

Activity Corner

Performing temporal tasks requires that humans remember information long enough to act upon it. This exercise exposes one of the fundamental usability challenges of a temporal user interface.

Materials Required

- Writing pad;
- Pen or pencil;
- Some friends or colleagues; and.
- A meeting place.

Step 1: Make a Series of Lists

Create several lists of various types as described below. Use the example list on the next page as a guide. Each list should be anywhere from 15 to 25 elements long. The following five should be enough, but make up more if you find you need them.

- Restaurants, as shown on the next page;
- Celebrity names;
- Names of animals;
- Colors; and,
- Nations of the world.

Step 2: Assemble As a Group

Gather your friends and colleagues together in a conference room or other meeting place. An appropriate group might be your IVR team, making this a good warm-up exercise before a brainstorming session.

Step 3: Read the Lists Aloud

- Don't share the lists in advance with the group.
- Appoint a reader.
- Have the reader read the list aloud.
- Pace the reading at about two seconds per list element or slightly slower. This means that you will make it through the entire list in under one minute.

Step 4: Write Down Everything You Remember

- Do not allow anyone in the group to write during the reading! You must wait until the list is complete before you pick up your pencil.
- Do not allow anyone in the group to speak—either during or after the reading.

- Instruct the group to listen closely to the list and to concentrate on remembering as much of it as possible.
- As soon as the list is complete, pick up your pencils and write down as many list elements as you can remember.
- Take a minute or more—sometimes you can remember something after concentrating that you missed on first listening.
- Pencils down when finished.

Following is an example list. I've deliberately made the list long but not too long—in this case, a list of 24 restaurants.¹

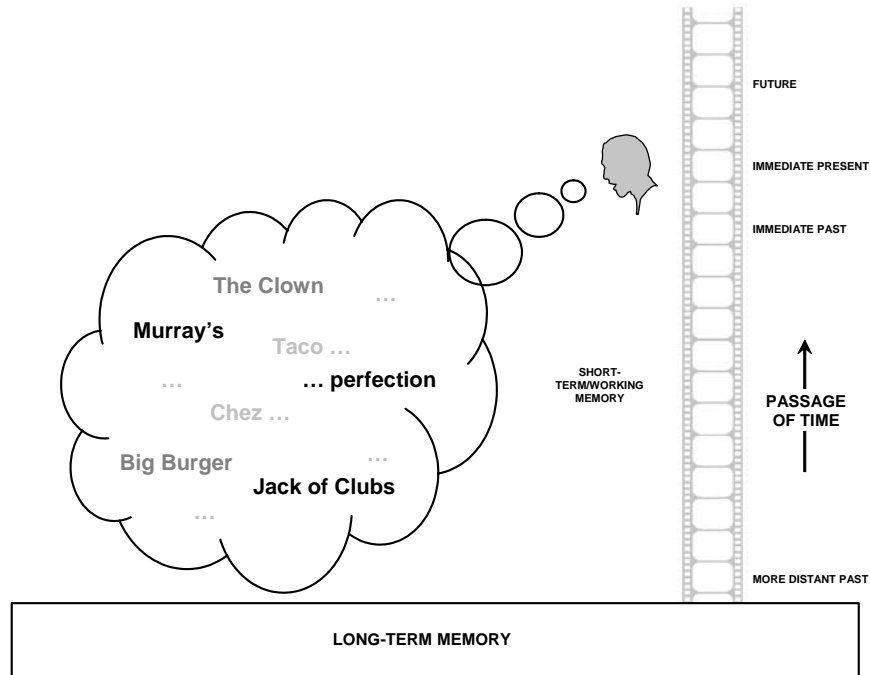
Murray's	The Fry Pan	Jack of Club Sandwiches	Taco Guano
JumpStart Cafe	Sea Sharp	The Radish	Star Ship
Yummies Greek Treat	Mustard Seed	On the Rocks	The Waffle Spot
The Clown	OneGreatBurger	The Boardroom	Home Grown Heaven
Chez Jacques	The Pumpkin Patch	Benny's	The Fox and Hound
Dottie's	Moon Dance	Li'l Dumplin'	Aged to Perfection

Step 5: Compare Notes

Reflect on your own performance and discuss it with the group. *Introspect* on what it felt like when you were hearing the list—let's say the 24 restaurant names above. The reader is like an IVR system presenting information to a user via telephone. When the recitation by the system is over—that is, the *presentation services* component of the user interface has presented its data to you—your internal mental context may look something like the following illustration.

