SELF-ORGANIZATION IN SOUND SYSTEMS: A MODEL OF SOUND STRINGS PROCESSING AGENTS

ROLAND MÜHLENBERND, JOHANNES WAHLE
Seminar für Sprachwissenschaft, University of Tübingen
Tübingen, Germany
{*roland.muehlenbernd, johannes.wahle*}*@uni-tuebingen.de*

Several typological universals of sound systems of human language are assumed to be a result of self-organization in a population's communication performance. This has been shown for human vowel system in diverse studies (c.f. de Boer, 2000b; Jäger, 2008). In these studies computational models were designed of agents communicating with single vowel sounds. In our study we present a computational model where agents communicate with concatenations of single sounds forming complex expressions. The goals of this study are i) to examine decisive factors that contribute to the emergence of realistic sound systems in artificial societies of interacting agents, and ii) to discuss ways to evaluate artificially emerged sound systems.

1. Introduction

Human sound systems reveal various global regularities on different levels. There are implicational universals describing the composition of sound inventories (c.f. Maddieson, 1984; Plank & Filimonova, 2000; Hyman, 2008) and particularly the structure of human vowel inventories (Schwartz, Boë, Vallèe, & Abry, 1997). There are also universal tendencies characterizing the structure of syllables (c.f. Bell & Hooper, 1978; Vennemann, 1988; Hammond, 1997). Finally, the existence of so-called *cross-linguistic phoneme correspondences* (Tiberius & Cahill, 2000) is assumed, i.e. regularities of the realization of the phonemic inventory, particularly the relationship of allophones and phonems. A number of studies addresses a synthetic approach – a simulation model of interacting agents – to show that a number of regularities might not be a consequence of an innate human disposition, but the result of *self-organization processes* of interacting individuals (c.f. Lindblom & Maddieson, 1988; de Boer, 2000b; Jäger, 2008). In line with Blevins (2006), we address the following research question: *Can self-organization explain particular universal tendencies of human sound systems and syllable structures?* Additionally, we are interested in the application of procedures for the evaluation of synthetically emerged sound systems.

2. The Simulation Model

Simulation models proved themselves as a valuable technique to gain insights into the way some properties of human language emerge, such as compositionality (c.f. Nowak & Krakauer, 1999), complex syntactic patterns (c.f. Kirby & Hurford, 2002), or regular patterns in syllable structure (c.f. de Boer, 1997, 2000a), and sound inventory (c.f. Jäger, 2008). One essential goal of such studies is to show that the presence of particular features can be explained by functional factors – such as self-organization effects in communication (Lindblom & Maddieson, 1988; de Boer, 2000b) – without the need for postulating innate dispositions.

Following this line of research, we want to create a model of interacting agents to study the concurrent emergence of i) a sound inventory and ii) a lexicon. We apply a model framework that has proven fruitful for studying the emergence of vowel inventories: the *imitation game* model (de Boer, 2000a, 2000b).[a] Furthermore, we want to analyze how realistic particular aspects of the emergent language systems are. We evaluate the quality by two factors: i) the syllable structure of the agents' lexical entries, and ii) the composition of the agents' sound inventories.

2.1. *The Model Conception*

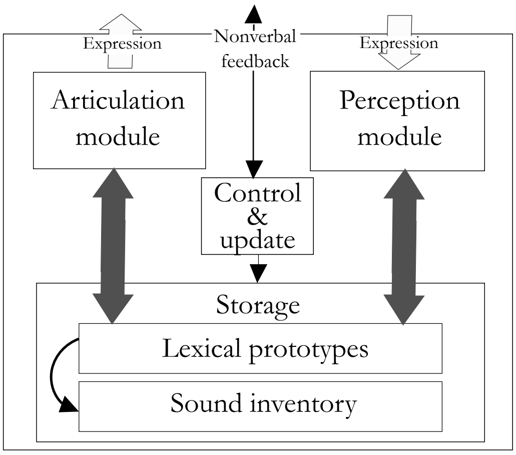
The conception of our model is an altered version of de Boer's (2000b) imitation game model. It is altered in two main aspects:

1. Agents produce expressions instead of vowels. As a consequence they have two different types of inventories: i) a sound inventory of the sounds used to build expressions; ii) prototypes of expressions the agent produces.
2. Instead of having a very exact model of phonetic information, such as an articulatory and perceptual space, we abstract away from this level of detail by taking phonetic codes as acoustic units.

