Learning novel sound patterns

Gary Jones, Fernand Gobet & Julian M. Pine
{gaj,frg,jp}@Psychology.Nottingham.AC.UK
School of Psychology; University of Nottingham,
Nottingham, NG7 2RD, England

Abstract

The acquisition of vocabulary represents a key phenomenon in language acquisition, yet it is still poorly understood. Gathercole and colleagues have recently provided a rigorous test of vocabulary knowledge (the nonword repetition test, Gathercole, Willis, Baddeley, & Emslie, 1994) and have adapted the phonological loop part of the working memory model (Baddeley & Hitch, 1974) to explain the nonword repetition findings (e.g. Gathercole & Baddeley, 1989). However, there are two major failings in their explanation: there is no description of how words are learned, and no definition of how the phonological loop interacts with long-term memory. We present an EPAM based computational model which overcomes these problems by combining the phonological loop approach with the EPAM/chunking approach (Feigenbaum & Simon, 1984). Trained on naturalistic phonemically coded speech (from mother's utterances to 2-3 year old children), the model provides a good match to the nonword repetition data from 2-3 year old children. The model is also able to show the effect on nonword repetition when the model is trained using different sets of input. Implementing the phonological loop within EPAM represents a parsimonious approach to learning novel sound patterns and provides a more precise definition of how vocabulary acquisition may occur.

Introduction

Children are remarkably adept at learning new verbal information. After an initial slow period from about 12 to 16 months when most children learn around 40 words, the learning rate increases such that in the next four months children will have learnt 130 more new words (Bates et al., 1994). By the beginning of their school years children learn up to 3,000 words per year (Nagy & Herman, 1987).

A major part of learning new words is learning the novel sequences of sounds that represent the word. However, it is difficult to directly examine the processes involved in learning the sound patterns of new words because it is impossible to be certain that the new sound pattern has never been encountered before. The use of *nonwords* (e.g. nate) which conform to the phonotactic rules of English provides a good test of vocabulary learning because it ensures that the (non)word to be learned is novel.

The Nonword Repetition Test

The nonword repetition (NWR) test (Gathercole, Willis, Baddeley & Emslie, 1994) was designed to investigate the role of phonological memory in word learning. The test involves the experimenter speaking a nonword and the child attempting to repeat the nonword. The nonwords in the test vary in length from one to four syllables. Gathercole and Baddeley (1989) found that the NWR test was a good predictor of vocabulary size, even after vocabulary scores (measured using the British Picture Vocabulary Scale, Dunn & Dunn, 1982) were partialled out of the correlation.

Gathercole and Baddeley (1990) used the NWR test to categorise children into two groups, those with low NWR scores, and those with high NWR scores. Children in the high NWR group were better at learning nonword labels than children in the low NWR group. Moreover, nonword repetition studies show the influence that vocabulary knowledge has upon the learning of new

words. For example, Gathercole, Willis, Emslie, and Baddeley (1991) found better NWR performance on nonwords that were rated high in wordlikeness than nonwords rated low in wordlikeness. Nonword length also influences performance: repetition accuracy decreases as the number of syllables in the nonword increases, excepting one-syllable nonwords (e.g. Gathercole & Adams, 1993; Gathercole, Willis, Emslie & Baddeley, 1991).

Based on these findings, nonword repetition ability would seem to provide a good test of phonological memory and is a good indicator of vocabulary size. Gathercole and colleagues have tried to explain the NWR findings within the framework of the working memory model.

The Phonological Loop Explanation of Nonword Repetition Findings

The working memory model (Baddeley & Hitch, 1974) has been adapted to account for vocabulary acquisition (e.g. Gathercole & Baddeley, 1989). The phonological loop part of the model is claimed to be a critical mechanism for learning new words. The phonological loop has two linked components: the phonological short-term store, and the sub-vocal rehearsal mechanism. Items in the store decay over time, and become inaccessible after around 2,000 ms (Baddeley, Thomson & Buchanan, 1975). The sub-vocal rehearsal mechanism (involving sub-vocal articulation in real-time) can refresh items in the store in a serial, time-based manner (Gathercole & Martin, 1996).

The phonological loop hypothesis is able to explain the basic nonword repetition findings involving nonword length because of the decay that takes place in the phonological store (items remain in the store for ≈2,000 ms unless refreshed). Longer nonwords take longer to rehearse and so their representations in the phonological store are not refreshed as often as shorter nonwords. Repetition ability will therefore decrease for longer nonwords.

