MODELING THE PERCEPTUAL MAGNET EFFECT AND CATEGORICAL PERCEPTION USING SELF-ORGANIZING NEURAL NETWORKS

[1]Kröger BJ, [2]Birkholz P, [1]Kannampuzha J, [1]Neuschaefer-Rube C

[1]Department of Phoniatrics, Pedaudiology, and Communication Disorders, University Hospital Aachen and Aachen University, Germany

[2]Department of Computer Science, University Rostock, Germany

bkroeger@ukaachen.de, piet@informatik.uni-rostock.de, jkannampuzha@ukaachen.de, cneuschaefer@ukaachen.de

ABSTRACT

Purpose: A neural model of speech production based on self-organizing neural networks and comprising motor and sensory modules for feedback and feed-forward control is introduced. The model is capable of describing speech acquisition stages as well as speech perception effects. **Method:** 20 instances of the neural model were trained as in early stages of speech acquisition (babbling and imitation) in order to create 20 different virtual toddlers. Perceptual experiments were performed using these virtual listeners. **Results:** Typical effects of speech perception occur during identification experiments on vocalic and consonantal acoustic stimulus continua. Consonantal categorical perception directly occurs during babbling while the vocalic perceptual magnet effect occurs later on during language specific imitation training. **Conclusion:** The introduced neural model of speech production using self-organizing neural networks is capable (a) of illustrating the close relationship between production and perception of speech and (b) of elucidating the formation of speech perception during speech acquisition.

Keywords: speech production, neural model, perceptual magnet effect, categorical perception, self-organization, speech acquisition

1.INTRODUCTION

The perceptual magnet effect (PME [1, 2]) as well as categorical perception (CP [3, 4]) indicate better perceptual discrimination of acoustic items at phoneme boundaries than within phoneme regions. A main difference of both effects is that PME is language specific and thus is acquired by toddlers during speech acquisition while CP of many phonetic features may result from general speech perception mechanisms and occurs without training for humans (adults and newborns) as well as for some animal species [2]. In this study a neural model of speech production [5, 6] is used to elucidate the formation of speech perception effects during speech acquisition.

2.THE PRODUCTION MODEL

The model used here [5, 6] is based on the Guenther approach [7]. It comprises self-organizing maps (SOM's) and mappings for modeling the relations between the phonemic, sensory, and motor level (Fig. 1). In addition to Guenther this model separates motor planning and motor execution [6]. The mappings for feedback and feed-forward control of articulation are established during the babbling and imitation phase of speech acquisition [7, 8]. A central feature or our approach is the phonetic map co-activating the phonemic, sensory, and motor state of a syllable currently perceived or produced. This map can be interpreted as a phonetic mirror neuron layer [6] and constitutes the central layer of SOM's (Fig. 2) for different types of syllables.

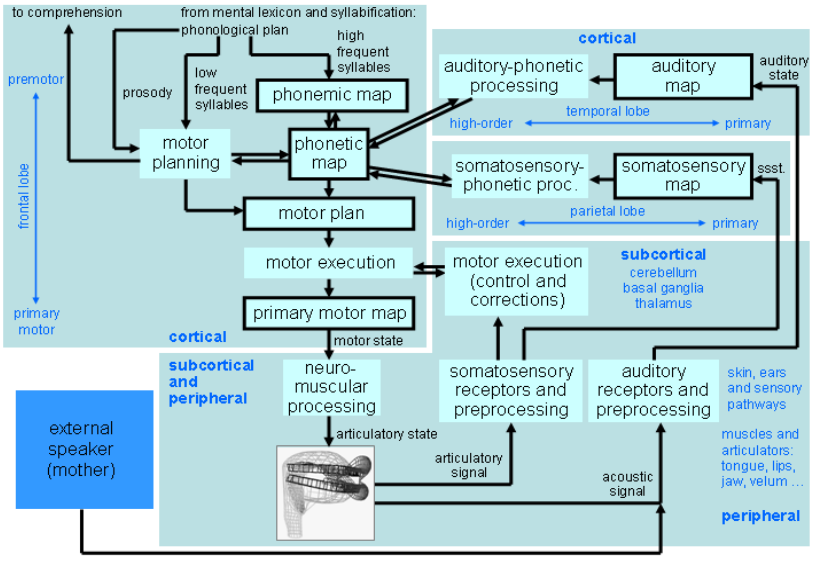
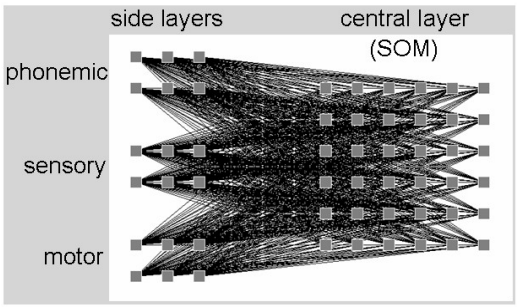


Figure 1: The neural model of speech production.

