In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, Lawrence Erlbaum, Hillsdale, New Jersey, 1993*

Exploring the Nature and Development of Phonological Representations

Prahlad Gupta
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
prahlad@cs.cmu.edu

Michael C. Mozer
Department of Computer Science
University of Colorado
Boulder, CO 80309
mozer@cs.colorado.edu

Abstract

Findings in infant speech perception suggest that early phonological perceptions may be syllabic in nature, and that there is a loss of sensitivity to nonnative contrasts toward the end of the first year of life. We present a neural network model that simulates these two phenomena. In addition, the model and simulations (1) demonstrate how information about stress can be utilized in generating syllable-like perceptions; (2) provide a simple means of extracting static representations from a dynamic and co-articulated signal; and (3) indicate that the development of "attractor" states may be necessary in network models of these phenomena.

Aims and motivation

This work explores certain aspects of the nature and development of phonological representations in human infants, using neural network modeling techniques. A "phonological representation" is taken to be a *phonologically organized encoding of speech sounds*[1]. This paper focusses on the idea that early representations are likely to be syllabic in nature.

There appears to be considerable agreement that the *syllable* is a readily accessible perceptual unit of phonological representation, both for children and for adults, i.e., that there is perceptual segmentation at the level of the syllable (Menyuk et al., 1986; Flores d'Arcais, 1988; Jusczyk, 1992). Evidence further indicates that the syllable is much more accessible than the segment and feature, for adults as well as developmentally (Menyuk et al., 1986). Moreover, children appear to have little awareness of the phonemic (i.e., segmental) structure of words (Jusczyk, 1986; Adams, 1990).

Two specific phenomena observed during the first year of life have been chosen for exploration here. One phenomenon suggests that two-month-old infants are more sensitive to syllable-sized units than to sub-syllabic units (Jusczyk, 1992). The other phenomenon indicates a loss of sensitivity to nonnative language contrasts somewhere between the ages of 6 and 12 months (Werker, 1991).

[1]This could denote either an output representation of a word, i.e., a representation involved in a motor program to articulate that word, or an input representation or percept yielded by auditory analysis of the word. For the purposes of this paper, however, "phonological representation" will mean an input representation.

The aim of the present work is to try and account for both these phenomena via simulation in a single model. A further modeling goal was to tackle the temporal nature of the speech signal, as well as its lack of linearity[2]. It seems vital for models of phonological processing to deal with these issues, especially if meaningful conclusions are to be drawn about such things as phonological representations.

Description of the model

Input representation

The input representation was intended to incorporate aspects of the temporal and non-linear nature of the speech signal available to a human learner.

The starting point was a phonemic feature representation proposed by (Shillcock et al., 1992), in which each possible phoneme is encoded in terms of a set of 9 feature values, which are intended to have physical correlates in the speech signal.

Several previous psychologically motivated connectionist learning models have utilized some such featural representation, simulating the temporal nature of the speech signal by inputting a sequence of feature vectors over time, representing a stream of phonemes (Elman, 1990; Norris, 1990; Gasser, 1992; Shillcock et al., 1992). However, such a representation imposes linearity on the input signal[3]. To represent some of the non-linearity of the speech signal, we devised the scheme illustrated in Figure 1.

A consonant is represented as three "time slices" long, while a vowel is ten time slices long (for simplicity, vowels are depicted as only six time slices long, in the figure). As shown for the nonsense word *bagi*, there is overlap of the various phonemes[4]. As a result, each time slice contains the feature vectors of up to three phonemes. For

[2]This refers to the fact that the acoustic information pertinent to identifying a particular sound segment is usually not in one piece of the signal, but is smeared over the continuous waveform, and overlaps with time slices that convey information about other segments.

[3]The TRACE model (McClelland and Elman, 1986) is explicitly concerned with dealing with such lack of linearity in the input. However, TRACE is not a learning model. Kohonen's "neural phonetic typewriter" (Kohonen, 1991) is a learning model, and is similar in some respects to the present work; the main focus of that model, however, was not on psychological or representational issues.

[4]This input representation is similar to that of TRACE-II (McClelland and Elman, 1986).

