Revealing the Structure of NETtalk's Internal Representations

Charles R. Rosenberg
Cognitive Science Laboratory, Princeton University,
Princeton, New Jersey 08542, USA

(April 3, 1987)

Abstract

NETtalk is a connectionist network model that learns to convert English text into phonemes. While the network performs the task with considerable accuracy and can generalize to novel texts, little has been known about what regularities the network discovers about English pronunciation. In this paper, the structure of the internal representation learned by NETtalk is analyzed using two varieties of multivariate analysis, hierarchical clustering and factor analysis. These procedures reveal a great deal of internal structure in the pattern of hidden unit activations. The major distinction revealed by this analysis of hidden units is vowel/consonant. A great deal of substructure is also apparent. For vowels, the network appears to construct an articulatory model of vowel height and place of articulation even though no articulatory features were used in the encoding of the phonemes. This interpretation is corroborated by an analysis of the errors or confusions produced by the network; The network makes substitution errors that reflect these posited vowel articulatory features. These observations subsequently led to the discovery that articulatory features of place of articulation and, to some extent, vowel height, are largely present in first-order correspondences between vowel phonemes are their spellings. This work demonstrates how the study of language may be profitably augmented by models provided by connectionist networks.

Introduction

Speech synthesis is the translation of written text into an acoustic speech signal. In most speech synthesis systems, two distinct knowledge sources are used in the determination of the pronunciations of words: rules which encode regularities, and a dictionary of exceptions to those rules which handle those cases where the regularities break down. When a correspondence can be predicted on the basis of regularities operating at some level, it can be encoded more efficiently as a generative rule, such as the rule that when the letter "c" occurs before a high, front vowel, typically spelled with the letters "i", "e", or "y", it is pronounced like an "s" as in "icy" or "center" but like a "k" in other contexts. But even the best letter-to-sound rules have exceptions, and so the rules must be augmented by a dictionary which is checked before any rules are applied. This dictionary is generally used with great frequency because the most frequent words in English are also the most irregular.

NETtalk is a connectionist network model which learns the pronunciation of English text, represented as phonemes, or distinctive speech sounds (Sejnowski & Rosenberg, 1986; 1987a; 1987b). No rules of pronunciation were provided to NETtalk. Rather, the network reaches a reasonable level of performance by being presented with a number of training examples and being incrementally corrected using the back-propagation rule (Rumelhart, Hinton & Williams, 1986).

Is NETtalk's knowledge of pronunciations divided up in this way, between dictionary-like knowledge and rule-like knowledge? Since all knowledge shares the same architectural/representational space in NETtalk, an answer to that question is not immediately obvious. However, there is some indication that both exceptions and regularities are learned. For example, it learns correctly that "of" is pronounced /xv/ (see Appendix), even though the letter "f" is not pronounced this way in any other case. On the other hand, when trained on a 16,000 word selection from the entire 20,000 word Webster's Pocket dictionary, the network is able to generalize to the rest of the corpus with an accuracy exceeding 90% correct phonemes. Moreover, the confusions are usually between phonemes that are phonologically similar. This ability to generalize to novel words indicates that the network does much more than memorize specific input-output pairs as found in a dictionary. But what exactly *are* the regularities that it discovers in English pronunciation?

Many network models to date have been simple enough so that the functionality of many of the hidden units could be discovered through direct visual inspection of the unit activations and weight values. In larger networks, where one may have hundreds of hidden units and tens or hundreds of thousands of weights, it becomes difficult or impossible to detect underlying structures by such direct methods. We must turn to more sophisticated techniques.

Analytical tools

Factor analysis

Factor analysis (eg. Harman, 1976; Rummel, 1967) is a well-known technique for attempting to account for the variance in a large number of variables in terms of a much smaller number of relatively independent underlying factors. Factor analysis is based on the assumption of a linear model, meaning that the observed variables must be predicted by a linear, weighted, combination of underlying factors. Specifically, a factor is defined as

$$F = \sum_{i=1}^{k} w_i X_i,$$

where the w's are the factor weights (to be estimated from the data) and the X's are the k original variables. The factor loadings are the correlations between the final factors and the original variables.

Factors are determined in two stages. The initial factor extraction is based on the method of principle-components analysis. The first principle-component is that weighted combination of variables that accounts for the greatest amount of the total variance in the data. The second principle-component accounts for the greatest amount of variance not accounted for by the first principle-component, and so on. The principle-components are chosen so as to be mutually uncorrelated or statistically independent, but are typically hard to interpret. Consequently, a second process is generally performed where these initial factors are rotated.