Recent findings have shown that children do not use subvocal rehearsal until the age of seven (see Cowan & Kail, 1996, for a review), which would seem to jeopardize an explanation based on the phonological loop. However, Brown and Hulme (1996) have shown that nonword repetition phenomena can be explained solely by a decay-based model (the phonological store is a decay-based store). Although this paper will refer to the phonological loop, for children under the age of seven the loop is in fact solely the phonological store because children under seven years of age do not use the rehearsal mechanism.

The phonological loop is able to explain a lot of the vocabulary acquisition phenomena using a very simple mechanism. However, it fails in two critical areas: there is no explanation of how words are learned, and there is no explanation of how the loop interacts with long-term memory. Speculative explanations have been given as to how the loop may interact with long-term memory (e.g. Baddeley, Gathercole & Papagno, 1998; Gathercole & Baddeley, 1989). Gupta and MacWhinney (1997) have proposed a formal specification but their model has not yet been implemented computationally.

The phonological loop hypothesis also lacks precision because it is part of a verbal theory. For example, there is no definition of how long rehearsal takes based on the length of the sound pattern being rehearsed (except to say that long strings take longer to rehearse than short strings). Incorporating the loop mechanism within a computational architecture forces precision because the theory is implemented as a running computer program. While several computational implementations of the phonological loop exist (e.g. Brown & Hulme, 1996; Burgess & Hitch, 1992), none offer an explanation of how words are learned and none give precise definitions of the link between the phonological loop and long-term memory. An explanation of both how words are learnt and how the loop interacts with long-term memory is required for the phonological loop to provide a full account of vocabulary learning. EPAM is a computational modelling architecture which is able to provide such a specification.

Implementing the Phonological Loop Within the EPAM Architecture

EPAM has been successful in simulating various areas of human cognition, such as learning, memory, and perception in chess (De Groot & Gobet, 1996; Gobet, 1993; Simon & Gilmartin, 1973), verbal learning behaviour (Feigenbaum & Simon, 1984), and the digit-span task

(Richman, Staszewski & Simon, 1995), and is currently applied by our research group in domains such as the acquisition of syntax (Gobet & Pine, 1997) and the acquisition of multiple representations in physics (Lane, Cheng & Gobet, 1999).

The EPAM Architecture

EPAM learns by building a discrimination network which is a hierarchical structure consisting of nodes connected to one another by links (see Figure 1 for an example; in the Figure, nodes are represented by boxes and links by lines between boxes). Nodes contain information and links between nodes contain tests which must be fulfilled before they can be traversed. For the purposes of modelling the learning of sound patterns, EPAM has been simplified and is henceforth referred to as EPAM-VOC.

When an input is given to the network, EPAM-VOC traverses down the hierarchy as far as possible. This is done by starting at the top node (the root node) and selecting the first link whose test is fulfilled by the first part of the input. The node at the end of the link now becomes the top node and the rest of the input is applied to all the links below this node. When a node is reached where no further traversing can be done (because the input fulfils none of the tests of the node's links, or the node is a leaf node), EPAM-VOC compares the information at the node with the input information. Learning now occurs in two ways.

1. Discrimination. When the input information mismatches the information given at the node (the *image*), a new link (i.e. test) and node are added to the tree below the node that has just been reached. The new test will relate to the mismatched part of the input.

2. Familiarisation. When the input information is under-represented by the information at the node, new features (from the input) are added to the information in the node (the image). In EPAM-VOC, familiarisation is simplified such that the image of a node always consists of the information in the links that lead to the node.

Discrimination therefore creates nodes and links, and familiarisation creates or modifies the information contained in nodes. Examples of the discrimination and familiarisation learning mechanisms will be given later.

Learning Sound Patterns in EPAM-VOC

EPAM-VOC learns using utterances from mothers' speech. The mothers' utterances are converted into a sequence of phonemes before being used as input. This is done using the CMU Lexicon database (available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict) which cross-references words with their phonemic representations. The use of phonemic input assumes that some form of phonemic feature primitives already exist to distinguish one phoneme from another.

EPAM-VOC begins with a null root node. When an input (a sequence of phonemes) is seen, new nodes and links are created. At first, most of the new nodes and links will be for single phonemes. As learning progresses, the information at nodes will become sequences of phonemes and therefore segments of speech (e.g. specific words) rather than just individual sounds (i.e. phonemes). The EPAM learning mechanism is altered in two ways to accomplish this efficiently. First, before a sequence of phonemes can be learnt, the individual phonemes in the sequence must have been learnt. Second, when individual phonemes are learnt, they are linked to the root node (in this way all sequences of phonemes are below the node which represents the initial phoneme in the sequence).

