A Model of Spatio-Temporal Coding of Memory for Multidimensional Stimuli

Todd R. Johnson (Todd.R.Johnson@uth.tmc.edu)
Hongbin Wang (Hongbin.Wang@uth.tmc.edu)
Jiajie Zhang (Jiajie.Zhang@uth.tmc.edu)
Yue Wang (Yue.Wang@uth.tmc.edu)
University of Texas Health Science Center at Houston
School of Health Information Sciences, 7000 Fannin Suite 600
Houston, TX 77030 USA

Abstract

This paper presents a model of memory for multidimensional stimuli. The model captures the independence of features in memory, their recovery using spatial location and temporal cues, and the role of verbal recoding in building integrative feature memories. The model fits data showing that object features may be retrieved independently when given a location cue, but that correct retrieval of missing features given a feature cue depends on the correct retrieval of location. The model also suggests that positional codes implicated in many memory models may be the result of the initial positional encoding of stimuli by perception.

Introduction

Although perception appears to integrate multidimensional stimuli, mounting evidence suggests that object features, including color, form, motion, orientation, texture and location are independently processed by our visual system and can even remain independent in memory (e.g., Healthcote, Walker, & Hitch, 1994). This paper reviews the evidence for the independence and re-integration of features in memory and proposes a model of the spatio-temporal coding of memory for multidimensional stimuli. The model is implemented as a modification of the ACT-R cognitive architecture (Anderson & Lebiere, 1998) and shown to fit the results of a representative experiment.

Feature Independence in Memory

Evidence for the independent encoding of features in memory typically involves conjunction errors in recall or recognition tests (Reinitz, Lammers, & Cochran, 1992). In a recognition test, a conjunction error occurs when a subject reports previously seeing a new stimulus that consists of a conjunction of features from old stimuli. In a recall test, a conjunction error occurs when subjects recall a stimulus that erroneously conjoins features of previously seen stimuli. Conjunction errors have been demonstrated for a variety of stimuli, including faces (Reinitz et al., 1992; Treisman, Sykes, & Galade, 1977), two syllable nonsense words (Reinitz et al., 1992), colored forms (Stefurak & Boynton, 1986), and colored bars at different orientations (Isenberg, Nissen, & Marchak, 1990). Presentation times for study stimuli in these experiments range from 100 ms to several minutes, hence the results show that features are independently stored in both short- and long-term memory.

Nissen (1985) reported an experiment that suggested that visual features of objects (in this case color and shape) are stored separately, but are indexed or bound by their spatial location. Subjects were presented with four different shapes, each of a different color, and each in one of four positions, followed by either a location or color cue. When given a location or color cue, subjects were told to report the other two values indexed by that cue (color and shape, and shape and location, respectively). Subjects were tested in separate location-cue and color-cue conditions with 64 unique trials in each condition. Color, shape, location and cue were systematically randomized so as to ensure statistical independence among the stimuli and cues.

Nissen found that when the cue was a location, correct recall of color and shape were statistically independent; however, when the cue was a color, correct recall of shape depended on correct recall of location. These results suggest that object features are represented independently, with each feature associated with the object's spatial location. Thus, retrieving the shape of an object given its color as a cue requires one to first retrieve the location containing an object with that color, followed by retrieving the shape at that location.

Nissen's results showing independence in the location-cue condition were questioned by Monheit and Johnston (1994) who argued that because of the effects of guessing, very little deviation from independence was possible. By increasing the number of colors and forms (using letters instead of shapes) they reduced the effects of guessing and increased the expected deviation from independence. They also increased the number of trials to increase the chance of detecting a smaller deviation from independence. In a series of experiments that were similar to Nissen's location-cue condition, they found consistent evidence for the dependence of color and shape given a location cue. They explained their results by arguing that selective attention to an object tightly binds all features, but that in the Nissen experiment subjects have only enough time to selectively attend to a subset of the objects. Features of attended objects tend to be reported correctly, whereas features of unattended objects must be guessed. The combination of correct conjunction trials and those involving guessing produced the amount of dependence observed in their experiments.