We adopted most of the other elements of de Boer's model, i.e. a similar agent architecture and an imitation game as interaction protocol (see Figure 1). To adapt this infrastructure to a communication model of complex expressions, we had to answer a number of design questions, of which three were most essential:

1. *What is the phonetic code of the sound inventory?* We used the ASJP code (c.f. Wichmann et al., 2013) (41 symbols), since it allows us to utilize real language data from the ASJP data base for evaluation. Subsequently, we reduced this set to 35 symbols, which enables us to integrate acoustic data from Mielke (2012) for setting up an agent's *production bias*.

[a] In his study, de Boer has shown that the prototypical structures of human vowel systems can be reproduced as a result of self-organization effects of agents that produce and imitate vowel sounds. In subsequent work, de Boer (2000a) also proposed a more realistic model by assuming agents to interact via more complex expressions.



Interaction protocol:

1. S chooses randomly $p \in L$ and produces $e = pb(p)$
2. R receives e and selects p', with $dis(p', e) = \min_{q \in L} dis(q, e)$

   R produces $e' = pb(p')$
3. S receives e' and selects p'', with $dis(p'', e) = \min_{q \in L} dis(q, e')$

   if $p'' = p$, S signals success, else failure as feedback
4. R receives feedback: if success, shift prototype closer to input expression, else add new prototype
5. both agents update L and X

Figure 1. **Left**: the architecture of an agent. The *storage* module entails a) lexical prototypes of expressions L that is a subset of all possible expressions E, and b) the sound inventory X that entails exclusively sounds that occur in at least one lexical prototype. The *articulation module* realizes the production bias function $pb : E \to E$ which alters an expression $e \in E$ in a speaker biased way. The *perception module* is a search engine that finds the most similar prototype $p \in L$ for a received expression e. The control unit is informed by feedback signals and regulates the update of the prototypes, also by access to the sound inventory. **Right**: the interaction protocol of a communication step. The sender S chooses a prototype p from her storage and produces an expression e chosen from her prototypes and altered by production bias. If the storage of lexical prototypes is empty, a random expression is generated. The receiver R receives e, searches for the most similar prototype in her lexicon, and produces it (e'). S does the same and compares the new with the old prototype. If the prototypes are the same, S gives a feedback for success, otherwise for failure. Prototypes are updated according to the feedback: if success, R shifts her prototype closer to the input expression; if failure, either R adds the input expression to her prototype storage or S deletes prototype p, depending on p's former communicative success. Finally, both agents update their storages (see de Boer, 2000a).

2. *How is the similarity between expressions computed?* To deal with strings of ASJP code, we used a *weighted alignment algorithm* (Jäger, 2013) to compute the distance between whole expressions.
3. *How is a random expression generated?* A random expression is generated by concatenating n randomly chosen symbols, whereby n itself is a random number between 1 and 10.

Note, that a realistic communication model has to entail *noisy signals* such that the produced signal does not totally resemble the intended prototype. To implement unbiased random noise seems to be an unrealistic assumption, since any alteration of the agent's production process of a complex expressions must be influenced by context effects. In this sense, we implemented a tendency of agents to produce a particular alteration to an expression as *production bias*: a probability function that determines the probability of each sound x of an expression to be replaced by another sound x' in the production process, defined as follows:

Definition 1 (Production Bias) *Let X be an alphabet of single sounds and let $e = x_1 x_2 \dots x_{n-1} x_n$ be an expression, whereby $n \in \mathbb{N}_{>0}$ and $x_i \in X$ is the symbol of expression e at position i, $1 \leq i \leq n$. The production bias pb_α for replacing x_i with x' is defined as follows:*

$$pb_\alpha(x_i \rightarrow x') = \frac{\alpha \times sim(x_i, x') + (1-\alpha) \times P(x'|x_{i-1}, x_{i+1})}{\sum\limits_{x'' \in X} (\alpha \times sim(x_i, x'') + (1-\alpha) \times P(x''|x_{i-1}, x_{i+1}))}$$

In a nutshell: the production bias defines the probability to replace a sound x_i in context $x_{i-1}x_ix_{i+1}$ with a sound x', which depends on a) how similar both sounds are to each other, thus $sim(x_i, x')$, and b) how conventional it is to position x' in such a context, given by probability $P(x'|x_{i-1}, x_{i+1})$. All in all, the production bias is defined as the weighted and normalized sum of similarity and conventionality. The free parameter α gives weight to similarity versus conventionality. The important issue in this part of the model is how to define i) similarity between two sounds $x, x' \in X$ and ii) conventionality of a sound being in a specific context.