Cluster analysis

A technique which makes fewer assumptions about the underlying form of the data is cluster analysis (eg. Everitt, 1974). Here, items are progressively grouped or clustered together based on relative similarity within a cluster, and relative dissimilarity between clusters. The method is very simple: Given some matrix of similarities, cluster analysis iteratively merges the two most similar clusters. Of course, there are many ways to determine which groups are most similar. The three most common methods are "centroid", where the distance between two groups is defined as the average of

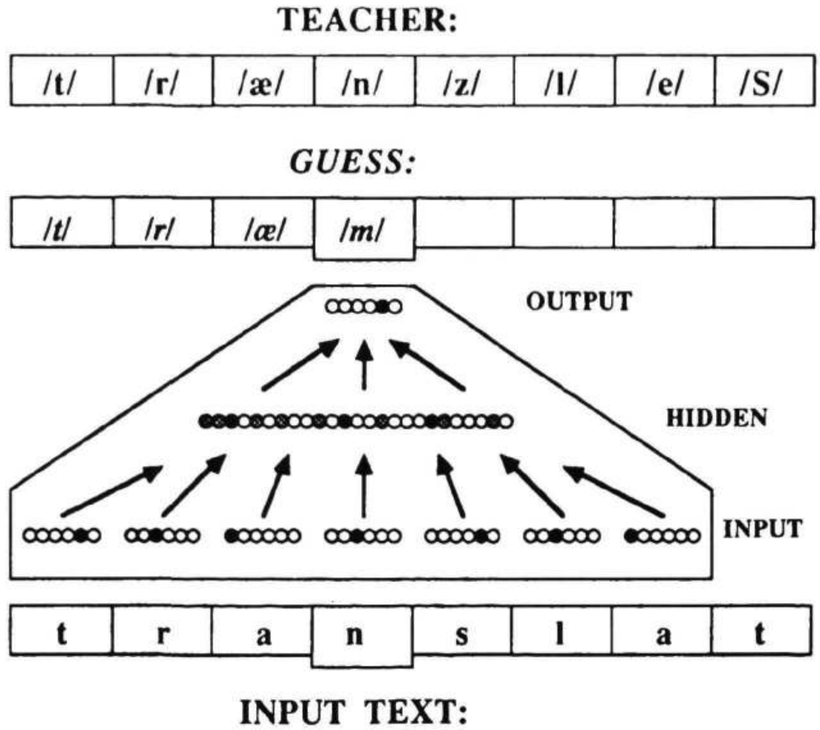

Figure 1: The Architecture of NETtalk.

distances between all pairs of members in each group, "complete linkage" or "farthest neighbor", where the two groups are merged that have the nearest most remote members, and "single linkage" or "nearest neighbor", where groups are merged on the basis of their nearest members. The resulting clusters can be graphed as a dendogram, and cuts through the dendogram yields the groups formed at that particular depth or distance.

A Brief Overview of NETtalk

NETtalk is composed of simple processing units arranged into layers. In the experiments reported here, I used an architecture with three layers of units and two layers of modifiable weights completely connecting successive layers of units (see Fig. 1). The activation of a unit in the network can take any value from zero to one. The connection strengths or "weights" can have any value, positive or negative. Each unit computes a weighted sum of unit activations times the weights from units in the layer below and determines its output value according to a nonlinear function that is zero for large negative inputs and monotonically approaches one for large positive inputs. (Hopfield, 1984; Rumelhart et al., 1986). Thus, information travels through the network, from the input layer to the intermediate layer of "hidden" units and finally to the output layer. The input layer receives information from an English text and the network produces a pattern of activation at the output layer of units that is a representation of how the input text should be pronounced. The representations on the input and output layers are fixed, but those of the hidden units is constructed by the learning procedure. The representations formed at this hidden layer are the focus of the present work.

The pronunciation of a given letter is generally influenced by the surrounding letters. Letter context was represented in the network by extending the input layer out over seven letters, where each of the seven letter positions was represented by separate groups of input units (see Fig. 1). At each sequential step, the network received input simultaneously from all the letters that fell within the fixed-sized input window. Based on this information, a guess was made as to the pronunciation of the middle letter of the window in two steps: 1. the values of the output units were determined by forward-propagating the activation of the input units through the network to the output layer, then 2. the sum of the squares of the differences was computed between the output units and each of the possible phoneme targets. The phoneme which minimized this distance was chosen as the guess made by the network. [1]

Changes were computed for the weights following each forward-propagation step, but these changes were actually incorporated into the weights only between words in order to reduce computation time. The output units were compared, unit-by-unit, with the correct phoneme supplied to the network, and the weight changes for internal connections were computed by recursively applying back-propagation from the output layer to the input layer. The seven-letter input window was then moved down one letter position in the input text, and the process repeated. When the end of the corpus was reached, the network continued at the beginning. (See Sejnowski & Rosenberg, 1986, 1987b for details on NETtalk and Rumelhart et. al, 1986 for details on the error back-propagation algorithm.)

[1] This procedure for selecting the phoneme is slightly different from that previously reported (Sejnowski & Rosenberg, 1986, 1987b).

The Analyses

These multivariate techniques were applied to the analysis of the patterns of hidden unit activities observed in NETtalk, where phonemes constituted the variables of the analysis. For each phoneme, average hidden unit activations were collected over all letter contexts. Since there were 80 hidden units in this version of NETtalk, the representation of each phoneme at the hidden layer was cast as an 80-dimensional vector. Pair-wise correlations between these vectors were computed and this correlation matrix was submitted as similarity data to cluster analysis and factor analysis. If phonemes were encoded as orthogonal dimensions in this space, then this correlation matrix would have one's along the diagonal (phonemes are correlated with themselves), but zero's everywhere else. On the other hand, if there is structure in the way the network represents phonemes, the phonemes might be found to cluster together in some way.