Despite Monheit and Johnston's results, Nissen's experiment still supports a special role for location in binding object features. Monheit and Johnston's critique of

Nissen's experiment should apply equally well to the color-cue condition, meaning that it would have been just as difficult to detect dependence. However, Nissen found dependence in both the aggregate data and 8 of the 9 individual subjects. The fact that Nissen's experiment was sufficient to detect dependency in the color-cue condition suggests that location still plays an important role in binding object features. However, it is possible that selective attention may increase the association among object features, making location less important in the recovery of feature conjunctions. Indeed, Wolfe and Cave (Wolfe & Cave, 1999) suggested that pre-attentive features are loosely bound, whereas features of attended objects are more tightly bound.

Additional evidence suggests that features may also be bound by temporal cues (Treisman, 1977). In one experiment a series of several letters, a number, and several more letters were rapidly presented either at the same location or were alternated above and below the fixation point (Keele, Cohen, Ivry, Liotti, & Yee, 1988). Subjects were told to report the color of the background surrounding the digit. When the items were presented at the same location, more errors came from reporting the color of letters at the –1 and +1 temporal positions, items that appeared just before and just after the target. However, when the items were alternated among two locations, more errors came from the –2 and +2 temporal positions, items that occurred at the same spatial location as the target, but that were temporally more distant than the –1 and +1 items. Based on the results of several similar experiments, the authors argued that spatial contiguity is the dominant requirement for binding features and that temporal contiguity is of use only when features appear in the same location. However, the dominance of location may be an artifact of the task. Other researchers have argued that subjects may use multiple strategies to recover feature conjunctions, depending on the available cues at study and test (Heathcote, Walker, & Hitch, 1994).

There is also evidence that conjunction errors are affected by the distance and similarity among stimuli. Several experiments have found that subjects are more likely to erroneously conjoin features of adjacent or similar stimuli (for a review see Ashby, Prinzmetal, Ivry, & Maddox, 1996).

Other lines of research have shown that verbalization can result in an integrated stimulus memory. In a recognition task using colored animal shapes and long presentation times, Stefurak and Boynton (1986) demonstrated that subjects had memory for feature conjunctions unless they were prevented from naming study stimuli by engaging in a secondary verbal task, in which case they appeared to have absolutely no memory of feature conjunctions. In addition to suppressing verbalization, their experimental task provided neither temporal nor location cues, because the study stimuli were presented simultaneously, and the test stimulus was not presented in its study location. As a result, the suppressed verbalization condition did not provide any of the cues (verbal, temporal, or spatial) that are thought to mediate feature integration.

Because verbalization appears to result in an integrated feature memory, it appears that verbal codes act in a different manner than spatial and temporal codes. Instead of acting as a tag for separate perceptual memories, it seems likely that the perceptual features are simply recoded as verbal cues. For example, the features "red" and "triangle" may be recoded as a verbal chunk "red triangle" that may be retrieved as a whole. Likewise, a display containing multiple stimuli may be recoded as a verbal list, such as "red triangle," "blue square" where each item is given a temporal position code.

To summarize, features of multidimensional stimuli appear to be represented independently in memory, but bound by temporal and spatial cues. Integrated feature representations are possible, but only if verbalization is possible.

ACT-R 5.0

Our model of feature integration in memory is embedded in the ACT-R 5 cognitive architecture (Bothell, 2002), where it adopts ACT-R's theory of memory and cognition (as described below), but slightly modifies ACT-R's perceptual system. This section describes ACT-R 5. Modifications needed to support the model are described in the next section.

Unlike previous versions of ACT-R, ACT-R 5 (hereafter called ACT-R) consists of several interacting, asynchronous modules for perception, cognition, memory, and action. The cognitive module consists of a procedural (production rule) long-term memory and a goal buffer that holds the current goal and goal-relevant information. The declarative memory module consists of declarative memory chunks and a retrieval buffer that holds the last item retrieved. Each declarative chunk has a unique identifier, a type, and zero or more attributes and values, such as:

Obj1 isa shape-map feature triangle location loc1

where "Obj1" is the identifier of the memory chunk, "shape-map" is the chunk type, "feature" and "location" are attributes, and "triangle" and "loc1" are their respective values.