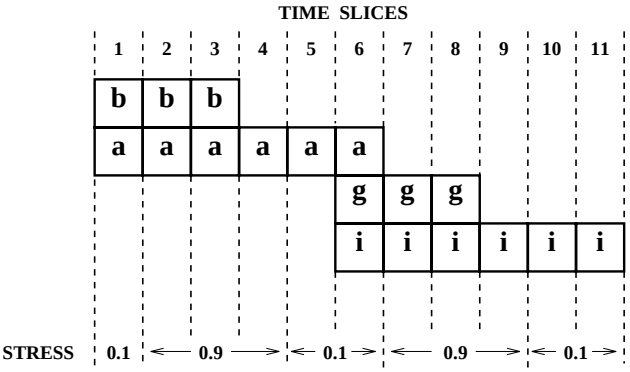

Figure 1: Input representation for the nonsense word *bagi*.

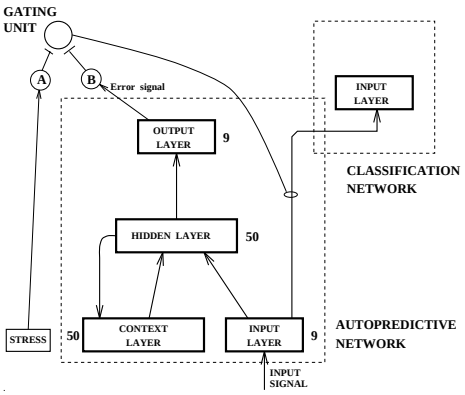

Figure 2: Overall architecture of the model. Numbers indicate number of units in a layer. Arrows indicate connectivity.

each time slice, a composite vector was created by taking the element-wise maximum value of the component feature vectors. Each time slice thus conveys "smeared" information about the phonetic content of the word. This scheme to some extent captures both the temporally extended nature of the speech signal, and its non-linearity.

Information about the *stress level* at each time step is also assumed to be available as input. It seems plausible to assume that such information is available to a developing system, as human sensitivity to prosodic features appears to be present at birth (Mehler et al., 1988). Each vowel has a *stress contour*, with no stress on its first time slice or on its last two time slices, and with peak stress on its middle time slices. For example, in Figure 1, the middle three time slices of /a/ and /i/ have a stress level of 0.9, while their other time slices have a stress level of 0.1, denoting no stress. The overall stress level for a time slice is the maximum stress of its constituent phonemes.

Architecture

The model, shown in Figure 2, consists of two networks and a *gating unit*. The *autopredictive* network is a simple recurrent network (Elman, 1990) whose input at each time step is the composite vector denoting the current time slice, presented to the input layer. In addition, hidden layer activations from the previous time step are copied to the context layer, and form part of the input. The network's task is to predict input at the next time slice as a pattern of activation over the output layer.

The stress level associated with each time slice is not encoded as part of the input signal, but instead provides input to the gating unit (see below), which is the second component of the system. The treatment of this information as qualitatively different from information about the actual content of the signal is consistent with the linguistic treatment of stress as a *suprasegmental* phenomenon (Kaye, 1989).

The third component of the model is a *classification* network. Its input comes from the same stream that feeds into the autopredictive network. However, these input connections are gated by the inhibitory gating unit, which is ordinarily active, thus preventing input from reaching the classification network. The gating unit receives inhibitory inputs from units *A* and *B*. Unit *A* fires if the stress level associated with the current time slice crosses a threshold, and unit *B* fires if the error signal from the autopredictive network crosses threshold. If both these units fire simultaneously, the gating unit is inhibited, allowing the input signal to reach the classification network. Input to the classification network is thus modulated by the results of processing in the autopredictive network, together with the level of stress of the current input.

For the classification network itself, three alternative architectures were investigated (Figure 3). These were (1) a Deterministic Boltzmann Machine (DBM) (Peterson and Anderson, 1987; Hinton, 1989), (2) a Competitive Learning network (CL) (Rumelhart and Zipser, 1986), and (3) a Multi-Layer Perceptron (MLP) (Rumelhart et al., 1986a).

The autopredictive network is meant to correspond to a level of processing that is more auditory in nature, while the classification network is intended to correspond to a more phonological level of processing. That is, at each point in the system's development, the responses of the classification network constitute its phonological perceptions.

Processing

To clarify the nature of processing in the model, we now step through part of the processing of the nonsense word *bagi*.