Initial Training

A network with a single hidden layer of 80 units was trained on an eleven thousand word selection from the full 20,000 word Websters Pocket Dictionary. Phonemes and letters were both encoded using a local encoding, where a single unit in each input and output group encoded a single letter or phoneme. There were seven groups of 29 input units per group and a single group of 55 output units with complete connectivity between layers. The network made ten complete passes through this corpus, thus being trained on a total of 160,000 words, of which 16,000 were different words. The order of the words was randomized. Performance at this point in learning was 92% correct phonemes and 56% fully correct words on the training corpus. From the remainder of the 20,000 word corpus, 1000 words were selected at random. These words were not included in the training set. Learning was turned off and the network was tested on this new corpus. 90% of the phonemes and 49% of the words were completely correct.

Data Collection

With learning turned off, data was collected from this network as it went through this 1000-word corpus of novel words. For each of 48 phonemes (the three "special" symbols, /./, /-/, and /_/, and four other phonemes /!/, /K/, /q/, and /|/ were dropped, the latter because they did not occur in the 1000 word corpus), hidden unit activations were averaged over all input windows that had that phoneme as it's pronunciation. Trials where the network failed to guess the phoneme correctly were not included in the means. Since the network guessed 90% of the phonemes correctly, roughly ten percent of the trials were thrown-out for this reason. [2] Thus, an 80-dimensional vector of mean activation values was constructed for each phoneme. The sample correlation coefficient, r, then was computed for each pair of phonemes, resulting in a 48 x 48 correlation matrix. Each cell in this matrix represented the correlation of the pattern of hidden unit activity that resulted for the ith and jth phoneme.

Results from Hierarchical Clustering

A hierarchical clustering of all the phonemes using the centroid clustering method is shown in Fig. 2, where the dissimilarities, were defined as $1 - r$. This analysis revealed two major clusters, corresponding to vowels and consonants. All phonemes on the left major branch are vowels, and nearly all of the phonemes on the right-side are consonants, with the exceptions of /*/ and /y/. Centroid clusterings are shown, though similar results were obtained using complete linkage clustering.

Different organizational schemes are apparent within the two major groups. Within the vowel group, the next major division is between the "front" vowels, /i/, /I/, /E/, /e/, and /@/, where the sounds are produced with the tongue towards the front of the mouth, and the "back" vowels, /u/, /U/, /a/, /c/, and /o/, produced more towards the rear of the mouth. Next, the central vowels, /x/ and /^/ split off from the "back" group. Thus, the major organizational principle within the vowel cluster appears to be place of articulation. The next major division is based on the *height* of the tongue when the sound is produced ("vowel height"), as the "high" phonemes, /i/ and /I/ split off from the "low" phonemes /e/ and /@/, with the mid-vowel, /E/, falling between the two groups. The same phenomena occurs for the back vowels, where the low /o/ and /a/ group are distinguished from the high /U/ and /u/.

The consonant cluster appears to quickly divide into anywhere from five to ten major groups. Unlike the vowels, the consonant clusters correspond to place of articulation only weakly. Rather, they seem to cluster around typical input letters. For example, one major grouping consists of the phonemes /T/, /D/, /C/, /S/ and /t/: these are the phonemes that the letter "t"

[2] However, the pattern of results does not change significantly if this requirement is relaxed.

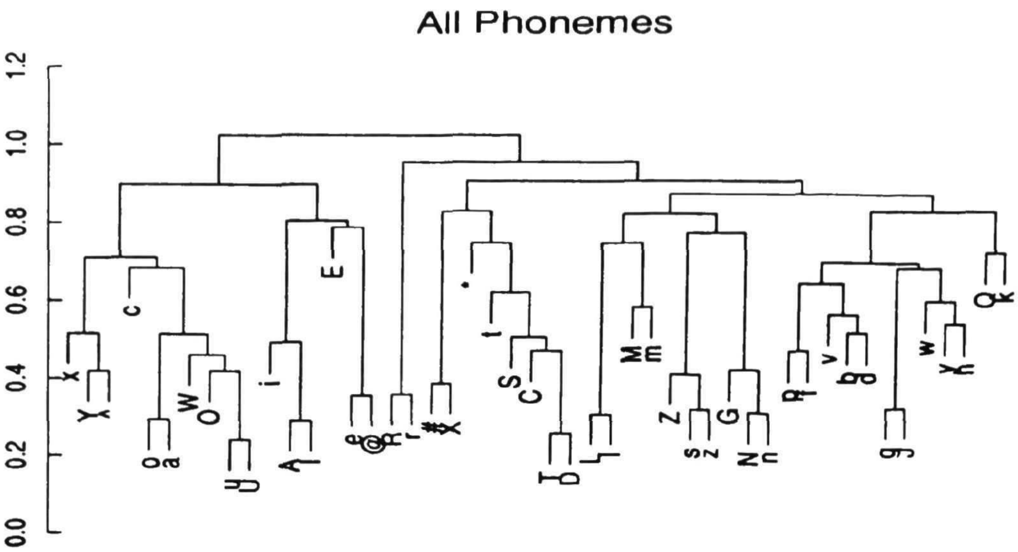

Figure 2: Hierarchical clustering results for 48 phonemes.

typically corresponds to. Another grouping are the phonemes /#/ and /X/. These are the phonemes for the letter "x". With few exceptions, the same pattern follows for the "m" group (/M/, /m/), the "s" group (/s/, /z/, /Z/), the "n" group (/n/, /N/, /G/), the "p" group (/f/, /p/), and the "g" group (/g/, /J/).