I've represented sensory input as a kind of filmstrip—a set of snapshots that plays across your attention and gives you an instantaneous picture of the world around you. In the immediate present—near the top—you have your current conscious view of that world. It includes information that has come in to your eyes, your ears, and your other

senses including internal sensations from your body. Time is flowing upward from the past into the future.



This filmstrip metaphor is not technically accurate in terms of psychology. But I find it convenient as a visual reference, and I don't think it is too misleading. In fact, I remember many years ago reading about a cognitive psychology concept known as Rapid Attentional Integration (RAI). I have borrowed this filmstrip metaphor from that reading. The idea is to imagine that all kinds of preconscious processes have already formed percepts for each sense, and that some higher-order process has now integrated all of those percepts into a pre-organized and partially interpreted snapshot of reality. It is this snapshot that is presented to your conscious mind, at a rate of some ten frames per second or so.²

At this moment—when you first become conscious of the sensory input and can attend to its many components—the information reaches the highest level of sensory memory and enters short-term memory—allowing you to think about it and to make decisions based on it.

As shown in the illustration, a lot of time has passed since the filmstrip appeared with the list elements—they are all fading from memory as you reflect on them. You are now in a present in which your attention is focused inward—on trying to remember the vanishing list of restaurant names. That is why it is important to turn down the gain on your incoming movie by remaining quiet. Otherwise, new information will interfere with your ability to remember the immediate past.³

Notice that the names—shown in the cloud that you are seeing in your “mind’s eye”—are already somewhat jumbled. The first thing that fades is the exact ordering of the names. You may remember one or two, but not be able to recollect which came earlier and which later. You are also more certain about specific ones—shown in a black font—that seem to be more vivid for no apparent reason. Others are dim and foggy. Maybe you only remember parts of some of the names, for example, “Chez something-or-other.”

As shown in the illustration, you can only hold a few items in memory at a time. So when you get to the end of the list, there are only a few “ghosts” remaining. You’ll find that’s true for your colleagues as well. Some people may remember as many as nine restaurants or so—although probably the number is much smaller. I wouldn’t be surprised to see some sheets with only two or three names on them. Many of your group may well remember the last restaurant, because it was the most recent to reside in sensory memory.

Perhaps—after some thinking, you might come up with some extras:

(illustration, a handwritten list as one may have written them down for the exercise—a couple of common franchises, plus a “wrong” memory that contains a partial restaurant name, and another “gummy something” instead of Yummies Greek Treat.) The names are: Benny’s, OneGreatBurger, Taco Guano, and Jack of Club Sandwiches.

Look at these last ones. Now look inside your own mind. You didn’t really *remember* them at all, did you? You *reconstructed* them from your world knowledge. Maybe the franchises—Benny’s, OneGreatBurger, Taco Guano—were a little easier because they’re commonly known. Does that make them easier to remember? No, probably not. Instead, what you have in your vanishing memory of the auditory list is a remnant

that you have reconstructed from long-term memory. In other words, your knowledge of fast-food chains, plus a little reference to an echoic memory that is rapidly degrading, allows you to *hypothesize* that you “heard” these choices—even though you don’t really *remember* hearing them per se.

You can test this effect yourself. Build two lists. One, a random collection of unrelated words with no context. The other, a list of related choices including some commonly-known items—perhaps restaurants as shown here—or colors, celebrity names, nations of the world, or a list of animals. Make the list too long to remember on a single presentation (15 or 20 will do fine). Then read the list aloud once to a friend. Have her write down everything she remembers—not during the presentation, but immediately afterwards. Remain silent while she writes.

