

SEMANTIC, PHONETIC, AND PHONOLOGICAL KNOWLEDGE IN A NEUROCOMPUTATIONAL MODEL OF SPEECH ACQUISITION

Cornelia Eckers¹ and Bernd J. Kröger^{1,2}

¹ *Department of Phoniatrics, Pedaudiology, and Communication Disorders,
RWTH Aachen University, Aachen, Germany*

² *School of Computer Science and Technology, Tianjin University, Tianjin, P.R. China
(ceckers, bkroeger)@ukaachen.de*

Abstract: Relevant literature of early language development is reviewed in the context of our neurocomputational model of speech acquisition. This literature confirms our hypothesis, that phonological knowledge acquisition depends on phonetic in combination with semantic learning. It has been shown that phonetic learning starts at birth, followed by semantic learning, which starts at around 5 months of age. With around 18 months of age the collection of phonological knowledge begins, which is evident from experiments in perceptive development: infants are able to discriminate two objects even if their corresponding speech sounds are similar. Evidence from experiments in productive development indicate that the frequency of producing words correctly increases if semantic knowledge is available in parallel. In consequence the extension of the training data set within our neurocomputational model is described in this paper.

1 Introduction

In this paper the perceptual and productive early language development will be reviewed. The goals of the present paper are (i) to introduce our neurocomputational model of speech acquisition and (ii) to give a guideline for improving its sets of training data by reviewing relevant literature on perceptive and productive early language acquisition. This literature underlines the hypothesis that phonetic and semantic learning are precursors of phonological knowledge acquisition. Thus far an action repository including feed forward control and feedback control loops has been implemented in our neurocomputational model of speech acquisition. To simulate early language acquisition so far phonological representations are trained in parallel to phonetic (i.e. sensorimotor) representations. Because we assume that phonological knowledge acquisition is a consequence of phonetic in combination with semantic learning, we hypothesise that beside phonetic information in addition semantic information has to be trained in our model, in order to establish phonological meaningful contrasts (phonemic representations). We will describe our neurocomputational model and its revised training data.

2 The first two years of early language development comprising phonetic, semantic and phonological aspects

2.1 Perceptual development

Phonetic learning

In experiments it has been shown, that infants are already capable of distinguishing different speech sounds (phonetic units) both in native and in nonnative languages by birth [4, 17]. This is termed as **categorical perception** (e.g. [16]). This skill will be lost from ten to twelve months in favor of discriminating native speech sounds, i.e. sounds of the target language, which will be required by the child. That means that infants are no longer capable of discriminating non-native sound contrasts [28]. E.g. Japanese infants are not able to distinguish /l/ and /r/ whereas German or English infants can [24]. This process can be labeled as **phonetic learning**.

Learning to segment words from continuous speech

In language acquisition one important step is learning to segment words from continuous speech. Prosodic as well as phonetic and phonotactic as well as statistical features promote word-segmentation [11]. To segment the speech stream into words, infants have to pay attention to speech. Growing up to four months suprasegmental differences become important. Now infants pay more attention to infant directed speech items like “ah” as well as words like “lovely” etc. [6] (Examples are given here for American English-acquiring infants). Now word-segmentation begins on the basis of prosodic cues [9]. The ability to break the speech stream into meaningful units by different language specific pattern, e.g. stress pattern of a native language, improves between six and nine months of age [9, 11]. At around eight months statistical learning, which is a sort of analysis of the frequency of occurrence, distribution and further statistical properties in speech, begins as well [22]. In consequence infants notice e.g. regularities in the co-occurrence of speech sound sequences and realize word boundaries on the basis of this information. For example on the speech item [beɪ] often follows [bi:] so that “baby” will be a word in an utterance like “pretty baby” whereas “tyba” will not be a word, because the probability that [ti:] is followed by [beɪ] is low [22]. Moreover with nine months infants become more and more sensitive to different phonotactic patterns, i.e. the rules of permissible vs. forbidden combinations of speech sounds in the native language. Thus, nonnative speech items can be differentiated from native speech items on the basis of phonetic and phonotactic patterns even without prosodic features [7, 10]. For example the speech item [ʃtrʊmpf] (German “Strumpf”) will not be perceived as an English word because [ʃtr] is not a permissible combination in English. A speech item like [fɜːstkɪs] will be segmented in “first” and “kiss” because [stk] is not a permissible combination in English. Thus during early language development, infants learn which language specific features are important.