For the similarity value we used a metric of production similarity that we gained from experimental data obtained by Mielke (2012). Based on different aspects such as nasal and oral airflow, larynx height and vocal fold contact area he computed similarities between sounds using a principal component analysis.

We define the conventionality of a sound in a given context as the probability of a sound x_i appearing between x_{i-1} and x_{i+1}, thus $P(x_i|x_{i-1}, x_{i+1})$, estimated on the basis of trigram frequencies in the ASJP data. The resulting counts were smoothed using an adapted version of Kneser-Ney-smoothing (Kneser & Ney, 1995).

Definition 2 (Conventionality) *Let X be an alphabet of single sounds and let $P_{ENV}(x_i)$ be the probability of x_i in the context of x_{i-1} and x_{i+1} this is $P_{ENV}(x_i) = \frac{|\{x_{i-1}, x_{i+1} : c(x_{i-1}, x_i, x_{i+1}) > 0\}|}{\sum_j^X |\{x_{j-1}, x_j, x_{j+1} : c(x_{j-1}, x_j, x_{j+1}) > 0\}|}$, where $c(z)$ is the plain count of z. Let $\lambda(x_{i-1}, x_{i+1})$ be $\lambda(x_{i-1}, x_{i+1}) = \frac{d}{c(x_{i-1}) + c(x_{i+1})} \cdot |\{z : c(x_{i-1}, z, x_{i+1}) > 0\}|$, then*

$$P(x_i|x_{i-1}, x_{i+1}) = \frac{max(c(x_{i-1}, x_i, x_{i+1}) - d, 0)}{c(x_{i-1}) + c(x_{i+1})} + \lambda(x_{i-1}, x_{i+1}) \cdot P_{ENV}(x_i)$$

The smoothing parameter d was calculated in accordance with Sundermeyer, Schlüter, and Ney (2011) $d = \frac{n_1}{n_1 + 2n_2}$, where n_1 is the number of observed trigrams whose count is one and n_2 is the number of trigrams whose count is two.

2.2. *Simulation Experiments*

To gain a lot of data in a reasonable time frame, we decided to conduct first experiments with a very small population of three agents. The gist of our investigation was to figure out how the implementation of the production bias influences the resulting lexicon and sound inventory. Thus, we made experiments with 4 different settings: 3 settings apply the production bias function pb as defined in Definition 1, each with different α parameters: 0.2, 0.5 and 0.8. The fourth setting does not involve a production bias, but a random alteration. In this setting randomly chosen sounds are occasionally replaced, deleted or added.

To neglect any assumption for a prior bias or disposition, each agent has an empty prototype storage and sound inventory at the beginning of a simulation run. With a probability of 0.01 an agent adds a new entry to her prototype storage. In each simulation step every agent communicates with every other agent according to the interaction protocol (Figure 1). As a primary result, in all runs the number of lexical prototypes and the size of the sound inventories both increased over time. Simulation runs were stopped after $4,000$ simulation steps, where the size of emerged lexicons (number of expression prototypes) and the size of the emerged sound inventories of the agents roughly corresponded with sizes of concept lists and sound inventories of the languages of the ASJP database.

2.3. *Evaluation*

The results of the simulation were evaluated with respect to two different aspects: i) to measure the characteristics of the emergent lexicons of the agents, we analyzed the syllable structure[b] of the resulting expressions (sound strings), and ii) to measure the quality of the emerged sound inventories of the agents, we analyzed the composition of sounds they entail. Therefore a list of rules for each of these aspects was compiled. These rules should represent universals regarding sound inventories or syllable structure. The rules were selected in such a way that they can be evaluated as true or false and can be answered within the provided framework. Some of the universals found in the literature could not be used for evaluation since they either focus on distinctions which are not measurable by using the ASJP code (rounded vs. unrounded vowels) or since the necessary information was not present in the system at all (minimal pairs). This results in 15 rules which were used to examine the sound inventory (Table 1) and five rules for the syllable structure (Table 2).

2.4. *Results*

We made 100 simulation runs for each setting (random alteration and α = 0.2, 0.5, 0.8) and analyzed the resulting sound inventories and syllable structures according to the rule set. To compare the results with empirical data, we applied the rules to the 6895 lexicons of the ASJP data base.

Figure 2 (left) depicts the results for the estimation of the sound inventories: the box plots over percentage values of satisfied rules (15 sound rules) for all

[b] We used a self-implemented syllable parser inspired by Brunson (1989).

Table 1. Examples of implemented rules to check the quality of the resulting sound inventories.