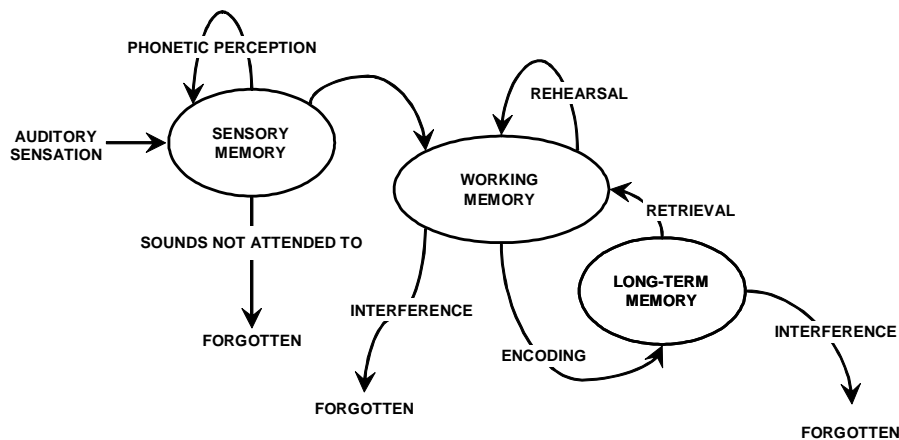
These kinds of tests are very easy to run, remarkably effective at teaching you the basic constraints of human memory, and surprisingly consistent. You will find natural variation from person to person, but on the whole the number of items remembered correctly will cluster into the shape of a bell curve as you sample more people. The longer the list, the more rapidly the curve will slope off. The tails of the curve will be fairly predictable—*everyone* will remember at least one item, and *no one* will remember all items.

If the list is long enough and contains commonly-known elements (as in the restaurant example), you might see *false positives*. That is, some people will “remember” restaurants that were not actually spoken.⁴ Such occurrences will increase with the length of the list and with the commonness of the items. If the list is arbitrary, false positives will go down but the total number of remembrances will also drop.

One final variation on this theme: try assembling two lists. Read the first one aloud. Then hand the second printed list to the friend. Have the friend check the words on the second list that were in the verbal presentation. Have the friend X the words that were not in the verbal presentation. On any for which she’s not sure, have her make no mark. You can repeat this across many friends and plot both the negatives and positives on a graph.

Human Memory

Designing a TUI requires an understanding of human memory. Most overviews of human memory represent brain structures using a kind of flowchart or state machine metaphor—like the illustration below. This can be a handy representation, and I share it here as a starting point for our journey through the world of human memory. Go to your favorite search engine and type in any one of the words you see in the illustration, and you will be literally inundated with similar figures.



I have to alert you that I don't care for these kinds of simplifications. They encourage the reader to form an image of the human brain as essentially a computer—with “buffers” and “temporary storage,” and a separation between data and actions. You tend to want to imagine that one “layer of processing” passes data upward to a next layer, eventually computing a model of reality based on sensory interpretation. Such models tend to mislead.

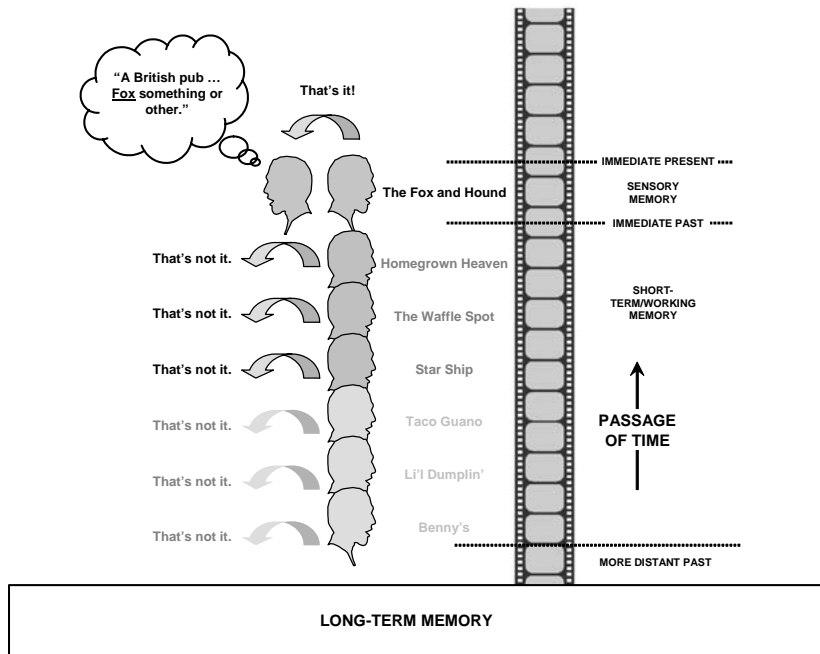
In its favor, however, the illustration presents a way of thinking about memory that can be convenient. And it does so by providing a visual metaphor that allows us to appreciate the structural and functional differences that have turned up in the huge volume of psychological data over the years. So I present it with the caveat that we're talking about human brains here, not a Von Neumann computer.