Semantic Learning

Beside phonetic and statistical learning, **semantic learning** is important during early language development. 5-month-old infants already prefer listening to their own name compared to other infants’ names. This seems to be the result of responding to frequently occurring items as prerequisite for semantic learning, i.e. relating sounds to meanings [18]. With eight months infants begin to understand the meaning of certain phrases, like “come here”, without knowing the meaning of the single words [5]. Referring objects to words increases at the age of ten to eleven months. The receptive vocabulary contains 11 to 154 words at this age [5]. At 14 months associating words to objects increases further, so that infants are capable of distinguishing newly learned words associated with objects if they are not sounding similar [26]. Here the attention

is based on connecting objects to language without phonological knowledge [27]. At the age of 16 months the receptive vocabulary size is 92 to 321 words [5].

Phonological Learning

From now on collection of **phonological knowledge** becomes more important. At the age of 18 to 20 months infants distinguish newly learned words even if they are phonologically similar [27]. Furthermore infants now move from a word-based to a segment-based phonological system [23] and **fast mapping** begins, i.e. the ability to learn and retain a partial knowledge of the meaning of new words by establishing an initial link between word and referent [3, 8]. Infants more and more learn the sound distinctions of their native language and learn the associated semantic contrast with those different sounds, i.e. phonological learning.

2.2 Productive development

During the productive language acquisition different stages between birth and 18 months of age are explored by [20]. From birth on to two months of age infants make **reflexive unconscious vocalizations**, i.e. different vegetative sounds like burp and sneeze. Further they make quasi-resonant sounds and sounds demonstrating discomfort like sustained crying or fussing. In the period of one to four months infants have an increasing control about phonation which results in more **speech-like phonation** of different single or multiple fully-resonant nuclei with vowel-like sounds longer in duration and sometimes combined with consonant-like segments. There are isolated consonant-like segments, glottal stops and glottal fricatives, which are still not comparable with adult sounds, too. Moreover infants now are capable to produce consonant-vowel-like sound combinations and reduplications, or at least two consonant-like segments in a series. At this point of development chuckles and laughter occur more frequently. Between three and eight months **expansion** takes place [20]. Within the period of expansion infants continue with cooing and laughing. Further there are more adult-like vowels isolated and in a row. There are more occurrences of vocal glides as well as of ingressive sounds, squeals and marginal babbling. From five until ten months production of basic canonical syllables begins. This period is also called reduplicative or **canonical babbling**. Babbling consists of single consonant-vowel constructions, consonant-vowel-consonant combinations (a silent gap between vowel and last consonant), and repeated syllables, such as [ba] and [gu:]. With nine months until 18 months of age the production becomes advanced forms, i.e. babbling changes into a **variegated babbling** period. This babbling consists of a mix of syllables, such as “ka-da-bu-ba” with varied stress and intonation patterns [20]. After twelve months the first **meaningful words** – simple in structure – are spoken that comprise the same sounds that were used in late babbling [25]. With 16 months of age the productive vocabulary size is around 50 words [5]. The consistence of producing these words becomes stronger with increasing association between semantics and phonetics. This occurs at the age of 18 months [25]. A further important step in productive development, which begins around 18 months, is the **vocabulary spurt** (e.g. [1]). From this point on infants learn 10 to 20 words a week [2].

3 Phonological contrast is based on phonetic and semantic learning

In our approach we assume that **phonological contrast** emerges on the basis of increasing association of phonetic and semantic information. From early language development as described before, it is clear that the first language skill of new born infants is phonetic learning [4, 17]. The next important ability is semantic learning, i.e. the development of item categories, which

is rudimentarily starting at five months of age [18]. Only afterwards phonological learning begins at around 18 months of age [27].

In consequence phonological categories are established. This is shown once more in [19, 27]. They determine that it is still difficult for 14-month-old infants to discriminate two objects corresponding to words if these words are similar on the phonetic level. But this promotes a challenging of phonetic abilities so that at around 18 months this task will be solved successfully [19, 27], i.e. phonological knowledge acquisition. During the productive language development this is shown when producing words correctly becomes better with more semantic reference at around 18 months as well [25].