The perceptual-motor module has subsystems for vision, hearing, speech production, and motor commands. The visual module has a buffer that holds the currently attended visual location and the visual stimulus at that location. It accepts commands from cognition (via production rules) to conduct visual search and shift visual attention. The motor module accepts commands from cognition to do simple computer-based physical tasks, such as moving the mouse to a certain location, pressing a mouse button, and typing commands.

Much of the coordination between perception and action is done by production rules. The condition side of a rule is limited to testing the buffers (including whether a particular

module is busy), whereas the action side can only initiate a limited set of actions that modify buffers or send commands to one of the other modules. When a rule fires, its action side initiates commands to the other modules, such as shifting visual attention or retrieving a red object from memory, after which the rule system is free to fire additional rules. The other modules in ACT-R handle these actions asynchronously, usually resulting (after some delay) in changes to the buffers. Rules can then detect these changes and take appropriate actions. Although more than one rule can match at a given time, ACT-R only fires one rule in each cycle. A psychologically realistic conflict-resolution mechanism, based on cost and probability of success, determines which of several matching rules will fire.

To understand how this works, suppose that ACT-R is given a cued recall task, where it must report a remembered shape with a cued color. Furthermore, assume that ACT-R is attending to a fixation point that changes to the cue word "red." When the visual system detects the change, it updates the visual buffer to indicate that the word "red" is now attended. A production rule that is conditioned on seeing a word in the visual buffer fires and initiates a memory recall request for a red shape, plus notes on the goal that such a request was initiated. As the declarative memory module begins to process this request, the rule system continues to check for and fire any matching rules. This allows ACT-R to engage in additional cognitive processing, including initiating commands to the perceptual-motor system, while the memory system processes the retrieval request. When the retrieval request is complete, the retrieval buffer is filled with either the newly retrieved chunk or an indication of a retrieval failure. Two separate rules, both sensitive to the goal annotation indicating the retrieval request, handle these possibilities. One rule tests for a shape in the retrieval buffer and initiates a speech command to say the name of the shape, the other rule tests for a retrieval failure and initiates a second retrieval to guess a shape.

To understand the model presented below, it is also necessary to understand how ACT-R processes retrieval requests. Retrieval requests specify a chunk type and one or more attribute-value pairs. The memory module returns the chunk of the specified type with the highest activation value, where activation of chunk i is determined by

$$A_i = B_i + \sum_j W_j S_{ji} + \sum_k P_k M_{ki} \qquad \text{(EQ 1)}$$

B_i is the base level activation of the chunk, reflecting how recently and frequently it has been retrieved. The first summation reflects associative priming of the chunk by chunks in the goal buffer, where W_j is the available activation and S_{ji} is the strength of association from chunk j to chunk i. W_j is typically set to 1/n, where n is the total number of chunks in the goal buffer. S_{ji} is initially set to S-ln(n), where S is a constant and n is the number of chunks that have chunk j as an attribute value. This setting produces the classic fan effect (Anderson, 1974).

The second summation in EQ 1 reflects similarity of the chunk i to the retrieval cue. M_{ki} is the similarity between the value of the *kth* attribute in the retrieval cue and the value in the corresponding attribute of chunk i. P_k (which defaults to 1) reflects the weighting given to the similarity of attribute k. By default, M_{ki} is 1 if the kth attribute value in the cue is identical to the corresponding value in chunk i, otherwise it is –10.

To model the random fluctuations of human memory, activations vary with time by adding noise as a logistic function of the parameter s, where s is related to the variance of the noise by

$$\sigma^2 = \frac{\pi^2}{3} s \qquad \text{(EQ 2)}$$

Finally, the activation threshold τ specifies the minimum activation value for retrieving a chunk. If all chunks matching the cue fall below this value, the retrieval request fails. As with chunks, the retrieval threshold varies from time to time according to the noise parameter s.