Memory is divided into three functional units—sensory memory, working memory, and long-term memory. In the illustration, we are dealing only with the auditory modality as it applies to spoken language.

I draw the illustration as a state machine for clarity.

“That’s It!”

Now here is the same list of restaurants that we used before to introduce the topic of human memory. This time, we’ll allow the user to interrupt. Read from bottom up to view the passage of time. The leftward-facing “you” is your mind’s eye, holding the target in mind. Your attention must return frequently to that target as you listen to the list, a process called “rehearsal.” Notice that the target is not precise—you’re looking for the “Fox something or other” and you’ll know it when you hear it. The curved arrow represents a comparison between the incoming speech and the target.



Back in the past, you heard, “Benny’s.” You then looked internally at your target and discarded the option. You are now free to forget “Benny’s”—you have no use for it

now that it's discarded. Then the same thing occurs as you listen to the rest of the list. "Li'l Dumplin'" [that's not it] ... "Taco Guano" [that's not it] ... "Star Ship" [that's not it] ... "The Waffle Spot" [that's not it] ... "Homegrown Heaven" [that's not it] ... after each choice you reflect on your target and then discard the option.

Finally you hear, "The Fox and Hound." That input—residing now in sensory memory—matches your target. The match triggers a "that's it" response.

As you can see, the user can easily choose if the task is one of *recognizing* a known target with "that's it" rather than one of making a more complex decision based on presented options. The distinction is important. If the user is recognizing, then all she must do is compare the item resident in echoic memory with her mental image (already constructed) of her desired goal. If the input resonates with the goal—even if it is not exactly the same word or phrase—then the emerging construct "lights up" inside the user's mind with a "that's it" response.⁵ "That's it" is the quintessence of recognition, and takes the form of an image that connects quickly with decisions, actions, changes to the organism, and the follow-on feelings that result from them.⁶

If the input does not resonate with the goal—that is, no recognition occurs in the user's mind—then it can be forgotten without penalty. The user stops attending to the current list item, and awaits the next input—all the while rehearsing the target to ensure that it remains in short-term working memory.

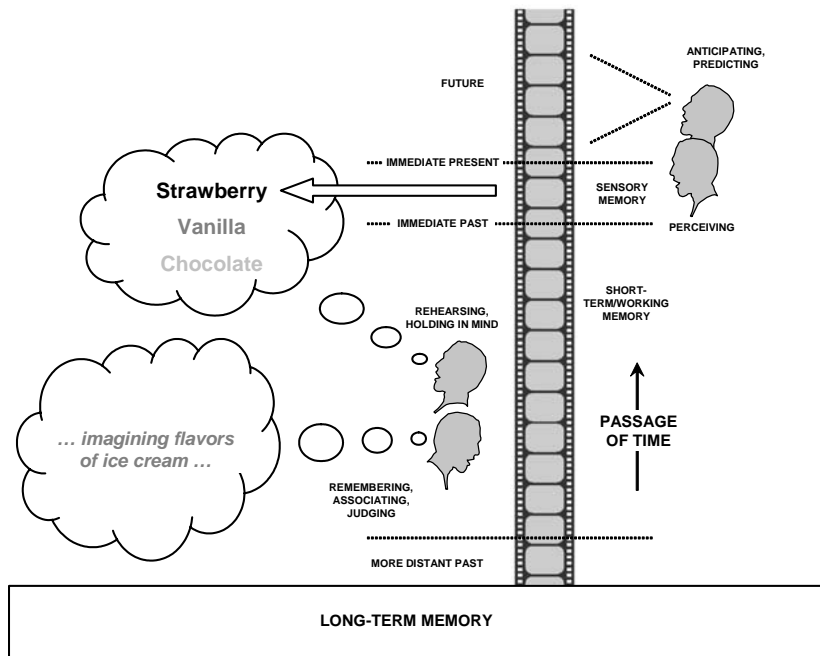
What this means practically is that lists known to contain a single item that the user can be expected to recognize and act upon can be quite long—perhaps 20 items. The user is not expected to hold any list elements reliably in mind, and the user cannot be expected to attend across more than two or three list elements at the very most. But recognizing is easy, so such lists can be longer.