Especially this is relevant for acquiring phonological categories and confirms our hypothesis that collection of phonological knowledge depends on phonetics in combination with semantics. In Figure 1 a broader view considering the influences of direct face-to-face communication for the process of acquisition is proposed. Within the triangular circuit caregiver and infant focus on an object. The infant listens to the infant-directed speech and in parallel follows the gaze and gestures of the caregiver. By this, the infant is capable of linking an articulated sound to an object. In consequence an auditory and semantic representation is acquired. And by imitating the caregivers articulated sounds, a motor representation is acquired as well (see also our model, Fig. 2). The bijective link between the sound chain of a word and an object, i.e. that exactly this sound chain means this object and this object is referred to as this sound chain, promotes the early phonological development. That is relevant to our neurocomputational model and its training data.

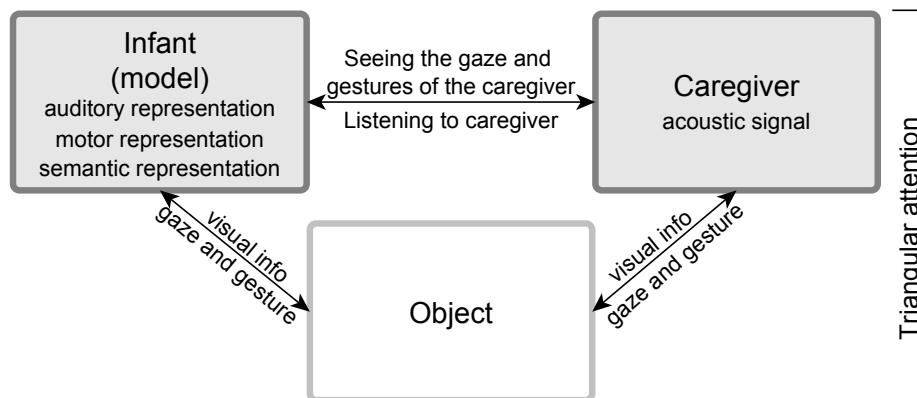


Figure 1 - Acquiring auditory, motor and semantic representation by communicating over an object in social context considering the triangular circuit. The infant (model) listens to the caregiver and imitates the acoustic signal. The infant (model) sees gaze and gesture of the caregiver and refers the acoustic signal to the object.

4 The neurocomputational model of speech acquisition

4.1 Organization of the model

The computer-implemented neurophonetic model of speech acquisition (Fig. 2) was introduced by [13, 14]. It consists of different **processing modules**, i.e. the semantic processing, the mental lexicon and the action repository, and two lower level processing modules, i.e. the articulatory processing, comprising the neuromuscular programming and execution and the articulatory acoustic model, and the sensory processing ending up in sensory short term representations (i.e. external sensory states, Fig. 2). Further there are two different types of maps:

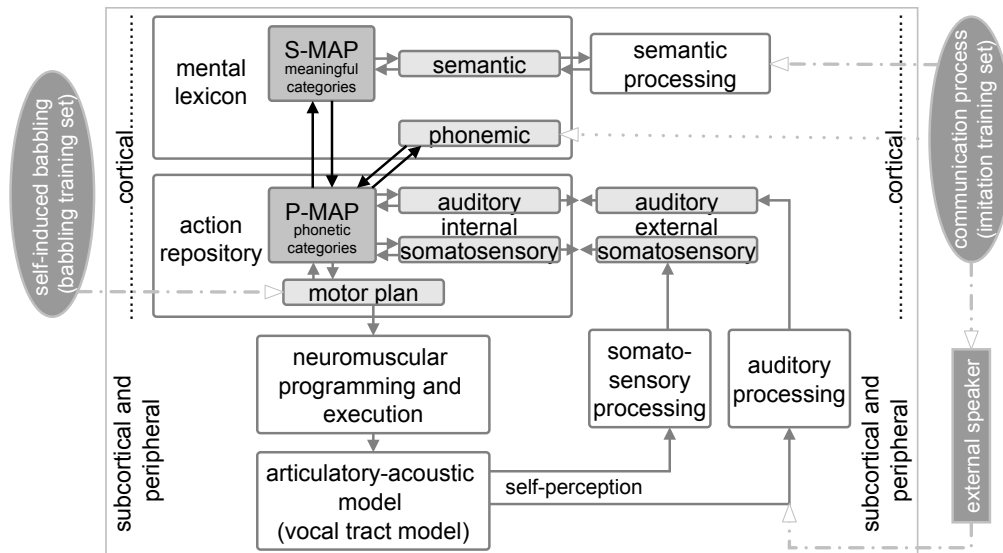


Figure 2 - Structure of our neurocomputational model of speech acquisition. White boxes indicate processing modules; grey boxes indicate self-organizing maps (S-MAP and P-MAP) and neural state maps, i.e. semantic, phonemic, auditory, somatosensory and motor plan state map. Grey blobs represent knowledge, which is learned by the model (reproduced from [13], p. 288)

self-organizing maps (SOM) and state maps. Self-organizing maps are the semantic map (S-Map) as central layer of the mental lexicon and the phonetic map (P-Map) as central layer of the action repository. These SOMs are interconnected by synaptic link weights with each other and with different state maps, e.g. internal auditory state map (Fig. 2 and Fig. 3). The SOMs and their link weights are part of the long-term-memory. They are capable of representing specific states, i.e. words and their meanings within the S-Map and sensory and motor information within the P-Map by local punctual neuron activations.

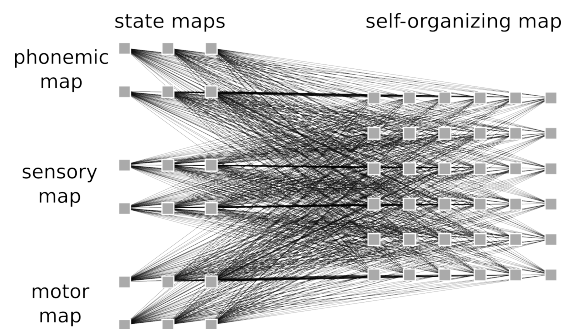


Figure 3 - Example of a self-organizing network. Grey squares = neuron collectives (i.e. neural maps). Black lines = neural connections between domain-specific state maps and a SOM (here: P-Map) (reproduced from [15], p. 797)

The state maps are part of the short-term memory that are the semantic, phonemic, auditory and somatosensory state map as well as the motor plan state map. The SOMs and state maps comprise ensembles of spatially closely connected model neurons (Fig. 3). By activating one neuron within the P-Map or S-Map (local activation) link weights are more or less activated and lead to an activation pattern of all neurons within a state map, which represents a word's or

syllable's, e.g. auditory, state (Fig. 3). The neural representation of e.g. an auditory state within the auditory state map can be assumed to be a neural representation of a bark-scaled acoustic spectrogram [13, 14].

4.2 Extended training and training data

For simulating early stages of language acquisition, different training data sets are needed. To train the self-organizing P-Map a sensorimotor training data set is needed. Each speech item is encoded as motor plan state (Fig. 4 left), encoded as auditory state (Fig. 4 middle), and encoded as phonemic state (Fig. 4 right). This combination of information has been trained by the model, associated with each other and finally formed the content of the P-Map [12]. So far no semantic information has been used.

From literature we know, that phonological knowledge emerges by referring phonetic information to semantic information. In consequence, the P-Map should be trained without a phonemic representation, because at this developmental stage, phonological information is not available. Thus, we need different training data sets, which comprise semantic information for training a further SOM, i.e. the S-Map (Fig. 2). Using these training data sets, semantic states will be encoded in addition; e.g. the concept "Mommy" (S-Map) will be associated with "woman", "loving", "caring", etc. (semantic state map in Fig. 2). These associations or semantic states will be trained, in order to form concepts, e.g. "Mommy", within the S-Map. Thus, **phonetic categories** are built-up at the P-Map level [12, 13] and **semantic categories** at the S-Map level [21]. Furthermore, as a consequence of P- and S-Map interaction, phonemic states (i.e. **meaningful contrast**) occur and allow the emergence of a phonemic state map.

