Modeling the Emergence of Syllable Systems

Melissa Annette Redford (mredford@utxvms.cc.utexas.edu)
Department of Psychology
The University of Texas, Austin, TX 78751

Chun Chi Chen (ccchen@cs.utexas.edu)
Department of Computer Sciences
The University of Texas, Austin, TX 78712

Risto Miikkulainen (risto@cs.utexas.edu)
Department of Computer Sciences
The University of Texas, Austin, TX 78712

Abstract

In this paper we present an approach to modeling emergent syllable systems using simulated evolution of a "vocabulary" of "words." The model is aimed at testing the general hypothesis that language-universal sound patterns emerge from selection pressures exerted on the system by the perceptual and articulatory constraints of language users. The model is able to distinguish between hypotheses about how specific, biologically-motivated constraints affect the sound structure of language. For example, it is shown that mandibular oscillation provides a strong constraint on the sequential organization of phonemes into words. Future work will explore the potential of other constraints that, with mandibular oscillation, will be sufficient to describe the emergence of syllable systems.

Introduction

Although much variability exists in the sound structures of languages, there are a number of common patterns. Several independent statements can be made that capture these universal (or highly frequent) sound patterns of language. For example, the most common vowel sounds in the world's languages are [i], [u], and [a], and voiceless stop consonants [p], [t], [k] are the most common set of consonants (Maddieson, 1984). According to a number of phonetic theories, such patterns emerge from perceptual and articulatory constraints. This idea is inspired by the biological theory of evolution by natural selection. Just as the morphological structure of an organism reflects the types of selection pressures to which it has been exposed, the sound structure of language reflects the pressures originating from the perceptual and articulatory systems of language speakers and listeners. For example, Lindblom (1986) has demonstrated that the criteria of maximal perceptual distinctiveness is sufficient to predict the systematic occurrence of [i], [u], and [a] in vowel systems of different sizes. Similarly, the prevalence of the voiceless stop series can be understood as resulting from a pressure for sound systems to maximize ease of articulation (Willerman, 1994).

In these examples, a language-universal sound pattern is identified and explained as resulting from a constraint that is either perceptual or articulatory. Like in the specific examples provided above, the sound patterns that are usually explained are context-free segmental patterns. One might wonder whether these types of perceptual and articulatory explanations will also be sufficient to account for patterns that arise from the organization of segments into larger units such as syllables. MacNeilage (1998) has proposed an articulatory-based hypothesis to explain one aspect of super-segmental sound patterns. According to MacNeilage, the serial organization of consonants and vowels into syllables emerges naturally from the basic close-open cycle of the jaw that characterizes all of speech production. In MacNeilage's view, the close-open cycle is basic because it reflects the evolutionary history of speech production in which human ancestors imposed phonation onto a cycle that existed (and is still currently in use across mammals) for a variety of ingestive processes.

To test a hypothesis about emergence, instrumental (i.e. experimental) phonetic research, which identifies the perceptual or articulatory constraints, must be complemented with demonstrations of how sound patterns can emerge. One way to accomplish such demonstrations is via computational simulations. For example, it is possible to set up a model with an initially randomly organized vocabulary. The phonemic structure of words changes over time by a process of selection to satisfy the specific constraints of the system. The final result is a well-organized and regularly-structured vocabulary.

In the first half of this paper we describe such a model for syllable systems that are evolved under the selection pressure of specific perceptual and articulatory constraints. In the second half of the paper, we present results from simulations in which the emergence of one universal aspect of syllable structure, namely, the occurrence of CV syllables (Bell and Hooper, 1978), is modeled. Specifically, we test whether mandibular oscillation, as described by MacNeilage (1998), provides a better constraint on the serial organization of phonemes than either a simple random concatenation of phonemes or than a constraint that ensures maximal perceptual distinctiveness between words in a vocabulary. Our findings indicate that mandibular oscillation does provide a strong organizational constraint on the system, but that other constraints must also be involved. In conducting these simulations we aim to demonstrate the utility of this type of optimization model in testing biological, constraint-based hypotheses on the emergence of linguistic sound patterns.

It should be noted that in spite of the use of terminology such as "fitness" and "selection" the model is not a model of language evolution. The goal is to demonstrate that specific perceptual and articulatory constraints, which are hypothesized to create the regular syllable systems of language via selection, are in fact capable of doing so.

The Emergent Syllable System (ESS) Model

In order to simulate the emergence of syllable systems, it is necessary to simulate the emergence of a vocabulary of words. This is because the syllable system of a language can be determined from the organization of phonemes in the words of that language. Thus our model, which will be referred to as the Emergent Syllable Systems model, or ESS, simulates the emergence of a "vocabulary" of "words."