The approximate probability of retrieving a chunk i given k competitors (including the threshold and chunk i) is given by

$$P(i) = \frac{e^{A_i / s\sqrt{2}}}{\sum_k e^{A_k / s\sqrt{2}}} \qquad \text{(EQ 3)}$$

where A_n is the mean activation of chunk n.

A Model of Memory for Multidimensional Stimuli

The model assumes that attending to a multidimensional stimulus results in a set of feature chunks in memory, where each chunk encodes one feature along with one or more temporal and spatial tags. While attention is fixed on the stimulus these chunks also appear in the corresponding perceptual buffer. If a stimulus is recognized (either identified or classified or both), perception may also deliver a separate chunk encoding the identity (or class) of the stimulus along with spatial and temporal tags.

Suppose that ACT-R attends to a red square at location Loc22 on a computer screen at time t1. ACT-R's visual buffer is then filled with chunks encoding red at Loc22 t1, square(shape) at Loc22 t1, and square(class) at Loc22 t1. These same chunks are also added to ACT-R's declarative memory. This is shown graphically in Figure 1, where squares represent chunks and arrows indicate chunk attributes. Locations (e.g., loc22) are chunks that correspond to unique locations using the computer-screen as the frame of reference.

A spatial tag encodes where the feature occurs and may be given in any number of frames of reference (e.g., Wang, Johnson, Zhang, 2001). For instance, one spatial tag might give object heading and distance in egocentric (body-centered) coordinates, whereas another spatial tag might indicate the exocentric heading and bearing of the object from another object. Because ACT-R's perceptual-motor

system is designed to work with two-dimensional computer displays, the model provides a spatial tag relative to the frame of the display. Evidence for frame-relative location encoding has been found in both monkeys and humans. Rolls (1999) found that some neurons in the monkey hippocampus responded to where the monkey looked on a screen independent of the position of the monkey relative to the location of the screen. Hock, et al. (1989) showed that subjects unintentionally retained frame-relative locations of circles forming patterns in a frame, such that they could estimate the frequency with which circles appeared at a particular location within the frame.

Given enough time, rules may recode the features. For instance, a set of rules may verbalize the visual features "red" and "triangle" resulting in a redundant verbal code with appropriate temporal tags. Rules may also recode the features into an integrated representation, such as a single chunk that binds "red" and "triangle." Whether or not a stimulus is recoded, and the nature of the recoding, is dependent on the production rules, which in turn depend on the current goal and the strategy being used to achieve it.

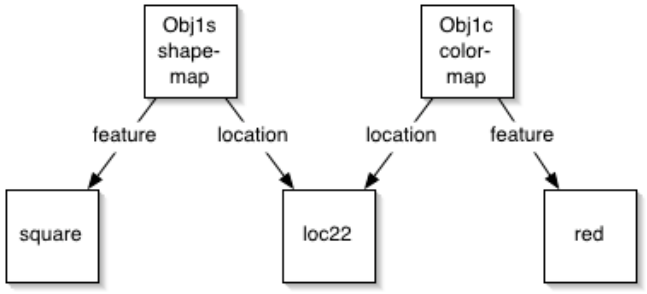

Figure 1. Representation of color, shape and location in the ACT-R model. Temporal cues are not shown.

The model assumes that the similarity (M in EQ 1) of temporal and spatial tags is inversely proportional to their temporal and spatial distance; however, the exact nature of this relationship is left to the model builder. As a result, the model will tend to confuse spatially adjacent features.

Applying the Model to the Nissen Task

As a partial test of the model, we applied it to the Nissen (1985) experiment described earlier. The critical phenomena in this experiment is that recall of color and shape is independent given a location cue, but when given a color cue, recall of shape is dependent on correct recall of location.

The ACT-R model contains production rules for attending to the four colored objects, attending to the cue, retrieving the answers, and pressing keys to record its responses. When presented with the 4 objects, the model visually attends to each object, resulting in automatic encoding of a color-map and shape-map chunk for each object. It then waits for the cue to appear, at which point it attends to the cue and begins the retrieval and response process.