The word is represented as a series of time slices, as shown in Figure 1. At the first time step, the composite vector for the first time slice is presented as input to the autopredictive network; at the second time step, the vector for the next time slice is presented, and so on. As noted in the previous section, the gating unit is ordinarily active, preventing input from reaching the classification network. Input will reach the classification network, however, if the gating unit is itself inhibited.

The input vector produces a pattern of activations at the output layer of the autopredictive network, representing a prediction of the input at the next time step. Comparison of this output with the actual input at the next time step

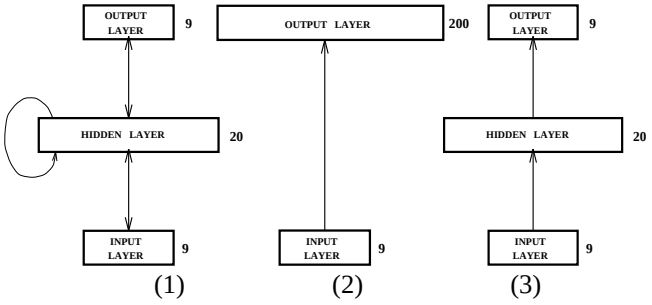

Figure 3: Classification network architectures. (1) Deterministic Boltzmann Machine; (2) Competitive Learning Network; (3) Multi-Layer Perceptron. Numbers indicate number of units in a layer. Arrows indicate connectivity.

yields an error signal (sum of squared error). If the error magnitude is greater than a threshold θ_B, unit *B* fires, resulting in an inhibitory input to the gating unit. If the stress level associated with the current time slice is greater than a threshold θ_A, unit *A* fires, providing another inhibitory input to the gating unit. If both *A* and *B* fire, the gating unit is itself inhibited. In this case, the input signal for that time slice will reach the classification network. Thus, as noted previously, input to the classification network is modulated by the error signal and the stress level. The thresholds were set at $\theta_B = 0.1$ and $\theta_A = 0.4$.

Suppose that the error signal on the first time-step for *bagi* is 0.5. The stress level is 0.1, as shown. Unit *B* fires, but not unit *A*. Inhibitory input to the gating unit therefore remains below threshold, the gating unit therefore maintains its inhibitory influence, and the input signal therefore does not reach the classification network.

On the next time-step, the input signal is the second time slice, which has an associated stress level of 0.9. Suppose the error in the autopredictive network is again 0.5. Both units *A* and *B* will now fire, the gating unit is therefore inhibited, and the input signal for the second time slice does reach the classification network. Processing for the remaining time slices of *bagi* proceeds in similar fashion.

As a result of this gating scheme, input reaches the classification network only when both stress and error are above threshold. Following training of the autopredictive network, it turns out that this happens only on time slices that are comprised of exactly one consonant and one vowel. In effect, the gating mechanism filters system input so that the classification network receives only invariant demisyllables as its input. This provides the basis for an account of certain early perceptual phenomena in human infants, discussed in the next section.

Input that does reach the classification network enters its input layer. As noted above, three architectures were examined for the classification network. For the DBM and MLP, the task of the network is to reproduce the input layer vector at the output layer (see Figure 3). For the CL architecture, the task is to categorize the input layer vector, by turning on exactly one of the output layer units.

For the DBM, the output activation is obtained by applying the input vector, and then performing synchronous updates of unit activations in repeated cycles until the magnitude of changes in unit activations falls below a specified criterion, i.e., until the network settles. The output unit activations at this time constitute the network's output. For the MLP, the output layer activation is produced by propagating the input vector forward in one pass.

For the CL architecture, the "winner" is chosen to be the unit with weights closest to those of the input vector, as in the standard algorithm (Kohonen, 1984), but with the additional requirement that the error for the winner be below a specified criterion. If it is not, an "uncommitted" unit is chosen to be the winner, in similar fashion to ART-1 (Grossberg, 1987). During training, this error criterion was progressively relaxed.

Weight adjustment for the autopredictive network and MLP classification network was via the back-propagation algorithm (Rumelhart et al., 1986a). Weights in the DBM classification network were adjusted via contrastive Hebbian learning (Hinton, 1989). In the CL classification network, the winner's weights were adjusted via the competitive learning equation given in Kohonen (1984).