Figure 2: Example of a multidirectional self-organizing neural network.



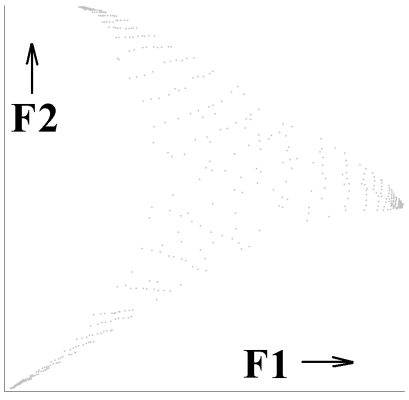
3.NEURAL REPRESENTATIONS

3.1.Proto-vocalic articulation and vowel phonemes

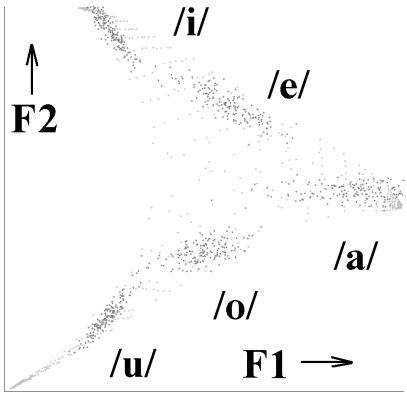
During babbling a proto-vocalic state set ([5] and Fig. 3a) and later on during imitation a set of vowel phoneme realizations for a hypothetical 5 phoneme vowel system (Fig. 3b) are used for training a self-organizing vocalic neural net (15x15 vocalic SOM neurons, standard learning parameters [9]). The whole proto-vocalic space is equally covered by SOM F1-F2 link weights after babbling training (Fig. 3c). A shift of link weights to phonemic F1-F2 regions and a concentration of link weights within these regions occurs after language specific imitation training (Fig. 3d). SOM neurons now represent phoneme regions within the vowel space (Fig. 3e).

Figure 3: (a) Proto-vocalic babbling training data (540 states). (b) Training data for babbling and for a five vowel phoneme system /i/-/e/-/a/-/o/-/u/ (100 states per phoneme). (c) Grid plot of F1-F2 link weights for each vocalic SOM neuron after proto-vocalic babbling and (d) after imitation training of the five vowel phoneme system. (e) Bar plot of phoneme link weights for each SOM neuron. Each box represents one of the 15x15 SOM neurons. Within each box: Bars from left to right represent the phonemic link weights for /i/-/e/-/a/-/o/-/u/.

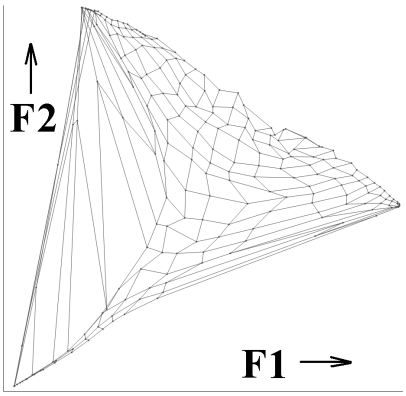
(a)



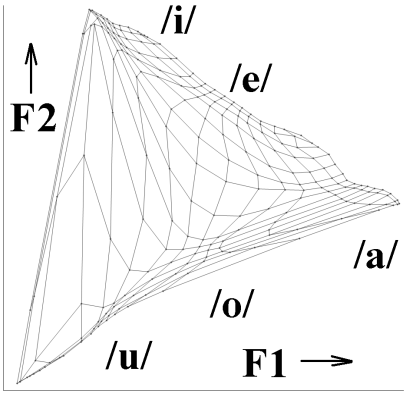
(b)