Results from Factor Analysis

A set of ten non-dipthong vowels were selected for factor analysis. Since dipthongs, such as /A/ in the word "bite", involve a change in the place of articulation and height during the course of their pronunciation, it was not desired to complicate matters by including these more ambiguous vowels in the data set. A 10 x 10 correlation matrix was prepared and submitted to a Varimax rotated orthogonal factor analysis. Three factors were extracted, which together accounted for 68% of the total variance. The rotated factors are presented in Table 1. Factor loadings less than 0.55 are generally considered unreliable (Comrey, cited in Kim & Mueller, 1978). Loadings less than 0.25 were deleted in Table 1.

The results from the factor analysis in general confirmed the analysis based on clustering, that the vowels organized according to place of articu-

SORTED ROTATED FACTOR LOADINGS

PHONEME	FACTOR 1	FACTOR 2	FACTOR 3
c /cAUght/	0.805		
a /fAther/	0.771	0.345	
@ /bAt/	0.746	-0.266	0.372
u /bOOt/		0.909	
U /bOOk/		0.901	
^ /bUn/	0.469	0.516	
i /bEAt/			0.808
I /bIt/			0.788
x /womEn/	0.411	0.278	0.676
E /bEt/			0.499
Var explained	23.04%	22.42%	22.37%
Cumm var	23.04%	45.46%	67.83%

Table 1: Three factors and accompanying factor loadings extracted from a Varimax rotated orthogonal factor analysis employing 10 vowels as variables. Zero factor loading was 0.25.

lation and vowel height. The three vowels that loaded highest on the first factor are the low vowels, /c/, /a/ and /@/, which suggests the interpretation of the first factor as a "lowness" factor. /u/ and /U/ both loaded high on Factor 2, leading to the interpretation that this factor relates to "backness". /I/ and /i/ both load high on the third factor, suggesting that this factor describes "frontness". A bit of an anomaly is the high loading of the schwa /x/ on this "frontness" dimension. Thus, Factor 1 was interpreted as representing vowel height, but two factors, Factors 2 and 3, represented place of articulation.

It was desired to collapse these three factors to two dimensions so that they could be compared to known relationships between place and height. Standard values for place of articulation and vowel height can be found in any introductory linguistics textbook. The set of values used for the purposes of this analysis are plotted in Fig. 3.

Using least-squares linear regression, weights were found to predict these true values of place of articulation and vowel height based on the factor loadings on the three factors for each vowel. Excellent fits were found in both cases. For "place", $r = 0.96$, $F(3,6) = 25.37$, $p < 0.001$, and for

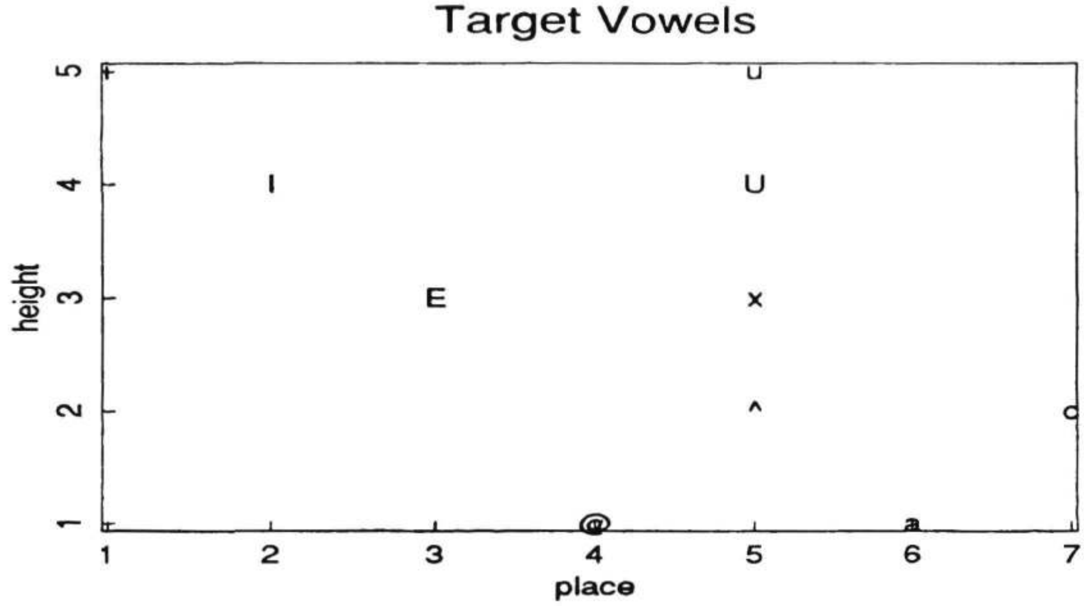

Figure 3: "Place of articulation" and "vowel height" attribute values for the set of ten vowels used the analysis.

