Functional Bases of Phonological Universals: A Connectionist Approach

Marc F. Joanisse and Mark S. Seidenberg
University of Southern California[1]

1 Introduction

The phonological inventories of the world's languages exhibit non-random patterns relating to how phonemes are grouped into natural classes, the types of processes targeting these classes, the direction of historical change, and the frequency with which phonemes occur and co-occur in languages. The present study investigated an important class of regularities concerning the distributions of vowels. Although the human vocal apparatus can produce many possible vowels, a large proportion of languages only use between 4 and 8 of them. In addition, languages with a given number of vowels tend to use similar sets of vowels. For example, most five-vowel languages employ the set [i e a o u]; a handful use similar inventories such as [i e a ə u]; and many possible sets of vowels are not observed at all, such as [e y œ æ u][2].

The standard approach within the generative tradition has been to view these phenomena in terms of the concept of markedness: vowel features are organized into a markedness hierarchy, such that vowels incorporating more marked features are less suitable in an inventory (Chomsky & Halle 1968; Clements 1985). The major drawback to this approach is the lack of criteria independent of mere frequency of occurrence for determining which vowels or features are "marked."

Our research addresses the hypothesis that vowel inventory patterns reflect functional constraints related to perception and production; specifically, languages tend to maximize distances between vowels. We focus here on the tendency to maximize the acoustic distances between constituent vowels, although featural or gestural distance may also be relevant. Inventories involving acoustically well-dispersed vowels are easier to both acquire and process because they are easier to discriminate, creating a tendency for languages to recruit such inventories. In contrast, less acoustically dispersed vowel inventories are more difficult to acquire and process because of the greater probability of misperceiving a constituent vowel, leading to languages shifting away from such inventories. On this view, inventories such as [i I e ɛ a] do not occur because they involve smaller distances than other, attested vowel sets (Jakobson 1941).

A similar idea has been explored using mathematical models (Liljencrants & Lindblom 1972; Lindblom 1986; Boë, Schwartz, & Valée 1994) in which the acous-

tic distances of vowels are expressed as a function of the Euclidean distance between their corresponding formant frequencies. Such models predict perceptually optimal inventories by maximizing the distances between all vowels in a given set. Although this approach can account for a considerable amount of data about the distributions of vowels, it is limited in several respects. First, it does not represent the variability with which vowels are produced; as we shall see, this variability can play a role in the frequency with which a vowel occurs in an inventory. Second, this approach is not a model of why this type of distance maximization occurs; we tie this optimization to constraints related to learning and processing. Finally, this approach does not easily allow the integration of other types of factors used to differentiate vowels, such as nasalization, diphthongization and vowel length.

Connectionist Models of Phoneme Acquisition

The present work builds on Lindblom's approach by situating it in a theory of how phonemic inventories are acquired and processed. In Joanisse & Seidenberg (1997) we described research in which connectionist models are trained to recognize the phonemes of pseudo-languages consisting of different vowel inventories. The approach is based on the premise that, like humans, connectionist models are not equally predisposed to learning and recognizing all types of patterns. By varying the characteristics of these vowel inventories and assessing the models' capacities to learn them, we generate predictions about the relative suitability of inventories. Inventories that are easier to learn are predicted to be more likely to occur in the languages of the world. Inventories that cannot be learned are predicted to not occur.

The general structure of the models used in the present research is illustrated in Figure 1. The ellipses represent groups of artificial neurons that encode information as patterns of activation. Lines between layers of units represent sets of connections through which activation is passed, in the direction of the arrows. The middle, so-called hidden, layer of units is used to enhance the representational capacity of the network[3].

The model is trained to recognize speech sounds based on a set of examples. The input to the model is the spectrographic representation of a given speech item. Identification results from passing activation through two layers of connections to the output layer. Learning proceeds through the adjustment of connection weights using the backpropagation learning algorithm (Rumelhart, Hinton, & Williams 1986); over the course of training, the model adjusts these weights in ways that facilitate accurate identification of the training stimuli and generalization to novel examples. The relative difficulty of learning the training set is reflected in the rate at which training proceeds, asymptotic levels of performance, and the number of items a trained network misclassifies.

We used this model to explore two types of vowel inventory phenomena. The

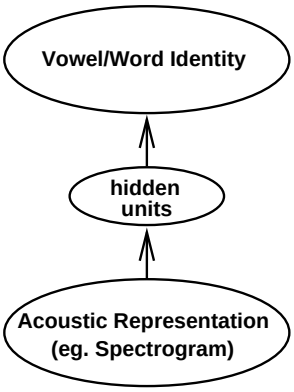


Figure 1: General structure of the models used in the present paper.

first relates to the preference of four-vowel languages to choose front mid vowels over back mid vowels, which we propose is related to differences in these vowels' variability. In the second set of simulations, we explore the interaction between the number of vowel quality contrasts in a language, and its tendency to use contrastive length. It is shown that both factors affect ease of learning and processing in the model, providing an explanation for why certain inventories are preferred.