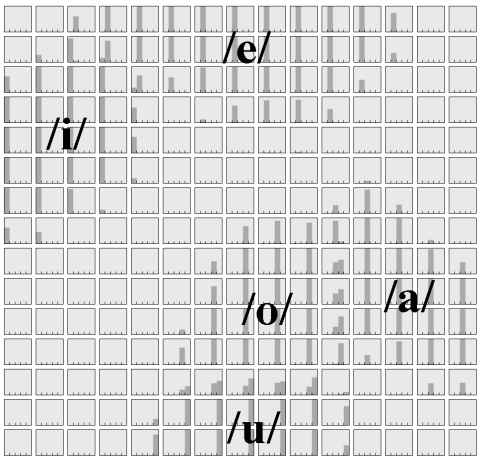
(c)



(d)



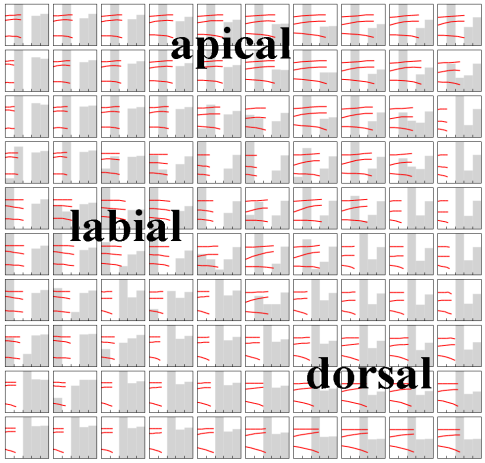
(e)



3.2.Proto-closing gestures

In parallel during the babbling phase a set of proto-closing gestures [5] is used for training a self-organizing neural network for proto-closing gestures (10x10 VC-SOM neuron layer, standard learning parameters [9]). The closing gestures (VC-gestures) start from different proto-vocalic states and produce labial, apical, and dorsal closures. After training a clear separation of labial, apical, and dorsal closing gestures can be observed in the VC-SOM neuron layer (Fig. 4).

Figure 4: Bar plot of somatosensory link weights for each VC-SOM neuron after babbling of prelinguistic proto-closing gestures. Each box represents one of the 10x10 SOM neurons. Each box comprises 5 bars; from left to right: bar 1 to 3 represent labial-apical-dorsal closure, bar 4 and 5 represent the front-back and high-low tongue position of the proto-vocalic starting vowel. The F1-F2-F3-trajectories represent the auditory link weights for each SOM neuron.



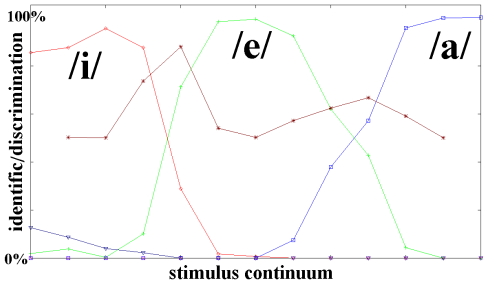
4.RESULTS

4.1.Perceptual magnet effect

The concentration of F1-F2 vocalic SOM link weights in phoneme regions of the vowel space after imitation training (Fig. 3b) is not responsible for the perceptual magnet effect. Rather this concentration of neurons within phonemic regions would lead to the opposite effect; i.e. better differentiation of items within phoneme regions, since more receptor neurons in a definite region lead to a better perceptual differentiation in this region [10]. A possible solution for this problem is given here: On the basis of the multidirectional architecture of our production model (Fig. 1) it can be hypothesized that the effect of better perceptual differentiation of auditory vocalic states because of the concentration of vocalic SOM neurons in the center of a phoneme region (Fig. 3b) is overridden by categorical phonemic knowledge: Neurons representing a definite phoneme region show a stable maximum link weight for this phoneme (Fig. 3e). This leads to a decrease in perceptual discrimination within this region since all these neurons represent one category or phoneme. In other words: Vocalic perception is not simply based on the association of the phonetic and auditory map but is also strongly influenced by the association of the phonetic and phonemic map (Fig. 1). It can be hypothesized, that this (top-down) association of phonemic to phonetic representations dominates speech perception.

In order to underline this hypothesis, an identification test was performed using a quasi-continuous [i-e-a]-stimulus continuum for 20 virtual toddlers, i.e. for 20 instances of our production model trained using different initial link weight values for all neural mappings within the model and using different random orderings of training items in all training sets. Vowel identification is done by calculating the most activated neurons (winner neurons) within all 20 phonetic vocalic SOMs for each acoustic stimulus from the auditory-phonetic associations and by identifying the appropriate phoneme from the phonetic-phonemic associations. The overall percentage of vowel identification and the subsequently calculated percentage of discrimination [11] is given in Fig. 5.

