Computational model of the process of learning vowel categories in language acquisition

MIYAZAWA Kouki  
Graduate School of Human Sciences, Waseda University  
Tokorozawa, Saitama, Mikajima 2-579-15  
m-kouki@moegi.waseda.jp  

KIKUCHI Hideaki  
Faculty of Human Sciences, Waseda University  
Tokorozawa, Saitama, Mikajima 2-579-15  
kikuchi@waseda.jp  

SHIROSE Ayako  
Tokyo Gakugei University  
Tokyo, Koganei, Nukuikita 4-1-1  
shirose@u-gakugei.ac.jp  

Abstract  
We examine how infants organize phonetic categories by using neural network model.  

In this study, we consider Self-Organizing Maps (SOM) model which was based on knowledge how infants organize their vowel categories. Our model learns values of Mel Frequency Cepstrum Coefficient (MFCC) of a natural continuation utterance from multi corpora and organizes a learning process of the vowel categories. This result demonstrated that this model could acquire number and boundaries of the vowel categories of an input language, and its learning process was achieved with a little quantity of data.  

Then, these results show the possibility that statistical frequency of the specific sound properties input by neighbouring adults may play an important role in language acquisition.

1 Introduction  
The early language acquisition doesn’t need explicit instruction and anyone can acquire vowel systems precisely and naturally. Moreover, anyone can acquire a phonetic system of his/her native language, and the speech that infants hear has few kinds and vocabulary is limited. This is one of the superior human abilities and difficult to be realized by technology of speech processing.  

Our approach is to build a computational model of language acquisition and examining some knowledge of human’s capacities to discriminate phoneme contrasts. So we aim at the realization of the speech recognition technology by the unsupervised learning.  

We consider dividing the process of language acquisition into ‘innate mechanism’ and ‘language experience’. Figure 1 is a diagram of our model’s performance. Our model optimizes for a specific phonetic system by the specific language input. We try to show how and how much information in the speech signals affect to the acquisition of a native language by this model.

Figure 1: Diagram of our model’s performance

2 Recent Studies  
2.1 Categorical Perception  
There is one of the human capacities known as "categorical perception." Figure 2 is a diagram of categorical perception. Perception changes suddenly when stimuli crosses the dashed line, and this is called categorical perception.

Figure 2: Categorical perception (Jusczyk, 2000)
For example, American adults can discriminate two liquid consonants contrast, /l/ and /r/, but it is difficult for non-native speakers of English, such as Japanese adults.

Interestingly, it is told that the 3-month-infants have the capacity to discriminate contrasts in any of the world's languages. In other words, the Japanese infants can discriminate the English /l/ and /r/ contrasts but Japanese infants lose their abilities to discriminate these contrasts gradually.

2.2 Development of phonetic organization

Infants gradually acquire phonetic systems of native language. Figure 3 is the process of infant's phonetic organization. A cross axis is the age of month of the infants. A vertical axis is number of the phonetic categories which the infants can discriminate. Newborns can discriminate in any of the phonetic categories. After that, 6-month-infants can discriminate only native vowel categories. And 9 month of age can discriminate native consonant categories. The pre-babbling infants can discriminate their vowel categories, so about the acquisition of the vowel categories, the feedback of one's speech is needless.

The native phonetic systems are acquired within first one year. In a process of the acquisition, phonetic categories are unified and category boundaries are shifted (Jusczyk, 2000).

2.3 Mechanism of the language acquisition

There is a hypothesis that language experience sets the perceptual boundaries. Nevertheless, there are indications that human had specific linguistic abilities (Kuhl, 2004). This is the question, whether the innate biased Mechanism optimized for a specific language by the experience or not.

The innate mechanism is the specific linguistic abilities which were decided by a gene. For example, chinchillas, a kind of rats, also showed categorical perception as well. So properties in the mammalian auditory system may be innate mechanism (Kuhl, 1975).

The experience is a exposure to speech information after birth. Statistical frequency of the specific sound properties input by adults may play an important role in language acquisition.