Let us consider an example of the network learning the utterance "What about that?" with an empty network. Using the CMU Lexicon database, this utterance is converted to the phonemic representation "W AH1 T AH0 B AW1 T DH AE1 T" (all of the phonemes used in the database map onto the standard phoneme set for American English). Note that the phonemic input to the model does not specify gaps between words, but does specify the stress of particular phonemes (0=unstressed; 1=primary stress; 2=secondary stress).

The first part of the input ("W") is applied to all of the root node's links in the network. As the network is empty, there will be no links. At this point EPAM-VOC must discriminate because the information "W" mismatches the information at the root node (the root node information is

null). The discrimination process creates a new node, and a link from the root node to the new node with the test "W". EPAM-VOC must then familiarise itself with the input, in order to create the "W" information in the image of the node. When the input is seen for the first time, just the phoneme "W" is learnt.

When encountering the input for the second time, the link "W" can be taken, and the input can move to the next phoneme, "AH1". As node "W" does not have any links, then under normal circumstances discrimination would occur below the "W" node. However, the model has not yet learnt the phoneme "AH1", and so discrimination occurs below the root node. Familiarisation will then fill the image of the new node with "AH1".

The third time the input is seen, the "W" link can be taken, with the input moving on to "AH1". No further links are available, but the input "W AH1" does not match the information at node "W" and so discrimination occurs. A new node "W AH1" is created below node "W". Familiarisation will fill in the image of the new node. After two further presentations of the input, the network is as shown in Figure 1. The simple example serves to illustrate how EPAM-VOC works; in the actual learning phase each utterance line is only used once, encouraging a diverse network of nodes to be built.
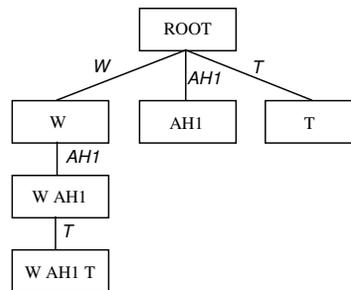

Figure 1: Structure of an EPAM-VOC net after seeing the input "W AH1 T AH0 B AW1 T DH AE1 T" five times.

Implementing the Phonological Loop and Linking it to Long-Term Memory

All of the nonword repetition test (the standard test of the phonological loop) studies used children below the age of six. As children below the age of seven are believed not to rehearse, the rehearsal part of the loop should not be used to simulate the nonword repetition findings reported earlier.

The storage part of the phonological loop is a decay-based store which allows items to remain in the store for about 2,000 ms. Consistent with this estimate, EPAM-VOC incorporates a time-limited store which allows 2,000 ms of input. The input is cut-off as soon as the time limit is reached, because there is no rehearsal to refresh the input representations.

The cumulative time required by the input helps to provide a theory of how the amount of information in the phonological loop is mediated by long-term memory. When an input is heard, long-term memory (the EPAM-VOC network) is accessed and the input is represented using the minimum number of nodes possible. Rather than the actual input being placed in the phonological loop, the nodes which capture the input are used. The length of time taken to store the input is therefore calculated based on the number of nodes that are required to represent the input. The time allocations are based on estimates from Zhang and Simon (1985), who estimate ≈400 ms to bring each node into the articulatory mechanism, and ≈84 ms to match each syllable in a node except the first (which takes 0 ms). As the input will be in terms of phonemes, with approximately 2.8 phonemes per syllable (based on estimates from the nonwords in the nonword repetition test), the time to match each phoneme in a node is 30 ms.

Using the example input "W AH1 T AH0 B AW1 T DH AE1 T" and the network as given in Figure 1, the actual input to the model will be "W AH1 T AH0 B AW1". The "W AH1 T" part of the input is represented by a single node, and is allocated a time of 460 ms. Most of the other

phonemes are not known to the model and are therefore assumed to take the same time as a full node (400 ms).[1] The actual input to the model therefore has a time allocation of 1,660 ms (the next phoneme in the input, "T", would increase the time to 2,060 ms which exceeds the storage time). When the EPAM-VOC network is small, only a small amount of the input information can be represented in the store, and so new nodes will not contain much information. When the EPAM-VOC network is large, a lot of the input information can be represented in the store and so the model can create new nodes which contain large amounts of information.