The critical production rules for modeling the experimental results are those for retrieving location given a color, and those for retrieving color and shape given a location. When given a location cue the model first attempts to retrieve a chunk encoding the color at that location (such as obj1c in Figure 1), and then attempts to retrieve a chunk encoding the shape at the given location (such as obj1s). When given a color cue, the model attempts to retrieve a color-map chunk containing that color (e.g., obj1c). It then uses the location in this chunk to retrieve a chunk encoding the shape at that location.

If a rule fails to retrieve a chunk, the model will guess an appropriate value. For example, if the model fails in retrieving a color-map chunk with color red, it will simply guess a location, and then use that location when it attempts to retrieve the shape.

Fitting the Nissen data requires estimating the parameters in EQ 1, as well as the activation threshold, and the noise parameter. The activation threshold and the base level activation B_i for all stimulus chunks were set at 0—the default value. The amount of activation available for associative priming was also set at 0, because chunks in the goal buffer are redundant with attribute values in the retrieval cue. Similarity among matching attribute values, including locations, was set to 1 with mismatching values set to 0. The noise parameter s was the only parameter tuned to fit the data. We used EQ 3 and the results from the Nissen experiment to determine an initial value of s, then iteratively refined it over several model runs to produce the fit reported below (where s = 0.39).

A. Location-Cue Condition

		Shape: Correct	Shape: Incorrect	
Color: Correct		0.485 (0.450)	0.212 (0.219)	0.697 (0.669)
Color: Incorrect		0.213 (0.175)	0.090 (0.156)	0.303 (0.331)
		0.698 (0.625)	0.302 (0.375)	

B. Color-Cue Condition

		Shape: Correct	Shape: Incorrect	
Location: Correct		0.477 (0.494)	0.221 (0.234)	0.698 (0.728)
Location: Incorrect		0.032 (0.051)	0.271 (0.221)	0.303 (0.272)
		0.509 (0.545)	0.492 (0.455)	

Figure 2: Results of simulating 50 subjects for each condition. Values in parentheses are the experimental results from Nissen.

The results of running the model for 50 subjects in each of the two conditions are shown in Figure 2 along with the Nissen data. As expected, shape and color recall are statistically independent in the location-cue condition (χ^2 = 0.131, p = 0.72), whereas location and shape are dependent in the color-cue condition (χ^2 = 901.23, p < 0.01). The proportions of correct and incorrect recall across both conditions produce a good fit to the Nissen data: R^2= 0.95.

Conclusion

The model described in this paper can account for the basic phenomena of memory-based feature integration. The special role of location was demonstrated by applying the model to the Nissen task. The use of temporal cues, such as that reported by Keele, et al. and discussed earlier, is supported by the model's use of temporal tags for each feature. If location is given a stronger association to the features than is the temporal tag, this would produce Keele's results showing a role of temporal contiguity only for items that appear in the same spatial location. The tendency to erroneously conjoin spatially adjacent features and features of similar stimuli is captured in the model using ACT-R's theory of memory retrieval which tends to confuse similar stimuli (see the discussion of EQ 1). This means that features of spatially proximal objects will be confused more often than those of spatially distant objects. It also means that if the model is trying to recall the color of an oval, it would be more likely to confuse its color with that of a circle than with a square.

The effects of recoding, including verbalization, are captured in the model by assuming that the names of individual features or the name or semantic identity of an object may be memorized instead of the individual visual features. If there is enough time for object identification, the subject need only remember an object's identity and location during study. At test, the object's identification provides a cue for reporting essential object features. Such recoding will produce integrated memories, because the individual features are already well-learned and "bound" to the object identity. If a subject remembers seeing a banana, conceptual knowledge of bananas is sufficient to recall that it was yellow and crescent shaped—there is no need to encode the specific perceptual features in a new memory trace.

