Journal of Phonetics ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

The emergence of phonetic–phonological features in a biologically inspired model of speech processing

Bernd J. Kröger[a,*], Mengxue Cao[b]

[a] Neurophonetics Group, Department of Phoniatrics, Pedaudiology, and Communication Disorders, Medical School, RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany
[b] School of Chinese Language and Literature, Beijing Normal University, No. 19 Xin Jie Kou Wai Da Jie, Beijing 100875, China

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form
17 September 2015
Accepted 21 September 2015

Keywords:
Neural model of speech processing
Speech production
Speech acquisition
Articulatory–acoustic modeling
Phonetic features
Phonological features
Distinctive features

ABSTRACT

Because speech acquisition begins with sensorimotor activity (i.e. babbling and imitation of speech items in order to learn articulatory–acoustic relations) as well as with semantic cognitive processing (i.e. linking phonetic items with concepts), distinctiveness as well as phonetic–phonological features emerge early in speech acquisition. Based on a biologically inspired model of speech processing, using interconnected growing self-organizing maps (I-GSOMs), the phonetic–phonological interface is described here in terms of a numerical computer-implemented model. By simulating early phases of speech acquisition, it can be shown that vocalic features like low–high as well as consonantal features describing the manner of articulation (like plosive, fricative, nasal, lateral, etc.) already arise at sensorimotor levels. This is reflected in our model by an ordering of syllables with respect to phonetic–phonological features at the level of an auditory based neural map. Other features like consonantal place of articulation (labial, apical, dorsal) as well as voiced–voiceless emerge at higher levels within our model, i.e. at the level of neural associations between a phonetic and a semantic neural map. It can be hypothesized from these findings that the phonetic–phonological interface does not appear as a clean cut within the speech processing system but as a broader zone within that system located between sensorimotor and semantic processing.

1. Introduction

Most models of speech production as well as of speech perception cover either cognitive linguistic aspects (Levelt, Roelofs, & Meyer, 1999; Levelt & Indefrey, 2004; Dell, Burger, & Svec, 1997) or sensorimotor aspects (Guenther, Ghosh, & Tourville, 2006; Guenther & Vladusich, 2012; Civier, Bullock, Max, & Guenther, 2013; Saltzman & Munhall, 1989). An often used interface between those models is the phonological representation (Hickok & Poeppel, 2007; Hickok, 2012). But if we intend to model speech learning, these approaches are not very helpful, because the phonological representation of a target language is not present at the very beginning, but emerges during speech acquisition (Eckers & Kröger, 2012). It is the goal of this paper to describe an integrative model of speech processing, exhibiting a connection between cognitive and sensorimotor parts, which allows to model these complex aspects of the phonetic–phonological interface including the option of an emerging phonological representation during speech acquisition (Kröger, Birkholz, & Neuschaefer-Rube, 2011; Kröger, Birkholz, Kannampuzha, Kaufmann, & Neuschaefer-Rube, 2011; Cao, Li, Fang, & Kröger, 2013; Cao, Li, Fang, Kaufmann, & Kröger, 2014).

In a first step the sensorimotor part of this model has been developed (Kröger, Kannampuzha, & Neuschaefer-Rube, 2009). Here we already separated premotor or motor plan representations (Kröger & Birkholz, 2007) and primary motor representations (i.e. complete specification of all control parameters of an articulatory synthesizer; Birkholz, Jackel, & Kröger, 2006) as a two-step feedforward processing model. Auditory and somatosensory feedback information together with the motor plan (feedforward) information was already used as input information in this approach in order to establish a self-organized map, linking sensory and motor information within pre-linguistic babbling training. Specific pre-linguistic vocalic and pre-linguistic closing–opening movement sets (i.e. pre-linguistic CV-syllables) were used for training initial versions of vocalic and consonantal self-organizing maps. In a

* Corresponding author. Tel.: +49 241 8085222.
E-mail addresses: bernd.kroeger@rwth-aachen.de (B.J. Kröger), mengxuecao@outlook.com (M. Cao).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.09.006

second training step, language specific training sets for vowels and CV-syllables were added, establishing subregions within the self-organizing maps which were identified as neurocomputational correlates of phoneme regions.

A main shortcoming of this approach was (i) that the size of the self-organizing maps needed to be predefined and (ii) that a phonological representation needed to be predefined. But it can be assumed that the size of a sensorimotor, of a phonological, as well as of a semantic representation increases during speech acquisition if the amount of speech sounds, syllables and words increases. Thus we introduced growing self-organizing maps (GSOMs) for modeling parts of the mental lexicon (i.e. semantic map or S-MAP, Cao et al., 2013) as well as for modeling the sensorimotor part of the speech processing system, here called phonetic map or P-MAP (Cao et al., 2014). Moreover a first version of a neural coupling between P-MAP and S-MAP has been described in Cao et al. (2014) as well.