Simulating the Nonword Repetition Results

There are two main sets of results for the nonword repetition test. One set is taken from studies of children of four and five years of age (Gathercole & Baddeley, 1989; these results were reported in the introduction). A second set of nonword repetition results is reported by Gathercole and Adams (1993), who used a simpler version of the test on children of two and three years of age. They found the modified test still allowed phonological memory skills to be reliably tested.

The EPAM-VOC model learns sequences of sounds from inputs that are strings of phonemes (converted from speech utterances). EPAM-VOC is therefore capable of simulating the nonword repetition data of 2-3 year olds and 4-5 year olds solely by modifying the input that is given to the model to reflect the type of input that will be received by these age groups. *No mechanism within the model was changed for any simulation.* However, simulating the nonword repetition data of 4-5 year old children is difficult because the input which they receive comes from a variety of sources (e.g. conversing, schooling). Selecting input to reflect this is difficult, and so this paper places more emphasis on the simulation of 2-3 year old children (whose input is much less variable).

The NWR test for the model was performed by presenting each nonword to the model (as a string of phonemes), and seeing if the components of the nonword could be accessed within the same time limitations that were used for the input (see earlier). By definition, the information at one node cannot represent all of a nonword (because the nonword will never have been seen as input). Information from several different nodes is required to represent the nonword. If the number of nodes and the phonemes in each node could fit into the time limit, the nonword was deemed to have been repeated correctly, otherwise the nonword was deemed to have been repeated incorrectly.
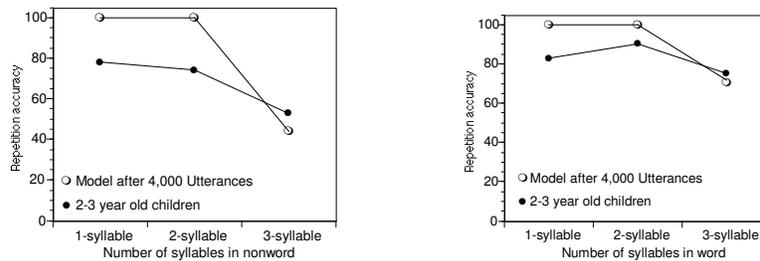
Simulation of Two and Three Year Old Children

EPAM-VOC used the most natural form of input for simulating 2-3 year old children's NWR results: utterances from five mothers interacting with their 2-3 year old children, taken from the Manchester corpus (Theakston, Lieven, Pine & Rowland, in press) of the CHILDES database (MacWhinney & Snow, 1990). The average number of utterances for each mother was 27,172 (range 23,583-32,558).

The model was run once for each of the mothers input. After every 500 utterances seen by the model, a nonword test was carried out. This consisted of presenting each nonword as input to the model and seeing if the nonword could be represented within the 2,000 ms time capacity. Note that Gathercole and Adams used a simplified version of the NWR test (using 1-3 syllable nonwords). They also performed a word repetition test. The results of the children and the model (after 4,000 utterances of input) for both the nonword and word repetition tests are shown in Figures 2 and 3.

The model performs at ceiling for the one and two-syllable nonwords and words. This is because a minimum of five phonemes can fit into the 2,000 ms time limit in the phonological loop (one phoneme uses up 400 ms at most). In the nonwords and words used, the average number of phonemes for one-syllable items is 3.2 and for two-syllable items is 5.0. The children do not perform at ceiling for any of the conditions.

---

[1] The time allocated to each phoneme is assumed to be constant.

Figures 2 and 3: Nonword (left) and word (right) repetition accuracy for 2-3 year old children and the model.

The children may not perform at ceiling for one and two-syllable items because of noise during either recognition or articulation of the item. In terms of the model, the children may not perform at ceiling because items in the phonological loop of young children decay quicker than for adults (all of the timing estimates used in the model are based on studies involving adults). The three syllable words and nonwords do not have these problems; EPAM-VOC provides a close fit to the children for both three syllable words and three syllable nonwords.

The results show that the EPAM-VOC model can produce repetition results which are comparable to young children using a simple learning mechanism and naturalistic input. In particular, repetition performance for three syllable items is closely matched by the model. The simulation also raises further questions about children's performance on repetition tests for one and two-syllable items.

Simulation of Four and Five Year Old Children

Given the problems in defining the input that 4-5 year old children receive, two different types of input were used to simulate their NWR results. The same utterances used as input to the model in the simulation of 2-3 year old children formed one set of input. The second set of input consisted of a random selection of words from the CMU Lexicon database. Figure 4 shows the comparisons of the NWR results of the simulations and the children.