Such a strategy will not work, however, if the task presents bananas in unnatural colors. In this case, verbal rehearsal and the resulting verbal memory may be of use. For instance, given enough time a subject might remember objects in the Nissen task by verbally rehearsing "red triangle, blue square…" and so on. Recall of these items would then be subject to serial recall effects, such as the serial position curve and positional errors. It seems likely that such a strategy would result in fewer conjunction errors, which would explain why verbalization results in integrated feature memory.

The model provides a possible explanation for the need to use positional codes (instead of integrated codes or associative chaining) in cognitive models of memory tasks. Positional codes were used to account for chunk position effects in alphabetic retrieval response times (Klahr, Chase, & Lovelace, 1983) and positional errors in serial recall (Anderson, Bothell, Lebiere, & Matessa, 1998). It is possible that these codes may be the direct result of the positional (temporal or spatial) encoding of stimuli by perceptual processes.

One limitation of the model is that it treats all errors as memory retrieval errors. However, the model could be extended to include probabilities for correctly perceiving features, or a theory of feature perception. However, since our emphasis is on the representation of features in memory and their later reintegration, we saw no need to introduce additional theory.

Monheit and Johnston's demonstration of dependence of color and shape given a location cue provides a challenge to the model presented here. To account for the dependence our model must be modified to provide for some strength of association between features of attended objects. In the present model, knowing the color of an object does not activate the object's shape (e.g., the strength of association between color and shape, S_{ji} in EQ 1, is 0). In the revised model, the color of a previously attended object would activate its shape, allowing for some dependence among features of objects. This would make our model consistent with Wolfe and Cave's view that preattentive features are loosely bound, whereas features of objects that have been attended are more tightly bound.

Finally, the model is meant to provide a foundation for a comprehensive theory of spatial cognition embedded in ACT-R. By embedding the model in ACT-R other researchers can use it in their models, where it may provide additional constraints and enable more realistic memory representations and behavioral predictions.

Acknowledgments

This research was supported by the Office of Naval Research, Cognitive Science Program under Grant No. N00014-01-1-0074.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*, 451-474.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341-380.

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Hillsdale, NJ: Lawrence Erlbaum.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103*(1), 165-192.

Bothell, D. (2002). *Act-R 5.0 Beta*. Retrieved February 6, 2002, from http://act.psy.cmu.edu/ACT-R_5.0/

Heathcote, D., Walker, P., & Hitch, G. J. (1994). Feature independence and the recovery of feature conjunctions. *The Journal of General Psychology, 121*(3), 253-266.

Hock, H. S., Smith, L. B., Escoffery, L., Bates, A., & Field, L. (1989). Evidence for the abstractive encoding of superficial position information in visual patterns. *Memory & Cognition, 17*(4), 490-502.

Isenberg, L., Nissen, M. J., & Marchak, L. C. (1990). Attentional processing and the independence of color and orientation. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 869-878.

Keele, S. W., Cohen, A., Ivry, R., Liotti, M., & Yee, P. (1988). Tests of a theory of attentional binding. *Journal of Experimental Psychology-Human Perception and Performance, 14*(3), 444-452.

Klahr, D., Chase, W. G., & Lovelace, E. A. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Pschology: Learning, Memory, and Cognition, 9*(3), 462-477.

Monheit, M. A., & Johnston, J. C. (1994). Spatial attention to arrays of multidimensional objects. *J Exp Psychol Hum Percept Perform, 20*(4), 691-708.

Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 205-219). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition, 20*(1), 1-11.

Rolls, E. T. (1999). The representation of space in the primate hippocampus, and its role in memory. In N. Burgess, K. J. Jeffery & J. O'Keefe (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 320-344). Oxford: Oxford.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. *Perception & Psychophysics, 39*(3), 164-174.

Treisman, A. (1977). Focussed attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics, 22*, 1-11.

Treisman, A., Sykes, M., & Galade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Atention and performance VI* (pp. 333-361). Hillsdale, NJ: Erlbaum.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24*(1), 11-17, 111-125.

Wang, H., Johnson, T. R., Zhang, J. (2001). The mind's views of space. In *proceedings of the 4th International Conference of Cognitive Science.*