Simulations and data

Like the human infant, the system is exposed to evironmental speech sounds. To model this, a set of mono- and di-syllabic words was constructed using consonants from the set {p, b, t, d, k, g, m, f}, and vowels drawn from the set {a, o, i, e}. No attempt was made to mirror the precise environmental distribution; the aim, rather, was to construct a limited sample incorporating some of the salient characteristics of the speech signal in English, using the representational scheme described earlier. This sample consisted of a set of 48 words such as *pot, dog, cat, big, pocket,* and will be referred to as the *input corpus*.

In the simulations described below, we assume a rough correspondence of 10 training epochs to one month of chronological age. Simulation experiments were performed at 20, 80, and 100 epochs of training, modeling empirical data from 2, 8, and 10 months, respectively. During the first 80 epochs, only the autopredictive network was trained. By this point, it had become stable in its predictions, and then in the next 20 epochs, weights in the classification network were also adjusted. This training procedure has the same qualitative effect as training both networks simultaneously from the beginning, because the classification network cannot establish stable categories until the autopredictive network stabilizes its responses.

The two-month-old infant

One particular focus of interest here is on experimental work examining the differential sensitivity of two-month-old infants to syllable-sized vs. non-syllable sized units. Jusczyk and colleagues (Jusczyk, 1992) have studied two-month-old infants who were exposed to a set of bisyllabic stimuli that either did or did not contain a common syllable (e.g., [bazi], [balo], [bamIt] vs. [pazi], [nalo], [kamIt]). After a two-minute delay period following exposure to one of these sets of stimuli, the infants were exposed to

a modified version of the original stimulus set. The first syllable of one nonsense word had been changed from *ba* to *da*, so that the stimulus set was now either [bazi], [dalo], [bamIt] or [pazi], [dalo], [kamIt]. Only infants who had previously heard the set containing the common syllable (i.e., the first set listed above) detected the subsequent change to the set.

In a similar test, the infants were initially familiarized with stimulus sets that either shared ([bi, ba, bu]) or did not share ([si, ba, tu]) a common phoneme. After the two-minute delay, they were presented with a modification of the original set, in which [ba] had been replaced with [da]. There was no advantage for the set with shared material—that is, there was no significant difference in responding to the changed stimulus set whether or not the original set had shared a common phonetic segment (Jusczyk, 1992).

These findings suggest that the presence of a shared *syllable* in a set of stimuli leads to a perception of similarity that does not arise when the set of stimuli shares a common *phonetic segment*[5]. This in turn suggests that, as noted earlier, the syllable is a more readily accessible perceptual unit than the segment.

To simulate these experiments, we constructed the sets of stimuli shown in Table 1. At the end of 20 epochs of training on the input corpus (with weight adjustment of the autopredictive network only), the various stimulus sets above were presented as input to the overall system, and the responses of the classification network were recorded (with all three architectures). This was meant to simulate the responses of the two-month-old infant.

Results from the CL architecture provide the clearest picture of how the classification network responds. Table 1 shows the indices of the sequence of winning competitive units for each stimulus. As an example, the first entry in the left column indicates that, when the stimulus *ba-ke* was being presented to the system, the classification network's response consisted of sequential activation of the units whose indices are 101 and 116.

For the SYL stimuli, unit 101 is repeatedly active (at the beginning of each stimulus), and this pattern of repeated activity is changed by the SYL-C stimuli. For the NOSYL stimuli, however, there is no repetitive activity of any particular unit, and so the NOSYL-C stimuli do not represent disruption of a regular pattern. It seems entirely reasonable that the repeated visiting of a particular network state (activity of unit 101) would produce a highly salient perceptual experience, deviation from which would be easily detected. This provides a basis for understanding why the SYL/SYL-C change might be more easily detected than the NOSYL/NOSYL-C change. Network states for the SEG/SEG-C and NOSEG/NOSEG-C stimulus sets, in contrast, illustrate the fact that the various syllables *ba, bi* etc. are categorized as distinct percepts. Accordingly, the patterning of network responses to the SEG stimuli is

[5]Since there are two segments of overlap in the case of [bazi], [balo], [bamIt], but only one segment of overlap in the case of [bi, ba, bu], it could simply be the greater degree of overlap that causes perception of similarity in the first case. The appropriate control has been performed to rule out this possibility (Jusczyk, 1992), although we will not describe it here.