2 Experiment 1: Front-Back Asymmetries

The first set of simulations concerns the observation that, in four vowel languages, there is a greater tendency for languages to use a front mid vowel than a back mid vowel, as illustrated in Figure 2. This is an interesting asymmetry, given that the dispersion characteristics of the two inventories are approximately equal. A simple dispersion model as in Liljencrants & Lindblom (1972) predicts there to be little difference in the occurrence of the two inventory types, but this is contradicted by the empirical data.

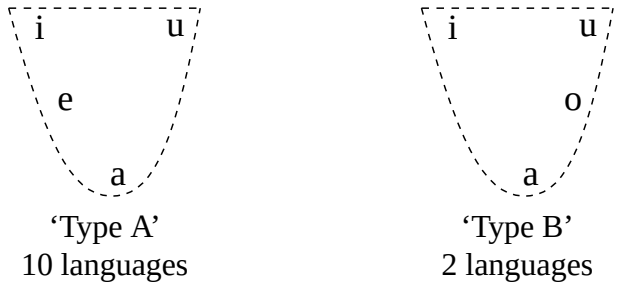


Figure 2: Asymmetry of back vowels over front vowels. Note that the relative dispersion of the two sets is roughly the same.

One explanation for this asymmetry derives from differences between front and back vowels with respect to their tendency to vary in production. As Figure 3 illus-

trates, there is a difference between nonlow front and back vowels in their tendency to overlap with each other. Speakers are able to produce the vowel /i/ with a better degree of precision, by stiffening the genioglossus muscle and propping the tongue laterally against the dental ridge (Beckman, Jung, Lee, de Jong, Krishnamurthy, Ahalt, & Cohen 1995). This is not possible for back vowels however, which leads to a greater F1 variability for /u/ than for /i/, causing more overlap for the /u/ – /o/ contrast than for the complementary /i/ – /e/ contrast.

This type of explanation is consistent with Stevens' Quantal Theory (Stevens 1989), in which a phoneme's distinctiveness is affected by nonlinearities in the relationship between the vocal tract and acoustics: phonemes with quantal articulations are those which can be produced with greater precision as a result of such nonlinearities. In the present account, a vowel inventory's frequency can be explained in part as the result of the overlap of its constituent vowels – inventories which minimize this overlap have more discriminable vowels, and are more likely to occur.

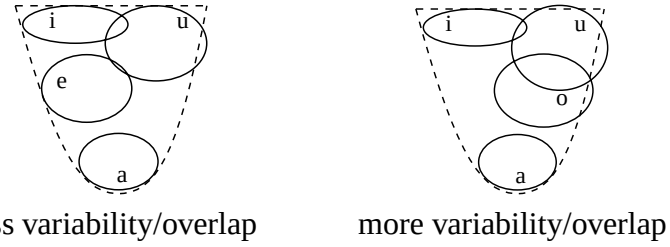


Figure 3: Schematization of vowel variability. Note the greater tendency for nonlow back vowels to overlap, compared to their front counterparts.

Method & Stimuli

To test the hypothesis that production variability has an effect on vowel inventory frequencies, we trained networks on the invented vocabularies of one of two pseudolanguages. These pseudolanguages differed only with respect to their vowel inventories: the artificial vocabularies consisted of all CV combinations of the consonant set [p t k b d g] and either the vowel set [i e a u] (Language A) or [i a o u] (Language B). Training stimuli were devised by extracting CV syllables from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus of spoken English. Raw waveforms were transformed into fourteen 8 ms frames of 96 spectral coefficients (bandwidth: 62.5 Hz, frequency range: 0-6000 Hz) using the Fast Fourier transformation. The result was a set of 1,344 spectral coefficients for each CV syllable in the training set.

Using waveforms drawn from the TIMIT database allowed us to obtain many instances of each syllable type (in most cases, greater than 25 of each type) as spoken by many speakers of both sexes and of many American regional dialects. Although networks were not explicitly trained to identify vowels and consonants,

the identification task at hand would force these networks to learn to identify the consonant and vowel components of the input set in order to perform the task at optimal levels. This is consistent with what we believe to occur in children as they learn to identify the words in their target language. It is hypothesized that the greater degree of overlap between nonlow back vowels compared to nonlow front vowels will cause slower learning rates and poorer generalization in models trained on Language B, compared to Language A.