"height", $r = 0.96$, $F(3,6) = 23.41$, $p < 0.005$. These two dimensions were not orthogonal, but were negatively correlated with each other, $r = -.079$. The factor loadings were then substituted back into the regression equations and the derived estimates for "place" and "height" are plotted in Fig. 4.

Thus, the pattern of results from the factor analysis largely agreed with the results based on hierarchical clustering: the representation of vowels at the hidden layer of NETtalk appears to be organized around two articulatory features of vowels, place of articulation and vowel height. In addition, the clustering of all of the phonemes suggests that the vowels and consonants are represented by distinctly different patterns of hidden unit activation.

Results from Confusion Data

If the internal representations of vowels and consonants are structured in the way suggested by the previous analysis of the hidden units, then these patterns ought to be reflected in the overt behavior of the network. For example, if vowels and consonants are as distinguishable as they appear to be, then the network should rarely confuse them, by guessing a consonant when the target was a vowel, and visa versa. We might also expect the structure of

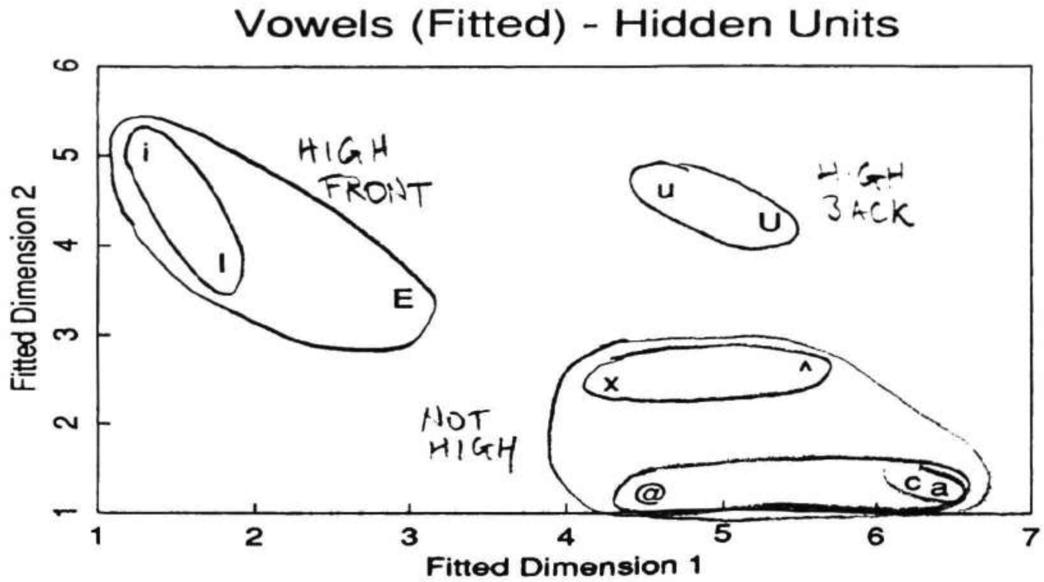

Figure 4: Fits based on linear regression of the three factor solution on the attribute dimensions "place of articulation" and "vowel height", plotted in Fig. 3. Data based on hidden unit activations. The groupings are based on a centroid hierarchical cluster analysis of the original correlation matrix.

the internal representation of vowels to be reflected in the pattern of confusions displayed, so that more confusions should be apparent between nearby vowels in articulatory space, than between more distant vowels. Of course, this would improve the intelligibility of NETtalk, since mistakes would be limited to similar phonemes. The following experiment was designed to test this prediction.

With learning turned off, the network went through the same 1000-word corpus used in the previous experiment. For each of the 48 phonemes examined previously, the guesses made by the network were recorded. These responses were cast as a 48 x 48 confusion matrix, where the rows were the target phonemes and the columns were guesses or responses made by the network to that target. The fraction of the time the network guessed phoneme j in response to target i was recorded in the ith x jth cell of the matrix. Correlations were computed between each of the target phonemes (rows) of the confusion matrix, and the resulting correlation matrix was clustered using the "centroid" or "average" clustering procedure. The results are shown in Fig. 5. Similar results were found using the complete

