Multidirectional mappings and the concept of a mental syllabary in a neural model of speech production

Bernd J. Kröger[1], Peter Birkholz[2], Jim Kannampuzha[1], Christiane Neuschaefer-Rube[1]

[1] *Department of Phoniatrics, Pedaudiologie and Comunication Disorders, University Hospital Aachen (UKA) and Aachen University (RWTH), Pauwelsstr. 30, 52074 Aachen, Germany*
Email: bkroeger@ukaachen.de, jkannampuzha@ukaachen.de, cneuschaefer@ukaachen.de

[2] *Institute of Computer Sciences, University of Rostock, Albert-Einstein-Str. 21, 18059 Rostock, Germany*
Email: piet@informatik.uni-rostock.de

Abstract

As a result from modeling cortical processes of self-organization occuring during speech acquisition, a comprehensive neural model of speech production has been developed by using self-organizing neural networks and feed-forward neural networks. This model is capable of generating acoustic speech signals and sensory feedback signals by using a high quality 3-dimensional articulatory-acoustic speech synthesizer as a front-end device. A mental syllabary forms the central layer within this model. The mental syllabary comprises a heap of neural SOM layers (phonetic map) which can be interpreted as a system of mirror neurons co-activating phonemic, sensory, and motor states of a syllable under production (feed-forward control). Feedback control is modeled by comparing the current sensory feedback state produced by the articulatory-acoustic model with the pre-stored sensory state (efference copy), activated during feed-forward control via the mental syllabary. This model can be integrated easily as a phonetic part within a more general linguistic model of speech production.

Introduction

A model of speech production can be subdivided into a control module – comprising feed-forward and feedback control – and an articulatory-acoustic module, i.e. into a controller and a controlled system or plant. While a lot of knowledge has been collected concerning the plant over the last decades – i.e. concerning static articulatory data and time-dependent kinematic articulatory data and concerning the articulatory-acoustic modeling – much less knowledge is availabe concerning the neural control of speech production. The goal of our current work is to develop a comprehensive neural model of speech production which closely reflects as many aspects as possible of the natural human speech production process including self-organization of cortical structures occurring during speech acquisition.

One-layer feed-forward mappings, self-organizing maps (SOM's) and multidirectionality

Two different kinds of neural networks have been used in our modeling thus far, i.e. *unidirectional one-layer feed-forward neural networks* [1] and *multidirectional self-organizing neural networks* [2]. While one-layer feed-forward networks can be used successfully for modeling the somatosensory-to-motor mapping, self-organizing neural networks are feasible for modeling many aspects of sensory-to-motor mappings for different tasks, e.g. for modeling the auditory-to-motor mapping for vowels and simple syllables like vowel-plosive combinations [2].

Self-organizing neural networks comprise one central layer (self-organizing map, SOM) and one or more side layers (phonemic, sensory, and motor layers, Fig 1). The central layer can be interpreted as a *cortical layer of mirror neurons* [3] leading to a multidirectional co-activation of all side layers – i.e. of the phonemic, the sensory, and the motor layers. Thus an activation of a syllable within the phonemic map leads to a co-activation of the appropriate auditory, somatosensory and motor states (production). Or an activation of the sensory state of a syllable leads to co-activation of the appropriate motor and phonemic state (perception). At least also the activation of the high-level motor state of a syllable (i.e. the motor plan of a syllable; covert speech) – for example induced by visual stimulation – leads to co-activation of the sensory and phonemic state.

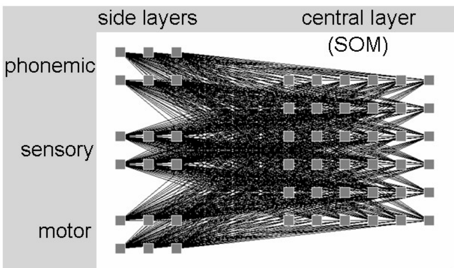

Figure 1: The organization of a multidirectional self-organizing network.

Towards a control model including multidirectional mappings and a mental syllabary

Our modeling results collected thus far for babbling and imitation can be subsumed in a comprehensive production model (Fig. 2). Within this control model the SOM layers of all mappings described above are called *phonetic map*. The *phonemic map* comprises the phonological codes of all sounds and frequent syllables of a language. Infrequent syllables are processed via a motor planning module. Multidirectional co-activation of phonemic, sensory, and motor states occur as described above via the phonetic map. The phonemic map and parts of the phonetic map form the *mental syllabary* (cp. [4]). Feed-forward control startes with activation of a phonemic state for a currently activated syllable, followed by co-activation of the appertaining sensory and (high-level) motor state (i.e. motor plan) of this syllable via the mental syllabary or motor planning module. Feedback control is activated if the (prestored) co-activated sensory state deviates from the current feedback sensory state. A separation of motor planning and motor execution is also included in our approach (Fig. 2)