Architecture of Vocabulary Evolution

The ESS model is based on Symbiotic Evolution (Moriarty and Miikkulainen, 1996). In this method, genetic algorithms evolve a population of partial solutions that combine to yield an optimal solution to the given problem. In the present adaptation of this model a set of words are randomly generated to form vocabularies and the "fitness" of these vocabularies is evaluated according to a set of specified perceptual and articulatory constraints. The fitnesses of all vocabularies that a word participates in are averaged to get a fitness for the word. Words are then sorted according to their fitness values and a percentage of the best words are bred to create a new population of words. The subsequent generation of vocabularies is then randomly selected from this population, and so on. In this way, words that participate in successful vocabularies are more likely to reproduce and become highly represented in the population. Via this process the search space of potential vocabularies is explored until a vocabulary with high fitness, satisfying the constraints as well as possible, is discovered.

Word Representation

The vocabularies each consist of a set of words (25 in these simulations) that have been randomly drawn from the entire population of words. In the first generation, the words represent random concatenations of phonemes. These phoneme strings have a maximum length (set to 15 in these simulations), but words may be of any length under the maximum length. In the subsequent generations, a new population of words is produced by "interbreeding" the best words (i.e., words with the highest fitness) from the previous generation. Interbreeding takes place through crossover where a portion of one word is concatenated to a portion of another word to form an offspring word. The offspring is novel, but no longer randomly organized, since it contains parts of words that have been found to be highly fit. A fixed rate of mutation is also applied. For each phoneme, there is fixed probability (1%) that the phoneme will be replaced by another, randomly-selected phoneme. Mutation ensures that diversity is maintained in the population of words outside of the crossover operations.

Phoneme Representation

In the present simulations, the phonemes include [i, a, u, p, t, k, s, l, n]. These phonemes are encoded in the model along the following dimensions: (1) vowel height; (2) vowel front-back dimension; (3) vowel roundedness; (4) consonantal place of articulation; (5) consonantal manner of articulation; (6) consonantal voicing characteristics; (7) segment class (either consonant or vowel); and (8) jaw openness scores. Parameters (1) - (7) are the standard distinctive features proposed by Chomsky and Halle (1968) to describe consonants and vowels. Parameter (8) jaw openness scores - reflects an articulatory characteristic of consonants and vowels that is not included in standard feature lists. This parameter was included in order to test the hypothesis that mandibular oscillation plays the main organizational role in syllable systems. The openness scores were derived from measurements made by Lindblom (1983).

Constraints

The fitness measure in the ESS model incorporates a number of constraints, each of which is based on the premise that the sound patterns of language emerge in response to selection factors provided by the perceptual and articulatory systems of language speakers and listeners. The constraints are weighted by the experimenter. The weight values determine the penalties that less fit vocabularies (and consequently their constituent words) receive during evaluation if they violate the constraints. The penalties are added together to determine the fitness for the vocabulary. In the present simulations the following constraints were used:

1. No two identical words are allowed in the vocabulary. This first constraint follows directly from the function of language. Different concepts will only be understood if they are labeled with different acoustic patterns. Given that words are represented as strings of phonemes and vocabularies consist of a set of these strings, the constraint is easily implemented by comparing each word in the vocabulary to every other word on a phoneme by phoneme basis. The penalty P_1 associated with this constraint is calculated as:

$$P_1 = W_1 \sum_{x} \sum_{y \neq x} s_{xy}, \qquad (1)$$

where W_1 is the weight associated with this constraint, x and y are words in the vocabulary, and

$$s_{xy} = \begin{cases} 1 & \text{if } \forall p \;\; x_p = y_p, \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

where x_p is phoneme p in word x.

2. Short words are preferred. This constraint assumes that the utterance of every segment in speech requires energy, and that natural systems try to conserve energy. The production of longer words therefore requires more energy than the production of shorter words. Consequently, all other things being equal, the system should favor shorter words. This constraint is implemented by counting the total number of segments in the vocabulary and dividing the sum by the largest possible number of segments in the vocabulary:

$$P_2 = \frac{W_2}{N_x N_p} \sum_{x} \sum_{p} 1, \qquad (3)$$

where W_2 is the weight associated with this constraint, x is a word in the vocabulary, p is a phoneme in this word, N_x is the number of words in the vocabulary (25 in these simulations), and N_p is the maximum length of the phoneme string (15 in these simulations).