Note that certain users do not interrupt, and so might fail when selecting from long lists—even if they are only attempting recognition.⁷

Holding a List in Mind

Extracting a "that's it" response from a user listening to a list of options is, although an easy task, not a very common one for IVR applications. More often, the user

is expected to make a *decision*—not just select from a list. And decision-making uses very different psychological processes. The cognitive load for such a task is illustrated in the following figure:



Imagine that you are asked to decide which of several items is the one you prefer at a given moment. For example, you want an ice cream cone, and you are presented with the question, “chocolate, vanilla, or strawberry?” It’s a 3-way list, right? But it is not as though you have a pre-conceived template that will light up with a “that’s it” when you hear one of the flavors. Rather, you must integrate the options and hold all of them in mind while “comparing” their properties against your internal desire.

Look at the illustration to see how much more complex that task is.

Now there are “four of you,” and attention must be distributed among all four. Part of you, upper right, is anticipating what might come next. That’s the “you” that the discourse marker “or” in the question, “chocolate, vanilla, or strawberry?” is designed to serve. Another part of you in the here and now is attending to sensory memory—perceiving and interpreting the current option. A third part of you, upper left, is rehearsing the list in order to hold it all in short term memory at once. And finally, lower

left, a part of you is retrieving experience memories to make a decision about your desire—in effect a “matching target.”

This fourth “you” is especially interesting. As you hear the choices, you take a moment to “imagine” (remember) what the first flavor, chocolate, tastes and feels like. Then you create the same fantasy about vanilla, and finally strawberry. You are really constructing a “taste test” for each flavor—from memory of course—and then comparing each test to make a final decision about *which flavor you want right now*.

Antonio Damasio tells us:

“... the images over which we reason (images of specific objects, actions, and relational schemas; of words which help translate the latter into language form) not only must be “in focus”—something achieved by attention—but also must be “held active in mind”—something achieved by high-order working memory.”⁸

This is the quintessential difference between recognizing and selecting a target (“that’s it”) on the one hand, versus *making a decision* (“this is what I want”) as discussed here. If you must decide based on several options, then you must accomplish the following:

- Listen to the whole list and retain it;
- Rehearse each element as the list progresses to ensure nothing is forgotten;
- Compare each element in the list with your need (perhaps multiple times);
- Make a decision that discriminates one list element from the others; and,
- Speak the chosen item (or press its corresponding key).

All of these actions must occur during and immediately after the list presentation. This is a case in which barge-in makes a smaller contribution. You need not interrupt the list, because you have to hold the entire list in mind before you can compare the items and make a choice.⁹

Users confronted with this challenge take longer to decide, make more mistakes, and more often must hear the list again. The problem grows with the length of the list. The design principle that emerges from this observation is as follows:

The maximum length of a list is shorter for lists that require “deciding” as a user action than it is for lists that only require “recognizing and selecting.” Rule of thumb: three or (at most) four.¹⁰ What’s more, timeout values need to be longer at the end of decision-list presentation than for the corresponding “that’s it” tasks.

Mapping a Concept onto a Verbal List

Tasks confronting users are sometimes more complex than can be represented by simple words or phrases. In such a case, the user has an even more difficult challenge trying to make what appear to be simple selections.

“Do you want claim status, coverage and benefits, or billing information?”

Here we have a set of formal categories—by and large not words that we use every day. This is because the task itself is not one that we pursue every day. The health-care application that asks such a question is an abstraction that derives from the complexity of our economy and the modern society that it has spawned. There is no “natural” or “evolutionary” model that makes this interaction even remotely basic or simple. So callers have to perform mental calisthenics to get through the question.