Figure 5: Percentage of identification and calculated discrimination for 13 vocalic stimuli ([i-e-a]-continuum) for 20 virtual toddlers illustrates the PME.

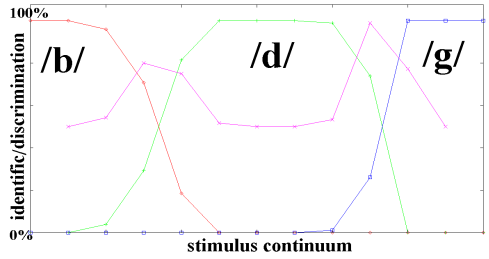


4.2.Categorical perception

The clear separation of labial, apical, and dorsal closing gestures in the VC-SOM (Fig. 4) directly accounts for categorical perception. An identification test was performed using a quasi-continuous [ba-da-ga]-stimulus continuum for the same 20 virtual toddlers. Identification of the closure-performing articulator (i.e. categories labial - apical -

dorsal) is done by calculating the most activated neurons (winner neurons) within all 20 phonetic VC-SOMs for each acoustic stimulus from the auditory-phonetic associations and by identifying the perceived articulator from the phonetic-somatosensory associations. The results are given in Fig. 6.

Figure 6: Percentage of identification and calculated discrimination of 13 consonantal stimuli ([ba-da-ga]-continuum) by 20 virtual toddlers illustrates CP.



5.DISCUSSION

On the one hand the PME occurs in our model of speech production as a result of proto vocalic babbling *and* imitation of language-specific vocalic items. Thus the PME needs language specific training data. On the other hand CP of place of articulation for consonants already occurs during babbling training of prelinguistic proto-consonantal closing gestures (raw gestures). Since our babbling training of raw gestures is simply based on labial, apical, and dorsal closing gestures, it can be hypothesized that phonetic knowledge acquired during consonantal babbling just results from the physiological fact, that just three oral consonantal closing raw gestures, i.e. labial, apical, and dorsal raw gestures can be produced. It can be hypothesized, that this physiological-phonetic fact of three oral consonantal articulators and its perceptual consequences is incorporated into general speech perception mechanisms arisen during evolution and thus is not necessarily a result of individual training during speech acquisition.

6.CONCLUSIONS AND FURTHER WORK

Our main goal was the development of a comprehensive neural model of speech production. As shown in this paper, a side effect is, that this model is capable of predicting effects of speech perception in a straightforward way. This fact demonstrates the high quality level of our production model, since the neural maps and mappings used in our model were designed primarily for modeling speech production.

During further work two topics should be focused on: (1) Modeling acquisition of voiceless consonants – i.e. acquisition of temporal coordination of glottal opening-closing and oral closing-opening gestures – may elucidate effects of categorical perception of VOT. (2) Modeling of prosodic structure should be included, since intonation patterns as well as specific segmental and syllabic structures are acquired early and easily by toddlers.

7.REFERENCES

[1] Kuhl PK (1991) Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50, 93-107

[2] Kuhl PK (2000) A new view of language acquisition. *Proceedings of the National Academy of Science* 97, 11850-11857

[3] Eimas PD (1963) The relation between identification and discrimination along speech and non-speech continua. *Language and Speech* 6, 206-217

[4] Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychological Review* 74, 431-461

[5] Kröger BJ, Birkholz P, Kannampuzha J, Neuschaefer-Rube C (2006) 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

[6] Kröger BJ, Birkholz P, Kannampuzha J, Neuschaefer-Rube C (2007) Multidirectional mappings and the concept of a mental syllabary in a neural model of speech production. *Proceedings of the Annual Meeting of the German Acoustical Society DAGA* (Stuttgart, Germany) see also: http://www.speechtrainer.eu

[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

[8] Oller DK, Eilers RE, Neal AR, Schwartz HK (1999) Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of Communication Disorders* 32, 223-245

[9] Kohonen T (2001) *Self-organizing maps.* Berlin: Springer, 3rd edition

[10] Kandel ER, Schwartz JH, Jessell TM (2000) *Principles of neural science.* New York: MacGraw-Hill

[11] Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368