3. The sound structure of different words should be as different as possible. This constraint assumes that the cognitive task of accessing the conceptual structure that underlies language is made easier if the different labels that encode different concepts are as perceptually distinct from one another as possible.

Given the elaborate representation of phonemes in ESS it is possible to construct a metric of similarity between words that is based, in part, on the similarity of the characteristics of the phonemes in words and, in part, on the position within the word in which the phoneme occurs. Each word in the vocabulary is compared with every other word in the vocabulary, and the number of common features shared by each of the phonemes in the corresponding positions is counted:

$$P_3 = \frac{W_3}{N_f N_l} \sum_x \sum_{y \neq x} \sum_p \sum_f r_{xypf}, \tag{4}$$

where W_3 is the weight associated with this constraint, x and y are words in the vocabulary, p is a phoneme location in these words, f is a feature in the phoneme, N_f is the number of features in the language (8 in these simulations), N_l is the number of phonemes in the language (9 in these simulations), and

$$r_{xypf} = \begin{cases} 1 & \text{if } x_{pf} = y_{pf}, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where x_{pf} is feature f of phoneme p in word x.

4. Mandibular oscillation. The mandibular oscillation constraint is based on work by Davis and MacNeilage (1995), who suggest that simple syllable structure is derived from the basic cyclic gesture of mouth opening and closing. The constraint implies that in the sequential organization of phonemes, each phoneme must move away from the position of the preceding segment either towards openness or closure until a maximum openness or closure is attained. Once at the maximum, the following segment must move in the opposite direction. In the present model, the constraint has been implemented to encourage a maximal difference in jaw height between adjacent segments. Openness scores (parameter 8 above) for adjacent segments are compared and the differences in openness are calculated:

$$P_4 = \frac{W_4}{\sum_x \sqrt{\sum_p (o_p - o_{p+1})^2}}, \tag{6}$$

where W_4 is the weight associated with this constraint, x is a word in the vocabulary, p is a phoneme in the word, and o_p is the "openness" feature of phoneme p. By this method, vocabularies with the biggest difference scores receive the smallest fraction of the penalty associated with the constraint.

Experiments

The set of simulations described in this paper explore how different constraints effect the sequential organization of phonemes within evolving vocabularies. Vocabularies were evolved under three separate conditions in the ESS model:

1. As a control, only constraints 1 and 2 were active, that is, no constraint was placed on the sequential organization of phonemes.

2. In addition to constraints 1 and 2, constraint 3 was included, that is, the system was constrained to produce a vocabulary of words that were maximally perceptually distinct from one another.

Table 1: Initial and final vocabularies in the mandibular condition.

Initial	Final
s	nunlpul
nkksnusupiipas	sipl
tast	suslps
ts	su
ukn	nuna
i	nana
-	tunika
upil	tusl
-	tusa
-	sapas
kusnitkau	sunas
-	tunl
usuksukanisaks	sunlsn
l	sitasp
n	pnsas
uuatlapl	kltlsn
nnktsksnpkktlk	kunls
s	slsis
tp	nipl
nlnial	pisls
kilnpauuutlaap	kltls
ainttuitukinl	kuts
np	tuslpul
atsnip	tlnin
ianaitlui	tusls

3. In addition to constraints 1 and 2, constraint 4 was included, that is, the sequential organization of phonemes was constrained by mandibular oscillation.[1]

The hypothesis was that a vocabulary with a simpler syllable system, that is, one with more CV syllables and fewer consonantal clusters, would be produced only in the condition where the organization of phonemes into words was constrained by mandibular oscillation.

In all three conditions, the best vocabulary of 25 words was taken from the 750th generation of evolution. The 750th generation was chosen because by this point the progress in fitness had leveled off. Twelve simulations were run in each condition. Each simulation began at a different starting point (generated with a different random number seed), but from the same point in each condition. Table 1 shows example initial and final vocabularies in the mandibular oscillation condition.

The final best vocabularies (of the 750th generation) were analyzed in terms of (1) the number of consonants and vowels they contained, (2) the number of CV syllables, and (3) word initial and word final consonant cluster sequences. Repeated measure analyses of variance (ANOVA) were conducted to test the strength of the differences between differentially constrained vocabularies.

Results and Discussion

Simple Syllable Structure

An analysis of the number of CV syllables present in each condition showed that, in keeping with the prediction, more CV syllables occurred in the condition in which the organization of phonemes was constrained by mandibular oscillation than in either of the other two conditions.