SYL			NOSYL			SEG			NOSEG		
ba-ke	101	116	*pa-ke*	182	116	*bi*	40		*fi*	57	
ba-to	101	17	*da-to*	67	17	*ba*	101		*ba*	101	
ba-mi	101	62	*ka-mi*	116	62	*bo*	204		*to*	17	
SYL-C			NOSYL-C			SEG-C			NOSEG-C		
ba-ke	101	116	*pa-ke*	182	116	*bi*	40		*fi*	57	
fa-to	180	17	*fa-to*	180	17	*da*	67		*da*	67	
ba-mi	101	62	*ka-mi*	116	62	*bo*	204		*to*	17	

Table 1: Responses of CL classification network to stimulus sets, i.e., indices of the sequence of winning competitive units in CL network in response to each stimulus. Description of stimuli: (SYL) {*ba-ke ba-to ba-mi*} (shared syllable); (SYL-C) {*ba-ke fa-to ba-mi*} (shared syllable, changed); (NOSYL) {*pa-ke da-to ka-mi*} (no shared syllable); (NOSYL-C) {*pa-ke fa-to ka-mi*} (no shared syllable, changed); (SEG) {*bi ba bo*} (shared segment); (SEG-C) {*bi da bo*} (shared segment, changed); (NOSEG) {*fi ba to*} (no shared segment); (NOSEG-C) {*fi da to*} (no shared segment, changed).

no more salient than for the NOSEG stimuli, and there would be no basis for differential sensitivity between the SEG/SEG-C and NOSEG/NOSEG-C changes.

It is important to note that equivalent results were obtained with both the other classification network architectures. Output unit responses for both the DBM and MLP architectures were projected onto the first two principal components. Trajectories in this two-dimensional space during presentation of the SYL stimuli all began at the same point, which can be thought of as representing the syllable *ba*. None of the other stimulus sets SYL-C, NOSYL or NOSYL-C had trajectories with this property. This behavior provides a rationale for why it would be easier to detect the SYL/SYL-C change than the NOSYL/NOSYL-C change, completely analogous to that for the CL architecture, except that the network states are distributed representations rather than discrete unit activations. The results for stimulus sets SEG/SEG-C and NOSEG/NOSEG-C were also analogous to those obtained with the CL architecture.

The classification network responses thus provide an account of the phenomena observed in two-month-old infants. This behavior is based on the fact that the classification network uses demisyllable-sized perceptual chunks. That is, demisyllables such as *ba*, *bi* and *bo* are all classified as distinct percepts. In consequence, *ba-mi* and *ba-to* are similar, whereas *ba* and *bi* are not.

Two features of the model are critical in establishing this perceptual unit. First, consonants in the input are never pure, but are always flavored by adjacent vowels, which is a realistic property of the input representation. As a result, there is no time slice from which the classification network could derive the percept of *b*. This excludes the possibility of consonantal perceptual units. Second, the establishment of *demisyllables* as the units results from joint modulation, by the autopredictive error and stress level, of input gating. As shown in Figure 1 for *gi*, the first time slice is flavored by both adjacent vowels (again representing coarticulation effects), which means

that there is no completely invariant input representation for any given demisyllable. The second and third time slices of the demisyllable are invariant, however, and the combined effect of the stress and error signals is to filter the input so that only these invariant time slices reach the classification net. Thus, realistic properties of the input representation, together with the filtering effects of stress and error, lead the classification network to see only demisyllables as its input.

Loss of sensitivity to nonnative contrasts

A number of developmental speech perception results have been described for the period between 6 and 12 months of age, relating to a loss of sensitivity to nonnative language contrasts. For example, it has been found that English-learning infants aged 6-8 months are able to discriminate Hindi and certain other nonnative contrasts, while infants aged 10-12 months are mostly unable to do so, as are adult native speakers of English. However, English-learning infants at all ages, as well as adults, retain the ability to discriminate certain other nonnative contrasts, such as that between two Zulu clicks (Werker, 1991). Part of Werker's account of these phenomena is that both sounds in the Hindi contrast (involving a dental vs. a retroflex [ta]) may by the later age have become assimilated to the native English alveloar [ta], and thus ceased to be discriminable. The Zulu clicks, on the other hand, may not be easily assimilable to any known category, and hence remain discriminable.