Nr	Rule	Source
01	All languages have place distinctions of high and low and of front and back in their vowel systems; hence vowel systems minimally include /i, a, u/	Plank and Filimonova (2000)
02	IF there are palatoalveolar consonants, THEN there are dental consonants.	
03	IF there are uvular stops, THEN there are velar stops.	
04	IF there is a glottal stop, THEN there must be a primary oral stop	
05	IF there is a voiceless palatal approximant, THEN there is also a voiceless labial-velar approximant	
06	IF there are fricatives, THEN there will be stops	
07	IF there are back consonants, THEN there will be front consonants.	
08	IF there is any other lateral, THEN there will be a voiced lateral approximant.	
09	All languages have a high or a lower high front vowel	
10	IF there is the voiceless labial fricative phoneme /f/, THEN there will be the voiced labial fricative phoneme /w-v/.	
11	Every phonological system contrasts phonemes which are [-cont] (= stops) with phonemes that are specified with a different feature.	Hyman (2008)
12	Every phonological system has coronal phonemes	
13	Every phonological system has at least one front vowel or palatal glide /y/.	
14	Every phonological system has at least one back vowel.	
15	Every phonological system has stops	

Table 2. Implemented rules to check the quality of the resulting syllables.

Nr	Rule	Source
R_1	Syllables must have onsets.	Hammond (1997)
R_2	Syllables can't have codas.	Hammond (1997)
	A syllable is more preferred, ...	
R_3	...the steadier speech sound is.	Vennemann (1988)
R_4	...the smaller the number of speech sounds in the coda.	Vennemann (1988)
R_5	...the closer the number of speech sounds is to one.	Vennemann (1988)

sound inventories per setting. The results show that the production bias produces inventories that are significantly closer to the ASJP data than random alternation, whereas alternation of the α-value does not have a significant impact. This result reveals that the production bias – although operating locally on single strings – affects the way human sound inventories are compounded. Nonetheless, the inventories that emerged with production bias have still significantly lower values that the ASJP data. As a next step, further assumption should be tested that might improve the results.

The evaluation of the syllable features did not reveal significant results. All four lines of experiments produced lexicons with syllable structure with similar quality, all lower than those of the ASJP data. Nevertheless, by analyzing the runs with the lowest values for each setting (worst case values), we found that random alternation generally produces the lowest values, whereas the intermediate production bias parameter $\alpha = 0.5$ guarantees the greatest lower boundary for almost all rules, as shown in Figure 2 (right). This result was not significant,

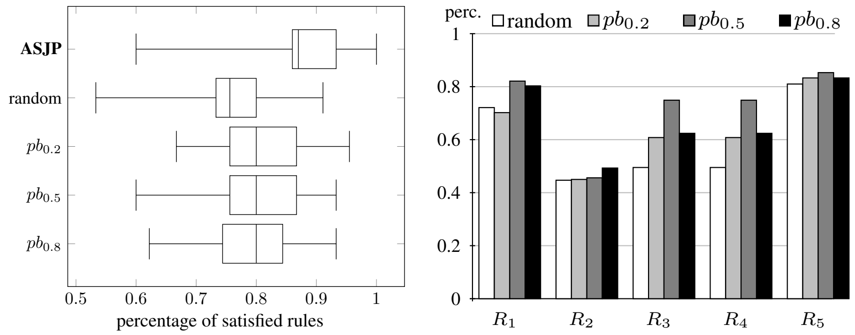


Figure 2. **Left:** result for the evaluation of the emerged sound inventories: box plots over percentage values of satisfied rules (15 sound rules, see Table 1). **Right:** result for the evaluation of the syllable features of the emergent lexicons: worst case values for each rule (see Table 2).

but might be a hint that an intermediate production bias prevents from producing syllables of low quality ($< .7$). Further investigations are necessary to test this suggestion and to detect more concretely the role of production bias in the process of syllable structuring.

All in all, the evaluation of the data reveals that a production bias supports the emergence of more realistic sound systems, though there is still room for improvement. Furthermore, a mid weighting (α-value) between similarity and conventionality for the production bias might increase the lower boundary of the quality of syllable structures.