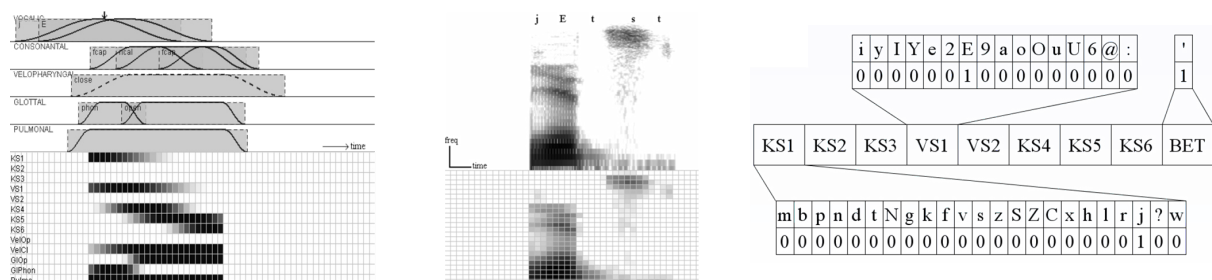


Figure 4 - Neuronal encoding of the training data for the phonetic self-organizing map (P-Map): motor plan representation (left), auditory representation (middle), and phonemic representation (right) for the syllable [jetst] (reproduced from [12], pp. 156–158)

5 Conclusions

By sifting the literature we concluded that phonological knowledge emerges by referring phonetics to semantics. In connection with objects, actions or situations (e.g. mom says: "come here!" and in parallel moves her palm in her direction) speech items become meaningful, and phonological categories emerge. To provide for a realistic simulation of the early language including the phonological development, our training data sets need to be extended. Beside phonetic information (i.e. sensorimotor states), that is summarized and stored within a P-Map, semantic information (i.e. semantic states) has to be trained in order to organize a S-Map as well. By the bidirected link between S-Map and P-Map the content of the S-Map will lead to an adaptation of the P-Map. Further this bidirected link promotes the assembly of a phonemic state map that comprises segments of meaningful contrast, i.e. phonemic representations or phonological knowledge. The S-Map therefore is a crucial part of the early language acquisition of

our model. In further studies this important step to phonological knowledge acquisition will be implemented.

References

- [1] BENEDICT, H.: *Early lexical development: comprehension and production*. J. Child. Lang., 6:183–200, 1973.
- [2] BERK, L.: *Child development*. Allyn and Bacon, Boston, MA, 2003.
- [3] CAREY, S. and E. BARTLETT: *Acquiring a Single New Word*. Papers and Reports on Child Language Development, 15:17–29, 1978.
- [4] EIMAS, P. D., E. R. SIQUELAND, P. JUSCZYK and J. VIGORITO: *Speech Perception in Infants*. Science, 171:303–306, 1971.
- [5] FENSON, L., P. DALE, J. REZNICK, E. BATES, D. THAL and S. PETHICK: *Variability in early communicative development*. Monographs of the Society for Research in Child Development, 59, 1994.
- [6] FERNALD, A. and P. KUHL: *Acoustic Determinants of Infant Preference for Motherese Speech*. Infant Behaviour and Development, 10:279–293, 1987.
- [7] FRIEDERICI, A. D. and J. M. I. WESSELS: *Phonotactic knowledge of word boundaries and its use in infant speech perception*. Perception and Psychophysics, 54:287–295, 1993.
- [8] HEIBECK, T. and E. MARKMAN: *Word learning in children: An examination of fast mapping*. Child Development, 58:1021–1034, 1987.
- [9] JUSCZYK, P. W., A. CUTLER and N. REDANZ: *Infants' Preference for the Predominant Stress Patterns of English Words*. Child Development, 64:675–687, 1993a.
- [10] JUSCZYK, P. W., A. D. FRIEDERICI, J. WESSELS, V. Y. SVENKERUD and A. M. JUSCZYK: *Infants' sensitivity to the sound patterns of native language words*. Journal of Memory and Language, 32:402–420, 1993b.
- [11] JUSCZYK, P. W., D. M. HOUSTON and M. NEWSOME: *The Beginnings of Word Segmentation in English-Learning Infants*. Cognitive Psychology, 39:159–207, 1999.
- [12] KANNAMPHUZA, J., C. ECKERS and B. J. KRÖGER: *Training einer sich selbst organisierenden Karte im neurobiologischen Sprachverarbeitungsmodell MSYL*. In KRÖGER, B. J. and P. BIRKHOLZ (eds.): *Elektronische Sprachsignalverarbeitung 2011*, pp. 154–163, Dresden, 2011. TUDPress.
- [13] KRÖGER, B. J., P. BIRKHOLZ, J. KANNAMPUZHA and C. NEUSCHAEFER-RUBE: *Towards the acquisition of a sensorimotorvocal tract action repository within a neural model of speech processing*. In ESPOSITO, A., A. VINCIARELLI, K. VICSI, C. PELACHAUD and A. NIJHOLT (eds.): *Communication and Enactment*, pp. 287–293, Berlin, 2011a. Springer.
- [14] KRÖGER, B. J., P. BIRKHOLZ and C. NEUSCHAEFER-RUBE: *Towards an Articulation-Based Developmental Robotics Approach for Word Processing in Face-to-Face Communication*. PALADYN Journal of Behavioral Robotics, 2:82–93, 2011b.