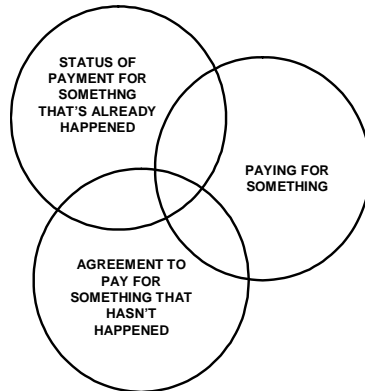
In this case, you are listening to words from the IVR and concurrently thinking about and making internal comparisons with your need or goal. This activity is similar to the previous essay, in which you are “deciding” which flavor of ice cream you want. But in this case, the task is even more difficult, because you must devote some additional brain power to figuring out the precise meaning of the words that stand for the tasks.

What is “Claim status?” What does it mean? Well, in isolation, with plenty of time to think about it, one could arrive at a definition—a “claim” must be a request for the insurance company to pay. And the following phrase, “coverage and benefits,” must be information about what the insurance company is willing to pay for. So you ask for claim status if you have already made a claim for a medical procedure in the past, and want to find out its status. But you ask for coverage and benefits if the medical procedure is in the future—you are in effect asking for permission to seek the procedure under coverage.

In other words, seeking payment in the past versus payment in the future is what you’re asked to decide. Now hold those concepts in mind as you hear the third choice—billing information. Well that’s a kind of payment as well, isn’t it?

What we have here is an exclusivity problem. As shown in the Venn diagram, the three classes have overlapping semantic implications—even though the *words are all different*. Each choice in the list is not fully exclusive of other choices, so you must hold all of these elements in mind at once—a challenge for short-term memory—and you must *think*—a challenge for both short-term memory and cognition. These chal-

lenges cause *interference* that limits the number of choices that can be processed at one time.



Now imagine that you're calling for your Dad, who goes in tomorrow for his fifth outpatient visit to adjust a leg brace following his knee and hip surgery a few weeks ago. You seem to remember that those co-pays that you make every time you take him have some kind of cap. That after a certain level, you have to pay less—or maybe it was more, it's so hard to remember these things. So you listen and must decide:

"Do you want claim status, coverage and benefits, or billing information?"

Although at first glance it is a "simple" question, you've actually got some real thinking to do here, don't you?

The interference is a version of *cognitive dissonance*, and is a key explanation for why healthcare applications have such low self-service numbers. The tasks themselves are complicated, and they overlap with each other across a number of dimensions.

Just In Time Learning

User interfaces are teachers as well as servants. I once worked in an office doing IT support. It was just after the release of the IBM Personal Computer, and a pool of typists was using EasyWriter II to prepare transcriptions of workshop presentations. The year was 1982. There were no hard disks, but these transcribers were thoroughly equipped with all of the contemporary tools needed for their work.

Not one but TWO diskette drives allowed both operating system and word processor to remain in drive A while the documents could be stored in drive B—a huge improvement over the less expensive systems that required constant swapping. Users simply inserted the master disk—amusingly labeled “DOS Boot”—into the left-hand drive and turned on the computer. In addition, all machines were stocked with the full complement of RAM—64K bytes—that DOS 1.1 was capable of supporting. Finally, all transcribers had a dictation device with a foot pedal, allowing them to play, transcribe, rewind, and replay the recorded audio.

In other words, it was a top-of-the-line operation. No expense was spared.

The typists were still drawing on IBM Selectric skills, and word processing was a bit new to them. But they were learning quickly, and it was one of the most productive environments that I’ve ever worked in. Each transcriber was able to generate the equivalent of a full-length book every few days. No local area networks, and so no shared peripherals, they each had a printer, so intermediate drafts could come easily.

One typist—I’ll call her Laura—was working on a document that had a lot of repetition in it. Similar tables and text kept reappearing, with only minor changes that made progressively detailed points. I watched her with interest as she worked.

At first, Laura would scroll backwards to a previous page, observe the text, and then return to the current working page to enter as much data as she could remember. But then it occurred to her that she could *print* the previous page. Remember that she’s a touch typist. Retyping something from a printed page is an easy and frequent activity. So she would print the page and then put it on her viewing stand to retype. No more scrolling back and forth. She was gratified that this solution was much more productive, and continued in this way for some time.