Network training consisted of 100,000 training trials. During each training trial, the model was presented with the spectral coefficients of a CV syllable randomly drawn from the training set, and trained to identify this syllable by activating the correct output unit. There were 24 output units, each corresponding to a different CV syllable type. All types were presented an equal number of times over the course of training.

Results and Discussion

Three different networks were trained on either Language A or B, for a total of six networks; averaged results for all six are reported here. To assess overall learning of training items, networks were tested at intervals of 10,000 training trials on the complete training set. This was done by presenting all items to the network, and calculating the resulting Sum Squared Error which is a function of the overall correct and incorrect activation of all output units (Rumelhart, Hinton, & Williams 1986). Results are plotted in Figure 4. Higher mean error rates for models trained on Language B indicate that these networks had more difficulty learning the training set consisting of the vowel set [i a o u] compared to those trained on [i e a u].

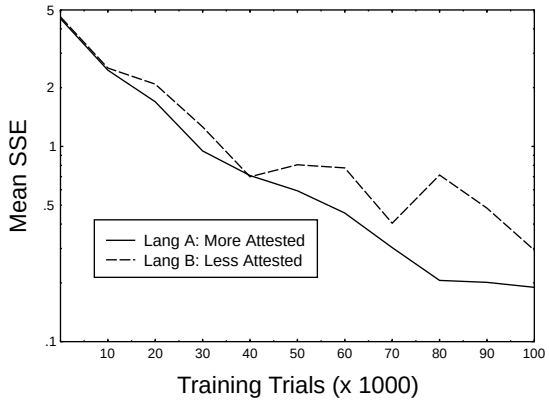


Figure 4: Sum Squared Error rates averaged over all networks trained on two pseudolanguages. Lower rates for Language A suggest less difficulty in learning the training set.

This result is consistent with the hypothesis that the learnability of a vowel

inventory affects its frequency of occurrence in the world's languages. In addition, these networks were capable of learning both pseudolanguages, consistent with the observation that both inventories are attested, although with different frequencies. Neither inventory is completely unlearnable within this architecture (unlike more extreme cases investigated in Joanisse & Seidenberg (1997), such as [i I e ɛ æ]). Given sufficient training, networks trained on either inventory can achieve near-perfect performance in categorizing items in the training set.

To further investigate how these inventories differ in their degree of suitability, we also tested networks on their ability to generalize to novel stimuli. This type of testing is comparable to the type of task confronting an adult language user, who must recognize novel tokens of familiar words in the course of everyday language processing. This was done by presenting fully trained networks with instances of CV syllables randomly withheld from the training set. Analyses of these errors indicated that networks trained on Language B misclassified novel vowels more frequently than Language A (Lang A: 71% correct; Lang. B: 65% correct).

Results from these simulations support the hypothesis that vowel systems are functionally optimized in ways that maximize the discriminability of their constituent vowels because of its effects on learning. In addition, these simulations suggest that accounting for these phenomena turns on examining stimuli that realistically represent variability in terms of production and overlap.

3 Experiment 2: Length and Quality Interactions

So far we have only considered vowel inventory tendencies in terms of quality contrasts related to formant frequencies. However, many languages use other types of cues to contrast vowels. Here we consider how one such cue, vowel length, interacts with spectral (formant-based) cues. Maddieson observes that 'The probability of length being part of the vowel system increases with the number of vowel quality contrasts' (Maddieson 1984:129). As such, 12.5% of languages with 4 to 6 vowel qualities and 24.7% of languages with 7-8 vowel qualities incorporate length contrasts, compared to 53.8% of languages with 10 or more vowel qualities. However, these data are incomplete, owing to the nature of the Maddieson (1984) database which does not consider contrastive length in cases where all vowel qualities participate in length contrasts.

We propose that this pattern is not accidental, and that a more extensive survey of length in vowel systems would reveal a similar pattern for languages not included in the Maddieson database. This pattern is attributed to the weak contrastiveness of durational cues for vowels, compared to spectral cues, owing to the degree of variability intrinsic to vowel length. Vowel length appears to be a useful cue in disambiguating various language contexts, for example in determining the voicing of an adjacent consonant (Chen 1970), rate of speech (Magen & Blumstein 1993) and a lexical boundary (Davis, Marslen-Wilson, & Gaskell 1997). Given this tendency

for vowels to vary in duration, it is plausible that adding a durational contrast would be dispreferred.