2.5. *Discussion*

The results show that the some of the tendencies of syllable structure and sound inventory can be modeled by self-organization. The analysis of the resulting syllable structures is of course affected by the properties of the syllable parser. But since the syllable parser is only used for the analysis and not part of the simulation process it will not introduce any bias into the emergence of the syllable features itself. Nevertheless, a more sophisticated parser, e.g. statistical syllable parsing, may be necessary to better identify a syllable and classify its parts.[c]

Another important aspect of the experiments is the presence of multiple agents, even though the population is considerably small. An innovation resulting from the production bias can only survive if it succeeds in communication. Thus there are two forces at work here. On the one hand, agents strive to produce strings in accordance with the *production bias*, i.e. strings which are more common and easier to produce; on the other hand, *communicative success* requires a

[c] Such an approach demands large corpora to successfully train the parser, which poses a problem in itself in the cross-linguistic case shown here.

stable code, i.e. maintenance of a form established in the population.

3. Conclusion and Outlook

The goal of this article is twofold: the first goal is the implementation of a model that can explain universal features of human languages by means of self-organization. The essential extension to former models in this area are i) a virtual society interacts through complex expressions, and ii) the members' lexicon and sound inventory emerges in parallel. In our study we used an extended version of de Boer's *imitation game*. The second goal is to establish evaluation methodology for synthetically emerged language system. In this study we proposed a first approach by implementing a rule-based matching system for universal tendencies of human sound inventories and syllable structure. This is but the first step to the development of a more elaborated evaluation system. Using the described evaluation methodology, it has been shown that self-organization can explain some particular universal tendencies in human sound systems and syllable structure.

Further work involves the development of a more fine grained evaluation mechanism. This involves a more sufficient syllable analysis mechanism and a more sophisticated and larger set of rules. Another aspect for improvement should focus on the production bias and its parts. Supplementary studies should investigate the influence of different similarity measures and agent structures.

Acknowledgements

The authors have been supported by the ERC Advanced Grant 324246 EVOLAEMP - Language Evolution: the Empirical Turn (`http://www.evolaemp.uni-tuebingen.de`), which is gratefully acknowledged. The authors especially thank Gerhard Jäger, Armin Buch, Igor Yanovich, Christian Bentz, Johannes Dellert and Marisa Köllner for insightful discussions.

References

Bell, A., & Hooper, J. B. (1978). *Syllables and segments.* North-Holland Publ.

Blevins, J. (2006). *Evolutionary phonology: The emergence of sound patterns* (3 ed.). Cambridge: Cambridge University Press.

Brunson, B. (1989). Parsyl: A computer model of syllable parsing and acquisition. *Toronto Working Papers in Linguistics*, *10*, 3-20.

de Boer, B. (1997). Emergent cv-syllables. *Vrije Universiteit Brussel AI-lab AI-memo*.

de Boer, B. (2000a). Imitation games for complex utterances. *Proceedings of BNAIC*.

de Boer, B. (2000b). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465.

Hammond, M. (1997). *Parsing in OT* (Tech. Rep.). University of Arizona.

Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, *25*(1-2), 83–137.

Jäger, G. (2008). Applications of game theory in linguistics. *Language and Linguistics Compass*, *2/3*, 408–421.

Jäger, G. (2013). Phylogenetic inference from word lists: Using weighted alignment with empirically determined weights. *Language Dynamics and Change*, *3*, 245–291.

Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). Springer.

Kneser, R., & Ney, H. (1995). Improved backing-off for M-gram language modeling. In *International conference on acoustics, speech, and signal processing* (Vol. 1, p. 181-184). Detroit, MI.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In M. Larry, C. Hyman, & N. Li (Eds.), *Language, speech and mind* (pp. 62–78). Routledge.

Maddieson, I. (1984). *Patters of sounds.* Cambridge University Press.

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, 145–163.

Nowak, M. A., & Krakauer, D. (1999). The evolution of language. *PNAS*, *96*(14), 8028–8033.

Plank, F., & Filimonova, E. (2000). The universals archive: brief introduction for prospective users. *STUF - Language Typology and Universals*, *53*(1), 109-123.

Schwartz, J. L., Boë, L.-J., Vallèe, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, *25*, 233–253.

Sundermeyer, M., Schlüter, R., & Ney, H. (2011). On the estimation of discount parameters for language model smoothing. In *Interspeech* (p. 1433-1436). ISCA.

Tiberius, C., & Cahill, L. (2000). Incorporating metaphonemes in a multilingual lexicon. In *Proceedings of the 18th conference on computational linguistics - volume 2* (pp. 1126–1130). Stroudsburg, PA, USA: Association for Comutational Linguistics.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change : with special reference to german, germanic, italian, and latin.* Berlin: Mouton de Gruyter.

Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., Holman, E. W., Sauppe, S., Molochieva, Z., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, M., Mailhammer, R., Carrizo, A., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A., & Valenzuela, P. (2013). *The ASJP Database (version 16).*