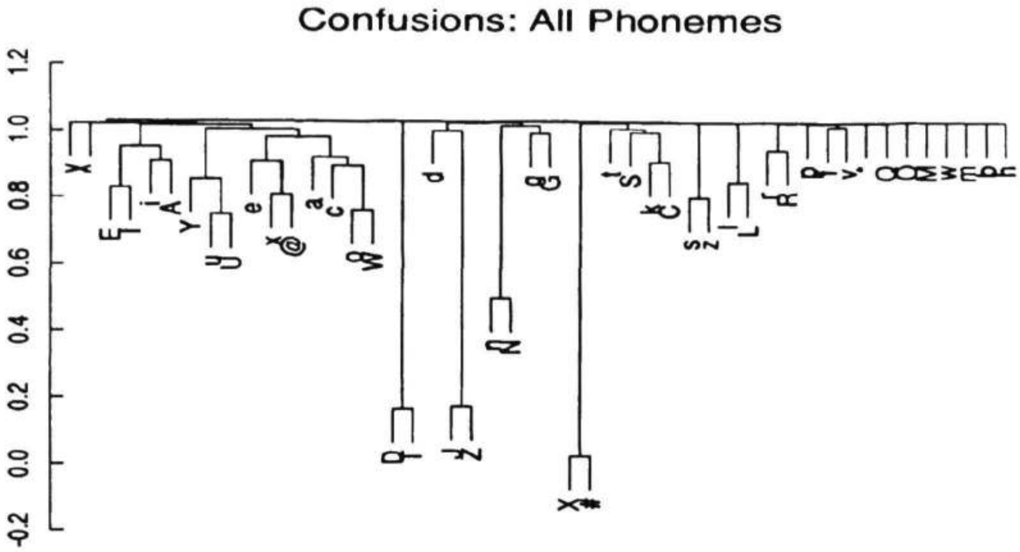

Figure 5: Hierarchical cluster analysis of ten vowels, based on the pattern of confusions produced by NETtalk.

linkage clustering procedure.

Besides four pairs of consonants that were consistently confused, /D/ and /T/, /J/ and /Z/, /n/ and /N/, and /X/ and /#/, nearly all of the dissimilarities are quite high - the network simply did not confuse very many of the phonemes. This lack of structure makes any analysis difficult. However, more confusions were made on the vowels than the consonants, and some structure is apparent. Three main clusters form, conforming fairly closely to "front" (/i/, /E/, /I/), "back" (/u/, /U/), and "not-high" (/x/, /@/, /a/, /c/, /o/). The dipthongs, /W/, /e/, /Y/, /A/ and /O/, are more difficult to interpret for reasons given previously. The "not-high" group then appears to split further into what might be called a "back, not-high" group (/a/, /c/, /o/) and a "not-back, not-high" group (/@/, /x/).

A factor analysis extracted a set of four factors which are very similar to the groups observed in the cluster analysis. As before, least-squares linear regression analysis was used to predict vowel height and place of articulation from the factor loadings. The fitted dimensions are plotted in Fig. 6. Fits to the attributes were not quite as good as in the hidden unit analysis. The correlation, r, between the fitted dimension for "place" and the attribute vector was 0.87, $F(4,5) = 3.92$. For "height", $r = 0.85$, $F(4,5) = 3.35$. Both

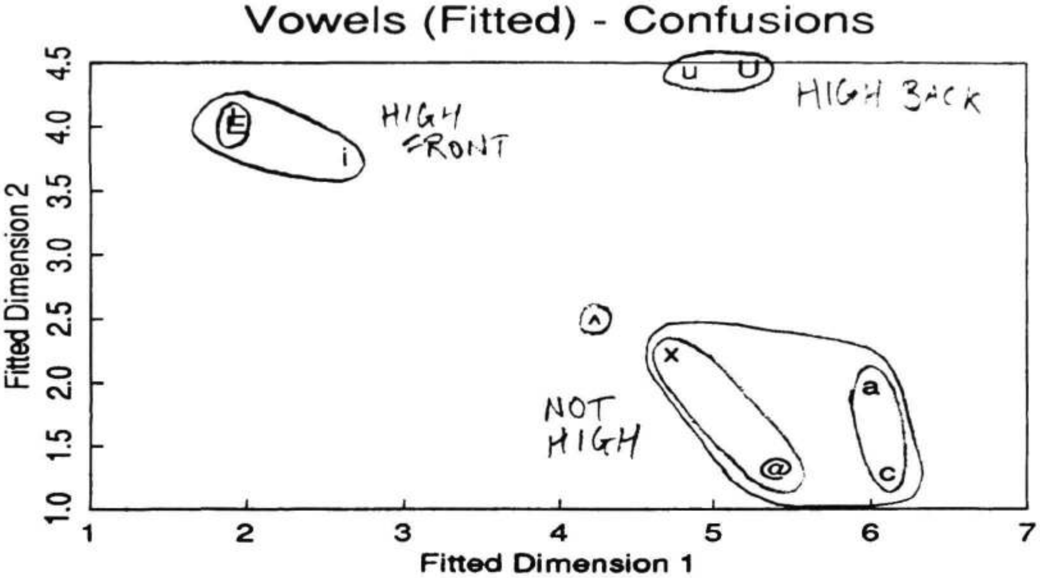

Figure 6: Fits of "place of articulation" and "vowel height", based on linear regression of the four factor solution on the attribute dimensions plotted in Fig. 3. Data was based on confusions. The groupings are based on a centroid hierarchical cluster analysis of the original correlation matrix.

F-tests narrowly failed to reach significant levels.

The overall pattern of results from the confusion data is similar to that found previously based on hidden unit activations: 1. vowels tend to be confused with vowels, and consonants with consonants, and 2. vowels that were similar to each other in terms of place of articulation and height were more likely to be confused.

Letter-to-Phoneme Correspondences

Why should vowels organize around these articulatory dimensions? Apparently, there is a "map" of articulatory features hidden in letter-to-phoneme correspondences for vowels, a map which does not exist (at least not to the same extent) for consonants. Upon closer inspection of first-order letter-to-phoneme correspondences, this was found to be true.