This analysis is additionally supported by the observation that length contrasts are often observed to accompany close quality contrasts (Maddieson 1984). This suggests that duration is a useful secondary cue to differentiating spectrally similar vowels, though on its own, it may prove to be less useful than a spectral contrast.

Connectionist models provide a way to explore these phenomena, because of their capacity to exploit multiple, simultaneous, probabilistic regularities in the service of learning to perform a task. The relative usefulness of cues can be assessed in terms of the model's ability to reliably learn and use them. In the present simulations, connectionist networks were trained to acquire vowel inventories that use spectral or length contrasts to different degrees. It was predicted that differences between network learning and generalization would reflect known facts about the occurrence of these cues in vowel systems of different sizes.

Method & Stimuli

Networks were trained on one of three vowel inventories which use length and spectral contrasts to different degrees. As Table 1 shows, these inventories seek to double the number of contrasts in the familiar [i e a o u] set by adding spectral contrasts, length contrasts, or both. For clarity, these inventories are also plotted in Appendix A.

inventory	*vowel set*
1 - quality only	[i ɪ e ɛ æ a ɔ o ʊ u]
2 - length	[i i: e e: a a: o o: u u:]
3 - both length and quality	[i: ɪ e: ɛ æ: a ɔ o: ʊ u:]

Table 1: Vowel inventories used in the present study. All seek to double the base [i e a o u] inventory, though in different ways.

Given the hypothesized differences in the contrastiveness of length and spectral cues, we predicted that networks would have more difficulty learning Inventory 1 compared to 2 and 3, because only spectral contrasts are used, and that Inventory 3 would be easier to acquire than Inventory 2 due to the interaction of duration and spectral cues in maintaining contrast in such a relatively large and crowded inventory. This would be consistent with the observation that a contrast like /i/ – /I/ is dispreferred compared to /i/ – /e/, though it might be more common than /i/ – /i:/. Ultimately, /i:/ – /I/ seems to represent a good compromise for languages with a crowded vowel space.

The architecture of the model was similar to those used in the previous simulations. Input consisted of 2040 spectral coefficients encoding the acoustic representation of a vowel. A total of 40 vowels of each type were devised for training

purposes. Training sets consisted of synthetic vowels, made to be highly realistic by using formant means and variances drawn from observed data in the TIMIT database and data published in Beckman et al. (1995). Contrastive vowel length was simulated by creating long vowels with a mean duration 1.66 times longer than short vowels, and with a standard deviation of 0.5 for the long vowels, and 0.33 for the short vowels[4]. In this model, vowel length was implemented by varying the number of non-empty frames presented to the input layer, such that longer vowels had more 'filled' frames, and shorter vowels had more empty frames.

Model training proceeded similarly to the previous experiment, though in the present simulations, the training task was simply to identify the vowel presented on the input by activating the appropriate output node. There were 10 output nodes for each network, each corresponding to a separate vowel identity; in cases where similar vowel qualities were differentiated by contrastive length, separate output nodes were used for the long and short versions of the two vowels. Three networks of each type were trained for 250,000 training trials. To compensate for the relatively simple nature of this task, 15 hidden units were used in each simulation.

Results

Sum Squared Error means are plotted in Figure 5. These results indicate that networks trained on Inventory 3 were learning the training set slightly better than those trained on Inventory 2, and much better than those trained on Inventory 1. To further assess model performance, fully trained networks were tested on a set of 10 novel vowels of each type. The mean percent of correctly generalized items was 62.3% for Inventory 1, 86.7% for Inventory 2, and 91% for Inventory 3.

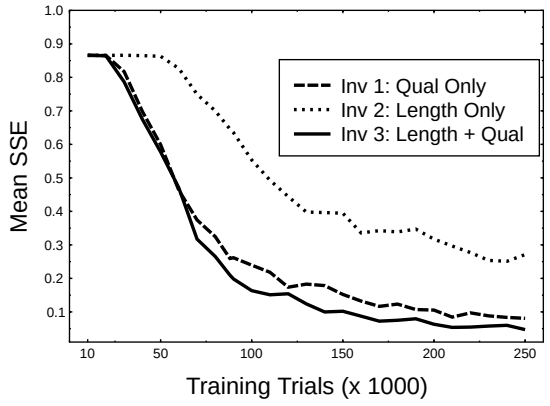


Figure 5: Mean Sum Squared Error rates for the three network types in this experiment. Lower values indicate better performance.

Summary

The present simulations support the hypothesis that contrastive vowel length represents a weaker cue than typical spectral differences for discriminating vowels, and affects languages' tendency to recruit contrasts which rely solely on length distinctions. This explains the paucity of languages with smaller inventories (3-8 vowel quality contrasts) that use vowel length distinctions. And while frequency data is missing for languages in which all vowel qualities participate in length contrasts, these results predict that the facts should not be different for such languages. Finally, these results are consistent with observations that length contrasts tend to accompany smaller quality contrasts (e.g. /i:/ –/I/).