[1]Which of the constraints are included in the simulation makes a big difference, but the relative weighting of the constraints is not crucial. Values $C_1 = 1000, C_3 = 5000, C_4 = 10000$ were used throughout, and $C_2 = 10000$ was used for the control, $C_2 = 7000$ for the perceptual, and $C_2 = 2500$ for the mandibular condition.

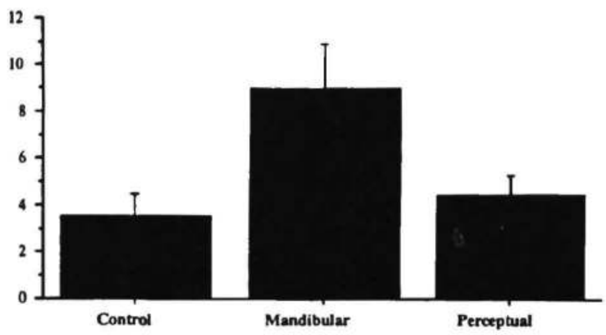

Figure 1: The average number of total CV syllable tokens in twelve final vocabularies is plotted as a function of simulation condition.

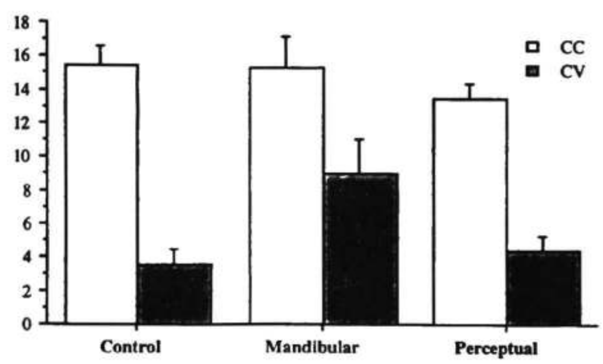

Figure 2: The average number of CC clusters and CV syllables are plotted as a function of conditions.

Figure 1 presents this result graphically. The average number of CV syllables per vocabulary is plotted on the y-axis and the different conditions are plotted on the x-axis. There was a significant difference between the mandibular condition on one hand and the random and perceptual conditions on the other ($F(1) = 7.9, p < .01$). This effect was not due to a higher ratio of vowels to consonants in the vocabulary. An analysis of the relative ratios of vowels to consonants in the three conditions indicated that a significant difference existed between the conditions ($F(2,11) = 21.48, p < .01$), but it was the perceptual condition, not the mandibular oscillation condition, that exhibited a higher ratio of vowels to consonants than the other two conditions. Thus, the large number of CV syllables present in the vocabularies that were constrained by mandibular oscillation indicate that this constraint provides a powerful organizational force.

Complexity of Organization

A second aspect of sound structure is the relative complexity of the syllable structures found in each condition. Clusters, or phoneme sequences that consist of two or more adjacent consonants, are typical of more complex syllable structure. Thus, the number of clusters at the beginning and end of words were counted for each of the vocabularies.

The results of this analysis did not support the hypothesis that vocabularies in which the sequential organization of phonemes was constrained by mandibular oscillation are simpler than vocabularies that were not constrained or that were constrained in another manner. All conditions had the same average number of nonmedial consonant clusters (Figure 2).

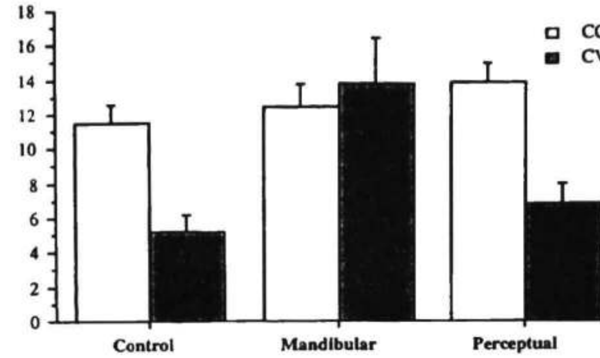

Figure 3: The average number of CC clusters and CV syllables are plotted as a function of condition in simulations where a fifth, mandatory vowel, constraint was included.

An even more disturbing finding was that in all conditions, the average number of nonmedial clusters exceeded the average total number of CV syllables. However, an examination of the types of consonant sequences indicated that clusters in the mandibular oscillation condition were systematically organized according to the close-open cycle such that the most "closed" consonants occurred at the edges of syllables (words) and the most "open" consonants occurred in the nucleus of the syllable (Table 1). This result contrasted with the results from the other two conditions where a single consonant was often repeated three or four times.