- [15] KRÖGER, B. J., KANNAMPUZHA and C. NEUSCHAEFER-RUBE: *Towards a neurocomputational model of speech production and perception*. Speech Communication, 51:793–809, 2009.
- [16] KUHLMAN, P. K.: *Early Language Acquisition: Cracking the Speech Code*. Nature Reviews, 5:831–843, 2004.
- [17] KUHLMAN, P. K., E. STEVENS, A. HAYASHI, S. K. DEGUCHI and P. IVERSON: *Infants show a facilitation effect for native language phonetic perception between 6 and 12 months*. Developmental Science, 9:F13–F21, 2006.
- [18] MANDEL, D. R., P. W. JUSZYK and D. B. PISONI: *Infants' Recognition of the Sound Patterns of Their Own Names*. Psychological Science, 6:314–317, 1995.
- [19] MILLS, D. L., C. PRAT, R. ZANGL, C. L. STAGER, H. J. NEVILLE and J. F. WERKER: *Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds*. Journal of Cognitive Neuroscience, 16:1452–1464, 2004.
- [20] NATHANI, S., D. J. ERTMER and R. E. STARK: *Assessing vocal development in infants and toddlers*. Clinical Linguistics and Phonetics, 20:351–369, 2006.
- [21] RITTER, H. and T. KOHONEN: *Self-Organizing Semantic Maps*. Biological cybernetics, 61:241–254, 1989.
- [22] SAFFRAN, J. R.: *Statistical Language Learning: Mechanisms and Constraints*. Current Directions in Psychological Science, 12:110–114, 2003.
- [23] SWINGLEY, D. and R. N. ASLIN: *Spoken word recognition and lexical representation in very young children*. Cognition, 76:147–166, 2000.
- [24] TSUSHIMA, T., O. TAKIZAWA, M. SASAKI, S. SHIRAKI, S. SHIRAKI, K. NISHI, M. KOHNO, P. MENYUK and C. T. BEST: *Discrimination of English /r-l/ and /w- j/ by Japanese infants at 6 – 12 months: Language-specific developmental changes in speech perception abilities*. Paper presented at the International Conference on Spoken Language Processing, 1994.
- [25] VIHMAN, M. M.: *Phonological Development: The Origins of Language in the Child*. Blackwell, Oxford, 1996.
- [26] WERKER, J. F., L. B. COHEN, V. L. LLOYD, M. CASASOLA and C. L. STAGER: *Acquisition of word-object associations by 14-month-old infants*. Developmental Psychology, 34:1289–1309, 1998.
- [27] WERKER, J. F., C. T. FENNELL, K. M. CORCORAN and C. L. STAGER: *Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size*. Infancy, 3:1–30, 2002.
- [28] WERKER, J. F. and R. C. TEES: *Cross-Language Speech Perception: Evidence for Perceptual Reorganisation During the First Year of Life*. Infant Behavior and Development, 7:49–63, 1984.