4 Discussion

This work has explored the idea that the vowel inventory preferences of the world's languages result from their functional optimization. The simulations described here demonstrate the utility of connectionist networks in exploring this type of hypothesis. The performance of these models is easy to assess, and allows us to directly compare results to empirical facts about the world's languages. Here we have explored only a few of the possible uses of this approach. Additional applications include testing the contrastiveness of other vowel cues, such as diphthongization and nasality. This type of approach could also be useful in assessing the role of discriminability in consonant inventory frequencies. For example languages' preference of velar and alveolar stops over palatal stops might also be a function of these phonemes' discriminability.

The results of such simulations can serve to explain phonological patterns based on articulatory, acoustic and computational constraints. The network discovers constraints in the course of learning to perform a task. In contrast to other approaches such as Optimality Theory (Prince & Smolensky 1997), constraints do not have to be specified in advance; they emerge in the course of acquisition given the nature of the architecture, the characteristics of the input, and the task being performed.

Notes

[1] Address correspondence to Marc Joanisse, USC Neuroscience Program, University Park, Los Angeles CA 90089-2520 (email: marcj@gizmo.usc.edu).

[2]See chapter 8 of Maddieson (1984) for an overview of facts concerning vowel inventories.

[3]See Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett (1996) for a discussion of how connectionist models work.

[4]Variabilities are estimates drawn from recordings of Finnish speakers, where lengths of vowels spoken in similar consonantal contexts – but differing prosodic contexts – were measured and compared. They represent a best guess at durational contrast and variability for a language with reliable length contrast, although languages will tend to vary along these parameters.

References

Beckman, M. E., T.-P. Jung, S. Lee, K. de Jong, A. K. Krishnamurthy, S. C. Ahalt, & K. B. Cohen (1995). Variability in the production of quantal vowels revisited. *Journal of the Acoustical Society of America 97*, 471–90.

Boë, L.-J., J.-L. Schwartz, & N. Valée (1994). The prediction of vowel systems: Perceptual contrast and stability. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition*, pp. 185–213. John Wiley & Sons.

Chen, M. (1970). Vowel length as a function of the voicing of the consonant envorinment. *Phonetica 22*, 129–159.

Chomsky, N. & M. Halle (1968). *The Sound Pattern of English*. MIT Press.

Clements, G. (1985). The geometry of phonological features. *Phonology Yearbook 2*, 225–52.

Davis, M. H., W. D. Marslen-Wilson, & M. G. Gaskell (1997). Ambiguity and competition in lexical segmentation. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Volume 19, Mahwah, NJ, pp. 167–172. Laurence Erlbaum.

Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, & K. Plunkett (1996). *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Jakobson, R. (1941). Kindersprache, aphasie und allgemeine lautgesetze. In *Selected writings I*, pp. 328–401. The Hague: Mouton.

Joanisse, M. F. & M. S. Seidenberg (1997). [i e a u] and sometimes [o]: Perceptual and computational constraints on vowel inventories. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Volume 19, Mahwah, NJ, pp. 331–336. Laurence Erlbaum.

Liljencrants, J. & B. Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Journal of Phonetics 48*(4), 839–862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala (Ed.), *Experimental Phonology*. Academic Press.

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

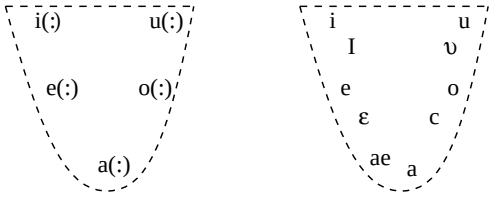
Magen, H. S. & S. E. Blumstein (1993). Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics 21*, 387–409.

Prince, A. & P. Smolensky (1997). Optimality: From neural networks to Universal Grammar. *Science 275*, 1604–1610.

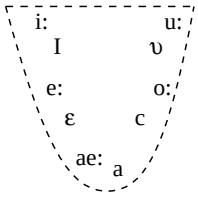
Rumelhart, D., G. Hinton, & R. Williams (1986). Learning internal representations by error propagation. In D. Rumelhart and J. McClelland (Eds.), *Parallel distributed processing, vol. 1*. Cambridge, MA: MIT Press.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics 17*, 3–45.

Appendix 1: Inventories used in Experiment 2



Inv. 1) Inv. 2)



Inv. 3)