"Syllabic consonants" exist in a number of languages including the famous examples of Bella Coola and Berber. Nevertheless, languages in which some syllable peaks are consonantal are much rarer than languages in which all syllable peaks must be vocalic. In addition, these exceptional languages have much larger consonantal inventories than the 6-consonant and 3-vowel inventories used to evolve these (tiny) vocabularies. If pure combinatorics is at work in shaping the structure of these actual language vocabularies, it might be expected that vocabularies with many more consonants than vowels would have syllables with consonantal peaks, but, again, this is not the case here. The fact that so many sequences of adjacent consonants are found, even in the vocabularies constrained by mandibular oscillation, therefore, indicates that other constraints may be operative in the organization of phonemes into syllables.

To verify this hypothesis, a second set of simulations was run with an additional constraint included in all three conditions. This constraint specified that each word in a vocabulary must contain a vowel:

$$P_5 = W_5 \sum_x t_x, \tag{7}$$

where W_5 is the weight associated with this constraint, x is a word in the vocabulary, and

$$t_x = \begin{cases} 1 & \text{if } \forall p \; x_p \in [\text{p, t, k, s, l, n}], \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where x_p is phoneme p in word x.

Table 2: Initial and final vocabularies in the mandibular condition with a fifth, mandatory vowel, constraint included.

Initial	Final
itap	pisansk
ussnappklsnnka	pist
ipnp	pisini
ipi	pisu
sus	lisukl
knplsssl	spnuksp
p	sasukl
ukn	sanuksp
klnuanu	sasu
tk	takunt
klnuanu	pusun
kuninpauklnnak	sasuns
stiuansatnskla	salap
sululal	spnuki
nps	snkunsa
iisuasanuspsup	siku
psnuti	pisunlk
pi	pikaks
apapkaillul	sisukls
-	pusukls
pnlsnlstilkapt	san
itunissu	pisani
nalpkuiusasunu	lisukls
psalpssiipla	siklni
lntialkik	punanu

The vocabularies evolved under this constraint should have an increased vowel-to-consonant ratio. This constraint was developed in a post-hoc fashion, but the results are nevertheless informative. Table 2 gives an example initial and final vocabulary in the mandibular oscillation condition, and Figure 3 shows the average number of CC and CV structures.

When more vowels are present in the vocabulary, and the serial organization of phonemes is constrained by mandibular oscillation, many more CV syllables occur while the number of initial and final clusters remains stable. The increase in CV syllables relative to the number of clusters results in a system with more realistic proportions than in the preceding simulations. Note also that even with a greater presence of vowels in the vocabularies, if serial organization is not constrained or if it is constrained following a perceptual criterion, few CV syllables emerge.

The results from this last set of simulations imply that, although mandibular oscillation provides an important constraint on the sequential organization of phonemes, other constraints are also operative in the emergence of a syllable system, and the ESS system is a useful tool in identifying them. The nature of these other constraints will be explored in future research.

Conclusion

The ESS model demonstrates that specific perceptual and articulatory constraints are capable of creating, via selection, the regular syllable systems of language. The model can be used to verify hypotheses about the emergence of particular sound patterns by assessing the relative power of specific, hypothesized constraints.

In future work the model will be scaled up so that larger vocabularies with larger phonemic inventories can be evolved. The purpose will be to provide comprehensive demonstrations that sound patterns could have emerged during evolution from the selection forces provided by the perceptual and articulatory systems of language users.

Acknowledgments

This research was supported by NSF grant #IRI-9504317 and by an NSF graduate research fellowship to the first author.

References

Bell, A. and Hooper, J. Bybee. (1978). Issues and evidence in syllabic phonology. In Bell and Hooper (Eds.), *Syllables and Segments*. Amsterdam: North Holland.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Davis, B.L. and MacNeilage, P.F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*, 1199–1211.

Lindblom, B. (1983). Economy of speech gestures. In MacNeilage (Ed.), *The Production of Speech*. New York: Springer-Verlag.

Lindblom, B. (1986). Phonetic universals in vowel systems. In Ohala and Jaeger (Eds.), *Experimental Phonology*. New York: Academic Press.

MacNeilage, P.F. (1998). The frame/content theory of the evolution of speech production. *Brain and Behavioral Sciences*, in press.

Maddieson, I.M. (1984). *Patterns of Sounds*. London: Cambridge University Press.

Moriarty, D.E. and Miikkulainen, R. (1996). Efficient reinforcement learning through symbiotic evolution. *Machine Learning, 22*, 11–32.

Willerman, R. (1994). *The phonetics of pronouns: articulatory bases of markedness*. Doctoral Dissertation, Department of Linguistics, University of Texas, Austin.