

Articulatory effort was measured at three stages of particular interest in the adaptation process: During the unperturbed session (where maximally optimised movements can be expected), when articulatory effort was maximal during the perturbation, and at the end of the perturbation (where again maximally optimised movements can be expected). Analyses of variance for the difference between the unperturbed value and the maximal value, and between the maximal value and the last perturbed value for each of the two parameters were calculated with SPSS 13.

Looking at the articulatory effort it has been found that the values of peak acceleration and jerk increase after perturbation onset, afterwards they decrease. The moment when the optimisation, the decrease in articulatory effort, starts depends on the speaker and is about the same for all sounds produced by the same speaker. For movement amplitude this result could not be found.

The results thus provide evidence that an optimisation resulting in a minimisation of jerk and peak acceleration, as proposed by Jordan, takes place whereas a reduction of the movement amplitude as proposed by the other models, could not be found. Furthermore, in line with the proposals by Jordan, this process is independent of the quality of the acoustic output. Assuming an internal model with a forward and an inverse component this means that the inverse model, permitting the gestural optimisation, can be trained even if the forward model, accounting for the new relation between motor commands and acoustic outputs, is not yet entirely correct.

References

Guenther, F. H., Hampson, M., and Johnson, D."A theoretical investigation of reference frames for the planning of speech movements", Psych. Rev. **105**, 611–633.

Jordan, M."Indeterminate motor skill learning problems", in *Attention and Performance, XIII. Cambridge, MA*, edited by M. Jeannerod (MIT Press, Cambridge).

Jordan, M."Computational aspects of motor control and motor learning", in *Handbook of Perception and Action: Motor Skills*, edited by H. Heuer and S. Keele (Academic Press, New York).

Kawato, M., Maeda, Y., Uno, Y., and Suzuki, R."Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion", Biol. Cybern. **62**, 275–288.

Nelson, W. L."Physical principles for economies of skilled movements", Biol. Cybern. **46**, 135–147.

Perrier, P., Ma, L., and Payan, Y."Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue", in *Proc. of Interspeech 2005*, edited by I. S. C. Association, 1041–1044.