To examine the model's responses to unknown sounds, a "dental" and "retroflex" *t* were simulated by modifying the value of the "coronality" feature, from 1.0 for the alveolar *t*, to 0.7 and 0.3 for the dental and retroflex versions respectively. These were used to create "nonnative" stimuli: a dental *ta*, which will be denoted by *t(d)a*, and a retroflex *ta*, denoted by *t(r)a*. The syllables *na* and *nga* were treated as a second nonnative contrast, since neither of them was included in the input corpus.

As mentioned above, adjustment of weights in the classification network began after the overall system had been exposed to the input corpus for 80 epochs, and continued for 20 epochs. Just prior to the start of this training (i.e., at 80 epochs), the overall system was tested on the nonnative stimulus set[6]. This is meant to simulate testing of the infant's abilities at age 8 months. Results from the CL classification network are shown in the left-hand part of Table 2. As shown, the nonnative stimuli are responded to by different units, indicating their discriminability.

The overall system was tested again on the nonnative stimuli, at the the end of the 20 epochs of training. Results from the CL classification network are shown in the right-hand part of Table 2. The same unit now responds to *ta*, *t(d)a*, and *t(r)a*, indicating that the nonnative stimuli have been assimilated to the known *ta* category. The *na* and *nga* stimuli are still responded to by different units, however, indicating that they are not assimilable to known categories.

Before training				After training			
ta	153	*ma*	92	*ta*	153	*ma*	92
t(d)a	153	*na*	149	*t(d)a*	153	*na*	100
t(r)a	50	*nga*	204	*t(r)a*	153	*nga*	39

Table 2: Responses of CL classification network to nonnative stimuli.

Equivalent results were obtained with the DBM, but not the MLP architecture. Output responses of the DBM classification network were projected onto the first two principal components. Before training, the network's responses were quite widely separated in state space, indicating discriminability of all the stimuli. After training, however responses to the stimuli were much less dispersed in state space. However, *na* and *nga* were considerably further dispersed than *t(d)a* and *t(r)a*. These results are analogous to those obtained with the CL architecture. With the MLP architecture, however, the opposite trend appeared: responses to the stimuli were more widely dispersed after training than before training.

These results can also be examined in terms of the average pairwise distance between members of the *ta–t(d)a–t(r)a* and *ma–na–nga* triples. With the DBM, the ratio of this average distance *after* training to the average distance *before* training was 0.42 for the stops, and 0.57 for the nasals, illustrating that discriminability had decreased for both groups, but more so for the stops. With the MLP architecture, however, the after-before ratio was 2.77 for the stops and 15.01 for the nasals, indicating that the members of each group had become *more* discriminable after training.

The results obtained with the CL and DBM architectures demonstrate lost sensitivity to certain nonnative contrasts as well as retained sensitivity to certain other nonnative contrasts. This not only simulates the observed developmental phenomena, it also provides a computational account of such a process, and thereby a basis for understanding why the observed selective loss of nonnative contrasts in infants might arise. As the perceptual ("classification") system develops, it becomes attuned to, and begins to categorize, sounds occurring in the environment. Other (nonnative) sounds now tend to be interpreted in terms of the categories developed for known, occurring sounds.

The DBM is an "attractor" network, in which the learned states represent basins of attraction. This means that inputs similar to those that have been learned will tend to result in one of these attractor states. The CL classification network discretely approximates this property of the DBM, in that an input is mapped to the output unit with most closely similar weights. The fact that the loss of nonnative contrasts is simulated with the CL and DBM architectures, but not the MLP architecture is therefore interesting, suggesting that the formation of attractor states is necessary to simulate this developmental trend. The MLP does not form attractors, and is therefore unable to capture this phenomenon.

[6]The syllables *ta* and *ma* were included in the testing, for purposes of comparison, being the closest trained stimuli to the *t(d)a-t(r)a* and *na-nga* contrasts, respectively.

Discussion

The present work provides a number of interesting demonstrations, in the context of two selected phenomena from infant speech perception. First, the simulations provide a good account of the selected developmental phenomena, and thus suggest computational mechanisms that could implement these processes. The simulations further suggest that the development of attractor states may be necessary in modeling these phenomena. Second, the simulations provide specific suggestions about the nature of early phonological representations; in particular, they suggest that these representations may be organized around demisyllable-sized units. Third, the work suggests that a simple mechanism utilizing information about predictive error and stress level can implement temporal integration over a demisyllable-sized window. This is interesting in two ways: (i) it provides a means of extracting static representations from a dynamic and nonlinear signal; (ii) although the significance of both the syllable and of stress in speech perception has long been noted (Gleitman et al., 1988), the present work provides a specific demonstration of how stress information might be utilized in generating syllable-like perceptions.