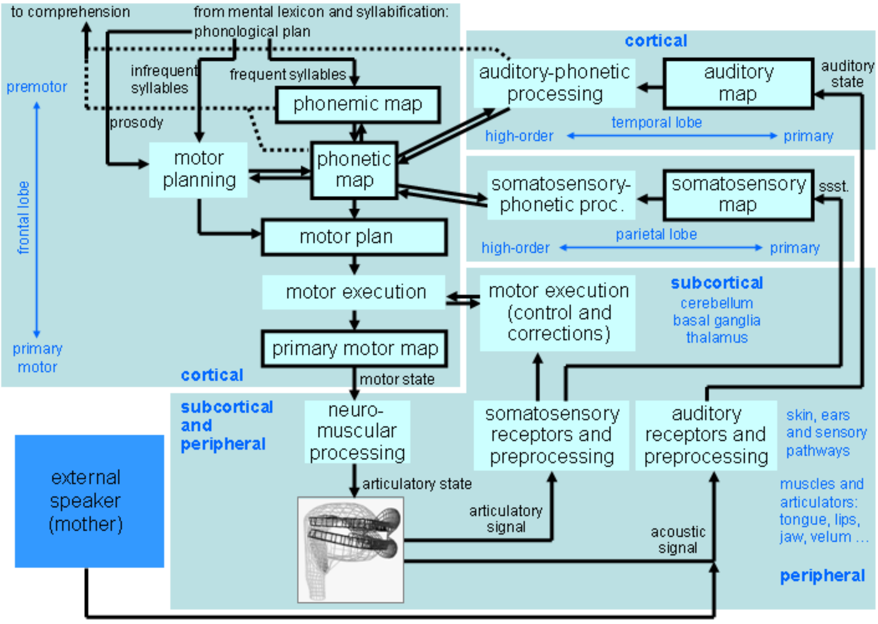

Figure 2: The organization of the multidirectional model of speech production (phonetic part).

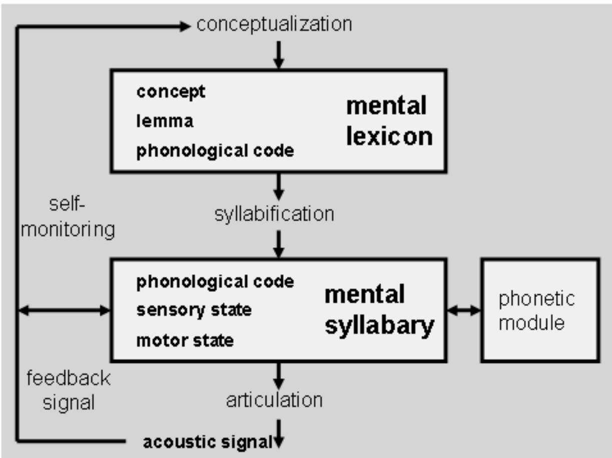

Figure 3: The organization of the complete linguistic and phonetic model of speech production.

A complete model of speech production

Our neurophonetic model of spech production can be seen as a part of the more general speech production model introduced by Levelt [4, 5, 6]. Two repositories – i.e. the *mental lexicon* and the *mental syllabary* are central components within this approach. Concepts, lemmas, and the phonological codes of lexical items are stored within the mental lexicon. The phonological codes and the motor plans of frequent syllables are stored within the mental syllabary. In our approach the mental syllabary in addition comprises the sensory representations of these syllables (efference copies) and is thus closely related to a *phonetic module*, i.e. to a heap of phonetic self-organizing maps, which are responsible for language specific production and perception effects (Fig. 3).

Discussion

A comprehensive neural model of speech production is introduced. While this concept is mainly compatible with the approach of Guenther (e.g. [7]), it strongly emphasizes the existence of multidirectional mappings on the level of the mental syllabary, leading to a co-activation of the phonemic state of a syllable with all co-occurring phonetic states, i.e. with sensory and motor states. A phonetic module – also occurring on the level of the mental syllabary – is introduced here as a central layer of mirror neurons, which are responsable for the co-activation of phonemic, sensory, and motor states. In addition a separation of motor planning and execution is introduced here. This model can be integrated as a phonetic part into a more general concept of speech production as introduced by Levelt [4, 5, 6].

Acknowledgments

This work was supported in part by the German Research Council Grant KR 1439/10-1 and JA 1476/1-1.

References

[1] Kröger BJ, Birkholz P, Kannampuzha J, Neuschaefer-Rube C (2006a) Modeling sensory-to-motor mappings using neural nets and a 3D articulatory speech synthesizer. *Proceedings of the 9th International Conference on Spoken Language Processing* (Interspeech 2006 – ICSLP) pp. 565-568

[2] Kröger BJ, Birkholz P, Kannampuzha J, Neuschaefer-Rube C (2006b) Learning to associate speech-like sensory and motor states during babbling. *Proceedings of the 7th International Seminar on Speech Production* (Belo Horizonte, Brazil) pp. 67-74

[3] Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297: 846-848

[4] Levelt WJM, Wheeldon L (1994) Do speakers have access to a mental syllabary? *Cognition* 50: 239-269

[5] Levelt WJM (1992) Accessing words in speech production: stages, processes and representations. *Cognition* 42: 1-22

[6] Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. *Cognition* 92: 101-144

[7] Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language* 96: 280-301