It is the goal of this paper to describe simulations of speech learning using the current version of the model (Cao et al., 2014) focusing on the emergence of phonetic–phonological features. This occurs on the basis of an ordering of speech items at the level of the self-organizing phonetic map and on the basis of the fact that these speech items are linked unambiguously with meanings (i.e. S-MAP nodes). Therefore, the model is introduced in detail in the next section. The training data base (speech learning data base) for simulating early stages of speech acquisition is introduced in Section 3. The speech acquisition scenario (learning scenario) is described in Section 4 in detail, learning results after occurring from this simulation process are given in Section 5, and results as well as the advantages and disadvantages of our approach are discussed in Section 6. Detailed information concerning our data set is given in Appendix A, a listing and explanation of phonetic and linguistic key items as they are used in the context of this paper and this neural model are given in Appendix B.

2. The model

2.1. Organization of the model

Our model of speech processing (Kröger et al., 2009; Cao et al., 2014) can be divided into a cognitive lexical and a sensorimotor or phonetic part (Fig. 1). The *sensorimotor part* comprises a syllable-based speech action repository (Kröger et al., 2009; Kröger & Heim, 2013; Eckers, Kröger, & Heim, 2013; Eckers, Kröger, Sass, & Heim, 2013; Kröger, Kannampuzha, & Kaufmann, 2014) as well as a feedforward and a feedback processing route for controlling an articulatory–acoustic vocal tract model (Kröger & Birkholz, 2007; Kröger & Birkholz, 2009). The center of the speech action repository is modeled as a neural map (i.e. an amount of model neurons located in the same cortical area, most likely in a few mm^2 regions within a part of the prefrontal as well as within a part of the temporo-parietal cortex, see Kröger & Heim, 2013) representing all trained or learned syllables of the target language by specific model neurons. This neural map is called *phonetic map* (P-MAP) and is implemented as a growing self-organizing map in our quantitative approach (GSOM, see Cao et al. (2014)). Each syllable is represented here by an individual model neuron within the phonetic map (Fig. 2 for P-MAP based on auditory training, following the learning scenario described in Cao et al. (2014)). The phonetic map is associated with an auditory, somatosensory and motor plan map. Here, the learned neural representation of the auditory and somatosensory impression as well as of the motor plan appear for short time intervals, i.e. for those time intervals during which the syllable is activated at the level of the phonetic map. The neural association between a model neuron within the phonetic map and all neurons within both sensory maps as well as between that model neuron and all neurons within the motor plan map determines the current activation patterns occurring in

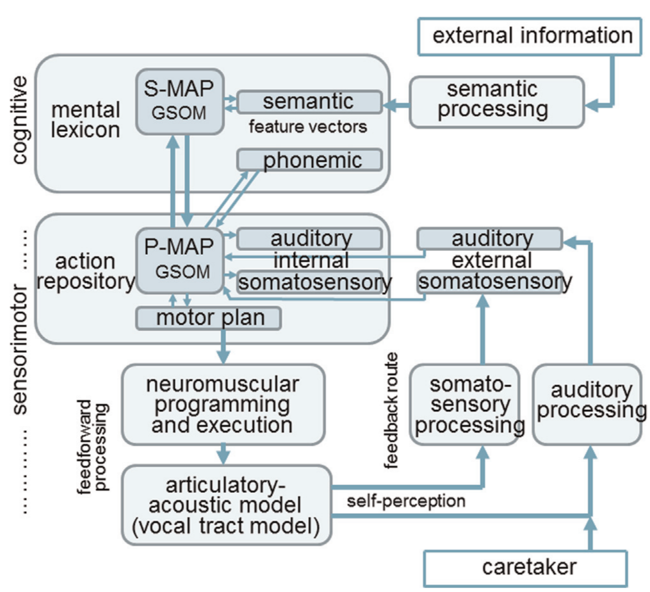

Fig. 1. Biologically inspired model of speech processing and speech acquisition. Rounded light blue boxes indicate processing modules and rounded blue boxes indicate neural maps. Squared boxes indicate external information (coming from caretaker or from other sources). In this neural model, knowledge is stored as link weights values in all bidirectional mappings (indicated by bidirectional arrows) as well as within the two self-organizing neural maps (semantic and phonetic map; see text). Unidirectional mappings simply forward information (forward patterns of neural activity). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)