Of course, much of the interest of such results lies in what they might reveal about the nature of phonological processing of more complex word forms, involving, for example, consonant clusters. It is unclear whether or how the present model would scale up to dealing with these. Also, the classification network essentially reponds to words by spelling them out as demisyllables, and it is not clear whether this provides a realistic account of syllabic sensitivities outside of the experiments considered here. These are obvious questions for further research.

Acknowledgements

We would like to thank Jeff Elman, Brian MacWhinney, Jay McClelland, and Dave Touretzky for helpful discussion at various points, and Joe Levy and Nick Chater for providing the code used to generate phoneme features.

References

Adams, M. J. (1990). *Beginning to Read*. MIT Press, Cambridge, MA.

Catford, J. C. (1988). *A Practical Introduction to Phonetics*. Oxford University Press, New York.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Flores d'Arcais, G. B. (1988). Language perception. In (Newmeyer, 1988).

Gasser, M. (1992). Learning distributed representations. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ. Lawrence Erlbaum.

Gleitman, L. R., Gleitman, H., Landau, B., and Wanner, E. (1988). Where learning begins: Initial representations for language learning. In (Newmeyer, 1988).

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11:23–63.

Hinton, G. E. (1989). Deterministic Boltzmann learning performs steepest gradient descent in weight-space. *Neural Computation*, 1:143–150.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In Perkell, J. S. and Klatt, D. H., editors, *Invariance and Variability in Speech Processes*. Lawrence Erlbaum, Hillsdale, NJ.

Jusczyk, P. W. (1992). From general to language-specific capacities: The WRAPSA model of how speech perception develops. Manuscript.

Kaye, J. D. (1989). *Phonology: A Cognitive View*. Lawrence Erlbaum, Hillsdale, NJ.

Kohonen, T. (1984). *Self-Organization and Associative Memory*. Springer-Verlag, Berlin.

Kohonen, T. (1991). The "Neural" Phonetic Typewriter. In Carpenter, G. A. and Grossberg, S., editors, *Pattern Recognition by Self-Organizing Neural Networks*. MIT Press, Cambridge, MA.

McClelland, J. L. and Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. In McClelland, J. L., Rumelhart, D. E., and the PDP Research Group, editors, *Parallel Distributed Processing*, volume 2: Psychological and Biological Models. MIT Press, Cambridge, MA.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29:144–178.

Menyuk, P., Menn, L., and Silber, R. (1986). Early strategies for the perception and production of words and sounds. In Fletcher, P. and Garman, M., editors, *Language Acquisition*. Cambridge University Press, Cambridge, England, 2nd edition.

Newmeyer, F. J., editor (1988). *Language: Psychological and Biological Aspects*, volume III of *Linguistics: The Cambridge Survey*. Cambridge University Press, Cambridge, England.

Norris, D. (1990). A dynamic-net model of human speech recognition. In Altmann, G. T. M., editor, *Cognitive Models of Speech Processing*. MIT Press, Cambridge, MA.

Peterson, C. and Anderson, J. R. (1987). A mean field theory learning algorithm for neural nets. *Complex Systems*, 1:995–1019.

Rumelhart, D., Hinton, G., and Williams, R. (1986a). Learning internal representations by error propagation. In (Rumelhart et al., 1986b).

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1986b). *Parallel Distributed Processing*, volume 1: Foundations. MIT Press, Cambridge, MA.

Rumelhart, D. E. and Zipser, D. (1986). Feature discovery by competitive learning. In (Rumelhart et al., 1986b).

Shillcock, R., Lindsey, G., Levy, J., and Chater, N. (1992). A phonologically motivated input representation for the modeling of auditory word perception in continuous speech. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 408–413, Hillsdale, NJ. Lawrence Erlbaum.

Werker, J. F. (1991). The ontogeny of speech perception. In Mattingly, I. G. and Studdert-Kennedy, M., editors, *Modularity and the Motor Theory of Speech Perception*. Lawrence Erlbaum, Hillsdale, NJ.