First-order correspondences between vowel letters and their pronunciations in the Websters corpus were investigated in much the same way as the hidden activations and confusions were analyzed previously. The full

	a	e	i	o	u	y	other
a	32.5	0.6	0.2	66.7	0.0	0.0	0.0
c	42.4	0.0	0.0	57.6	0.0	0.0	0.0
i	0.0	35.4	25.4	0.0	0.0	39.1	0.0
u	0.0	3.7	0.0	45.7	50.5	0.0	0.0
x	24.8	15.0	20.6	25.7	13.5	0.5	0.0
E	6.2	93.6	0.0	0.0	0.1	0.0	0.1
I	2.1	17.9	76.9	0.1	0.1	3.0	0.0
U	0.0	0.0	0.0	50.0	50.0	0.0	0.0
@	99.9	0.0	0.1	0.0	0.0	0.0	0.0
^	0.0	0.0	0.0	6.3	93.8	0.0	0.0

Table 2: Frequency with which the ten vowel phonemes are spelled by the letters "a", "e", "i', "o", "u", and "y".

Websters corpus of 20,000 words was searched for the percentage of times each of the ten vowel phonemes was spelled with each letter. The result was cast as a 10 x 7 matrix, where the rows correspond to the phonemes and the columns to their spelling. These percentages are presented in Table 2. As can be seen by inspection, there are obvious correlations between the phonemes in their spelling, and that these correlations agree, to some extent, with the articulatory dimensions of place and height. For example, the high-front vowels /I/ and /i/ both frequently correspond to the letters "e" and "i". The low-back vowels, /u/ and /U/ both frequently correspond to "o" and "u". Performing a factor analysis on this data, again using Varimax orthogonal rotation, four factors were found to account for 89% of the variance. As before, linear least-squares regression was performed in order to predict the attribute dimensions of "place of articulation" and "vowel height". "Place" was predicted well, the correlation between the attribute values and the predicted values being 0.95, $F(4,5) = 12.98$, $p < 0.001$. "Height" was predicted less well, $r = 0.79$, $F(4,5) = 2.06$, $p < 0.25$. Estimates for "height" and "place" based on the regression of the factor loadings on the attribute dimensions are plotted in Fig. 7.

These regularities inherent in these first-order letter-to-phoneme correspondences go far in explaining the previous hidden unit and confusion results: Similar phonemes, where similarity is based on place of articulation and vowel height, overlap in the letters with which they are typically spelled. This "map" of articulatory features is hidden within letter-to-phoneme correspondences for vowels. However, "height" was not as well represented in this first-order data as it was in the previous analyses using NETtalk. This suggests that that vowel height, unlike place of articulation, may require information from several letters to be adequately predicted. If true, then

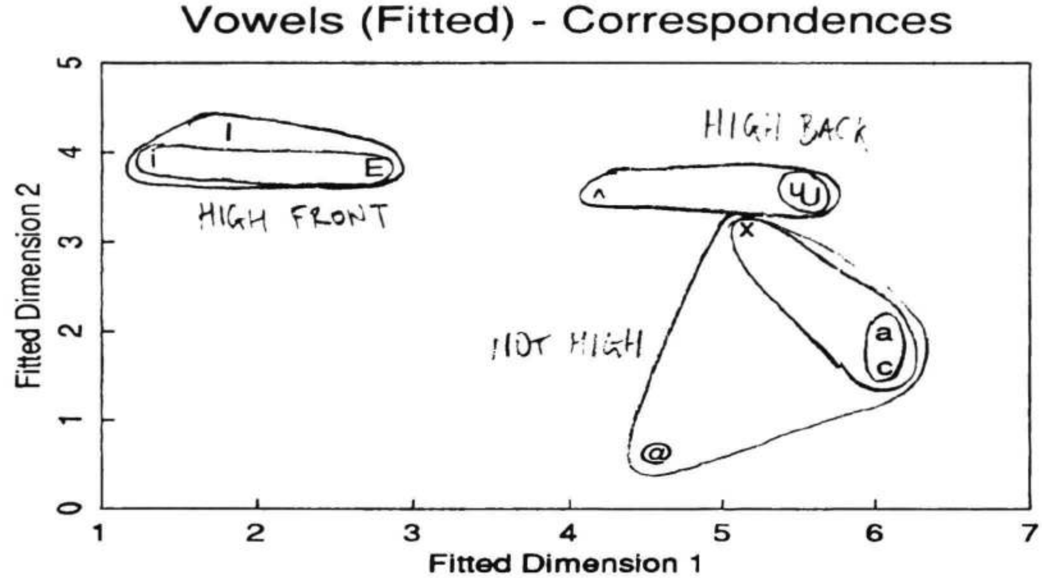

Figure 7: Fits of "place of articulation" and "vowel height", based on linear regression of the four factor solution on the attribute dimensions plotted in Fig. 3. Data was based on correspondences between phonemes and letters. The groupings are based on a centroid hierarchical cluster analysis of the original correlation matrix.

networks with only a single input group and output group ought to exhibit a similar disadvantage in the acquisition of vowel height.