3 The Model

3.1 Self-Organizing Maps (SOM)

SOM is a kind of the neural network model which reflects the information processing that the sensory area performs. SOM can classify input signals without instruction and estimating the categories.

So, we suppose that SOM is adequate as language acquisition model of the phonetic system. We try to simulate the cognitive ability of the infants. Concretely, we investigate relationship between, quantity of learning and accuracy in some languages. In this study, we introduce Terashima’s method of unification of the categories (Terashima, 1996). By using this method, we can unify the categories, so that we can simulate Categorical Perception by SOM.

We illustrate this method with examples. We consider classifying many kinds of Gaussian distribution. A number and boundaries of this categories are unknown. Figure 4-1 is an example of results of the learning by SOM.
Figure 4-2 is a result of SOM learning, and these categories can be unified by the Terashima’s method using data density histogram on SOM. Categories are unified and extracted three Gaussian distribution. But, the biological validity needs more discussion.

3.2 Example of the SOM learning

Figure 5 is an example of consonant categories learning. Input data of SOM are values of Voice Onset-Time, called VOT of two Voiced-Voiceless consonants, /d/ and /t/. We referred to data of VOT frequency, it was pronounced by American speakers. Our results demonstrated that SOM could categorize into two groups, /d/ and /t/, and the value of the categorical boundary, showed a good fitness to human data (Miyazawa, 2006).

For SOM which finished learning, we estimate the number and boundaries of categories. For Japanese data, we use phoneme labels offered in the Corpus of Spontaneous Japanese, called CSJ (Maekawa, 2003) and calculate the accuracy rate. Also, we can use reference data from the corpus (Figure 6-2). Then, accuracy of the result will be evaluated.

Table 1 is details of the language sounds that were used in our experiment. We use Japanese and Spanish data for the experiment. For Spanish data, we use the OGI multi-language telephone speech corpus (Yeshwant, 1992).

<table>
<thead>
<tr>
<th>Language</th>
<th>Vowels</th>
<th>Corpus</th>
<th>Details of sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>a, e, i, o, u</td>
<td>CSJ</td>
<td>Re-reading of speech, male and female</td>
</tr>
<tr>
<td>Spanish</td>
<td>a, e, i, o, u</td>
<td>CSLU (telephone speech)</td>
<td>Reading of story, male</td>
</tr>
</tbody>
</table>

Table 1: Details of the languages
The results of SOM learning are as follows. Table 2-1 shows relations between quantity of learning and number of categories. Table 2-2 shows the accuracy rates. These results demonstrate a high accuracy for Japanese vowel categories by the short-time learning. But, effects of the learning differ in each speaker.

Consonants which have stationary characteristics tend to form the independent categories.

Then, we use Spanish for an experiment. It is because the vowels of Spanish are usually pronounced clearly, and Spanish has five vowels resemble Japanese. The SOM acquires the vowel categories, like Japanese with a little quantity data.

<table>
<thead>
<tr>
<th>Language</th>
<th>Learning Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td>2.8</td>
</tr>
<tr>
<td>(female)</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>2.5</td>
</tr>
<tr>
<td>(male)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 2-1: Number of categories

<table>
<thead>
<tr>
<th>Language</th>
<th>Learning Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td>20</td>
</tr>
<tr>
<td>(female)</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>20</td>
</tr>
<tr>
<td>(male)</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 2-2: Accuracy rate [%]

5 Conclusion and Future Works

Our model can acquire number and boundaries of the vowel categories of an input language, and its learning process was achieved with a little quantity data. So our model shows that not only the Japanese but also other languages could show a high accuracy for vowel categories. Then, these results show the possibility that language experience and self-organizing learning process play an important role in the language acquisition. But, effect of the learning differs in each speaker. This is a future problem, we try to consider infants’ innate mechanism and process of cognitive development.

We will try to improve accuracy, of the language which is different in the number of vowels. So, we will try to choose parameters, and time window length, suitable for learning of multi languages. Then, we will also improve our model to reflect infants’ cognitive ability.

References