The children's performance for one-syllable nonwords is actually worse than for two-syllable nonwords (which should be more difficult). The poor performance for children's repetition of one-syllable nonwords is partly due to the "intrinsic acoustic characteristics of the stimuli constructed at this length" (Gathercole & Baddeley, 1989, p.209).

The model again performs at ceiling for one and two-syllable nonwords. However, the three and four-syllable performance by the model is more interesting. The model tends to over-perform when the lexicon is used as input, yet tends to under-perform when the mothers' utterances are used as input. *This highlights the role of the input as a mediating factor in repetition performance.* Even after the model receives 25,000 of the mothers' utterances as input, its performance is still poor (this is because the vocabulary used in the input is not sufficiently diverse; only 3,046 words on average are used in the mother utterances in each of the mother-child interactions). For more diverse input (5,000 words from the lexicon, with each word being treated as a separate utterance), the model over-performs. The difference in performance across different inputs shows that the input which a child receives may strongly influence repetition performance. The results show that creating a realistic reflection of the input that 4 and 5 year olds receive is difficult, but suggest that a more realistic input would produce a good match to the data.
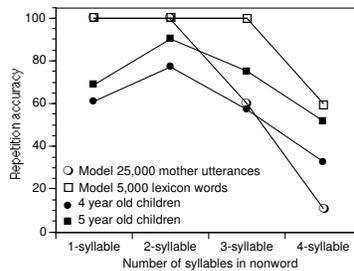
Figure 4: NWR accuracy for 4 and 5 year old children, and the model.

Discussion

The simulations have shown that EPAM-VOC is able to approximate the NWR performance of both 2-3 and 4-5 year old children using a combination of a learning mechanism, naturalistic input, and a simple implementation of the phonological loop. This represents a parsimonious approach to learning novel sound patterns.

EPAM-VOC overcomes the major failings of the phonological loop hypothesis by providing a specific account of how existing vocabulary knowledge influences the acquisition of new sound sequences, and of how new words may be learnt by combining sound sequences. EPAM-VOC is also able to make precise predictions because of its sensitivity to the input that it receives. For example, the sound patterns of the most frequent words in the input will be learnt first.

This work represents the first steps in a modelling research program which aims to examine the extent to which the linguistic input a child receives can account for the child's language development, both for vocabulary and syntax.

References

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.)*, The psychology of learning and motivation,* (Vol. 8), 47-90. New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21,* 85-123.

Brown, G. D. A., & Hulme, C. (1996). Nonword repetition, STM, and word age-of-acquisition: A computational model. In S. E. Gathercole (Ed.)*, Models of short-term memory,* 129-148. Hove, UK: Psychology Press.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31,* 429-460.

Cowan, N., & Kail, R. (1996). Covert processes and their development in short-term memory. In S. E. Gathercole (Ed.)*, Models of short-term memory,* 29-50. Hove, UK: Psychology Press.

De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Assen: Van Gorcum.

Dunn, L. M., & Dunn, L. M. (1982). *British Picture Vocabulary Scale*. Windsor: NFER-Nelson.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8,* 305-336.

Gathercole, S. E. (1997). Models of short-term memory. In M. A. Conway (Ed.)*, Cognitive models of memory,* 13-45. Hove, UK: Psychology Press.

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29,* 770-778.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439-454.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove, UK: LEA.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.)*, Models of short-term memory,* 73-100. Hove, UK: Psychology Press.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349-367.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2,* 103-127.

Gobet, F. (1993). A computer model of chess memory. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society,* 463-468. Hillsdale, NJ: LEA.

Gobet, F., & Pine, J. (1997). Modelling the acquisition of syntactic categories*. Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society,* 265-270. Hillsdale, NJ: LEA.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59,* 267-333.

Lane, P. C. R., Cheng, P. C-H. & Gobet, F. (1999). Learning perceptual schemas to avoid the utility problem. In *Proceedings of the Nineteenth SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence.* Cambridge, UK: Springer-Verlag.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language, 17,* 457-472.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-93.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.)*, The nature of vocabulary acquisition,* 19-35. Hillsdale, NJ: LEA.

Richman, H. B., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review, 102,* 305-330.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5,* 29-46.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (in press). The role of performance limitations in the acquisition of 'mixed' verb-argument structure at stage 1. In M. Perkins & S. Howard (Eds.)*, New directions in language development and disorders.* Plenum.

Zhang, G., & Simon, H. A. (1985). STM capacity for chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition, 13,* 193-201.