Conclusions and Directions for Future Research

Summing up, vowels and consonants are encoded at the hidden layer of NETtalk as distinctly different patterns of activation. Within these two categories, vowels organize according to place of articulation and vowel height, whereas the consonants seem to organize around input letter. The pattern of confusions or errors produced by the network are similar to those observed at the hidden layer.

These observations led me to go directly to the training corpus, the Webster's Pocket dictionary, to investigate letter-to-phoneme correspondences for similar regularities. Surprisingly, articulatory features were found to be largely present in first-order correspondences between phonemes and letters; It is possible to reconstruct a fairly accurate map of vowel height and place of articulation using distances derived from the dissimilarities between the vowels in terms of their spellings. In short, similar vowel phonemes are spelled in similar ways. To my knowledge, this regularity has not been described previously but can be seen as a useful property of language since it makes it possible to make a good guess about the spelling of a word by "sounding it out" phonetically.

All of the hidden unit analyses reported here were based on averaging activity levels over all inputs for a given phoneme. A similar analysis has been performed for letter-to-phoneme correspondences (Sejnowksi & Rosenberg, 1987b), and even more could be learned by examining graphemes and other letter combinations at the hidden layer. For example, we have clustered hidden unit activation patterns in contexts in which the letter "c" is at the center, and have found regularities and subregularities in the coding scheme even for irregular cases (Sejnowski & Rosenberg, 1987a).

There is a great deal that could be learned about language and about representation from analyzing larger connectionist networks, once analytical tools have been appropriately adapted for this purpose. Judging from previous studies of small networks for problems such as XOR and the encoder problem, it is possible that NETtalk solves certain problems in letter-to-phoneme translation in elegant and perhaps novel ways. It also seems possible that similar kinds of techniques to those found useful in analyzing model systems such as NETtalk, may eventually be applicable to the understanding of how knowledge is organized in real neural systems.

Acknowledgments

I would especially like to thank Dr. Terrence Sejnowski, with whom I collaborated on the construction of NETtalk. I would also like to thank Drs. Stephen Hanson and George Miller, who helped me in thinking about many of the issues reported here, Drs. Ken Church, Mitch Marcus, Mark Liberman, Judy Kegl, and Chris Tancredi. Dr. Jerome Feldman suggested the use of local input and output representations, which simplified the analysis and interpretation of the hidden representations. Bell Communications Research generously provided computational support. Support was provided by a grant from the James S. McDonnell foundation to the Human Information Processing Group at Princeton University.

Appendix

Table of Phonemes

Phoneme	Sound	Phoneme	Sound
/a/	father	/C/	chin
/b/	bet	/D/	this
/c/	bought	/E/	bet
/d/	deb	/G/	sing
/e/	bake	/I/	bit
/f/	fin	/J/	gin
/g/	guess	/K/	sexual
/h/	head	/L/	bottle
/i/	Pete	/M/	absym
/k/	Ken	/N/	button
/l/	let	/O/	boy
/m/	met	/Q/	quest
/n/	net	/R/	bird
/o/	boat	/S/	shin
/p/	pet	/T/	thin
/q/	uh-oh	/U/	book
/r/	red	/W/	bout
/s/	sit	/X/	excess
/t/	test	/Y/	cute
/u/	lute	/Z/	leisure
/v/	vest	/@/	bat
/w/	wet	/!/	Nazi
/x/	about	/#/	examine
/y/	yet	/*/	one
/z/	zoo	/|/	logic
/A/	bite	/^/	but

Output representations for phonemes and punctuations. The symbols for phonemes in the first column are a superset of ARPAbet and are associated with the sound of the italicized part of the adjacent word. Compound phonemes were introduced when a single letter was associated with more than one primary phoneme. The continuation symbol, /-/, not shown here, was used when a letter was silent.

References

[1] Everitt, B., (1974). *Cluster Analysis.* Heinemann: London.

[2] Harman, H. H. (1976). *Modern Factor Analysis.* University of Chicago Press: Chicago.

[3] Hopfield, J. J. (1984). Neurons with graded response have collective computation abilities, *Proceedings of the National Academy of Sciences USA*, **81**, 3088-3092.

[4] Kim, J. & Mueller, C. W. (1978). *Introduction to Factor Analysis*, Sage University Paper series on Quantitative Appliciations in the Social Sciences, 07-013: Beverly Hills.

[5] Rumel, R. J. (1967). Understanding factor analysis, *Conflict Resolution*, **11**, 444-480.

[6] Rumelhart, D. E., Hinton, G. E. & Williams, R, J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland, (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, Cambridge: MIT Press.

[7] Sejnowski, T. J. and Rosenberg, C. R. (1986). *NETtalk: A parallel network that learns to read aloud* (Technical Report 86/01). Baltimore: Johns Hopkins University, Department of Electrical Engineering and Computer Science.

[8] Sejnowski, T. J., and Rosenberg, C. R. (1987a). Connectionist Models of Learning. In M. S. Gazzaniga (Ed.), *Perspectives in Memory Research and Training.* Cambridge: MIT Press.

[9] Sejnowski, T. J. and Rosenberg, C. R. (1987b). Parallel networks that learn to pronounce English text, *Complex Systems*, **1**, 145-168.