After two or three repetitions of this behavior, I intervened. “Laura, I see that you’re entering repetitive information. Can I show you a trick?” I taught her how to highlight a block of text, select it, and then copy it to a new place in the document. This “block copy and paste” feature was revolutionary, and Laura eagerly mastered the technique. Later she taught all of her officemates how to do it as well. The newly-copied text could then be individually edited, and Laura found the feature to be much more productive than the printing and retyping technique that she had invented as a solution to her problem.

Laura learned the technique on first trial because it appeared as a solution at just the right time. Had she attended an EasyWriter II class, the feature of “block copy and

paste” would have faded into the same memory dump as the indexing and header/footer features. She would have forgotten them as “not relevant to my immediate needs.” But as she was typing and retyping the same text, Laura’s set and setting were such that the problem became immediately pressing and urgent. She was therefore perfectly primed for the block copy and paste solution. It resonated with her needs, and she learned—in fact mastered—the technique after having been told about it only once.

One-trial learning is a feature of “just in time” instruction. By watching her work and even letting her work somewhat unproductively for a time, I was able to hand Laura a piece of power and control that she immediately recognized and valued.

The learning was easy because the need was great and the moment was right.

There are similar techniques in IVR design for withholding certain information until the user is clearly in need of it. Once the need is great, then delivery of the information leads to one-trial learning. The transmission of power to the user is vivid and memorable—callers not only use the tools, they *remember* from session to session when and how to use them. When you deliver just-in-time tools, you’ve made a friend for life. In IVR, the tools include:

- Notification of global navigation words: cancel, main menu, operator, start over, “I’m done” or goodbye;
- Syntactical templates for multi-token (slot-filling) natural language;
- Stable turn-taking behaviors for feedback and re-prompting;
- Proper and frequent grounding—instant and self-explanatory;
- Alerts prior to difficult regions of the dialogue;
- Easy outs from difficult regions of the dialogue;
- Respect for the user’s time in the form of express lanes and problem lines; and,
- Clarity of current status.

Endnotes

¹ These names are all fictional and were invented just for this essay. Any similarity to existing restaurants is purely coincidental.

² I doubt that consciousness is quantized in this simplistic a way, but the filmstrip metaphor is good enough for the discussion, and it gives you a way to visualize a temporal process in somewhat familiar terms.

³ The fact that ongoing speech interferes with working memory is well-known, and represents yet another explanation for why speech is not a very good user interface modality. Shneiderman reports experimental results from as early as 1993 that demonstrate how speech consumes cognitive resources that could otherwise apply to problem solving:

“The part of the human brain that transiently holds chunks of information and solves problems also supports speaking and listening. Therefore, working on tough problems is best done in quiet environments—without speaking or listening to someone. However, because physical activity is handled in another part of the brain, problem solving is compatible with routine physical activities like walking and driving. In short, humans speak and walk easily but find it more difficult to speak and think at the same time.”

Shneiderman, B., “The limits of speech recognition,” *Communications of the ACM*, 43(9), 2000.

⁴ A colleague of mine in a psychology 101 class describes such an experiment. “The professor read us a list of words that all had to do with sleep—nap, bed, snore, and so on. But “sleep” was not one of the words. Most of the class, however, included “sleep” as one of the words that they remembered hearing.”

⁵ I like to visualize a “that’s it” response as an emergent property that synthesizes itself through interactions between auditory input (in echoic memory), nascent semantic implications (in working memory), and goal (in conscious short-term memory). The synthesis generates a unique image—the “that’s it” itself—that exerts a pull on attention and sets the stage for willful action.

⁶ Be careful to distinguish “that’s it” from the “Aha!” response. “Aha!” describes *epiphany*, not recognition. Although the internal feeling can be similar, an epiphanous moment results from a very different internal reorganization of information based on newly-discovered relationships. At its core, “Aha!” derives from information reduction, while “that’s it” derives from a simpler form of pattern-matching.

⁷ See IBTBAGM to learn about various call types. The Director’s Cut of that book includes an essay about Lisa, the *prudent* user, who always has to consider every option before making a selection—even when the selection is obvious. These call types turn up in IVR data as timeouts or zero-outs. Although they heard and recognized their target, the subsequent elements in the list interfered with retention. They experienced “that’s it” and then promptly forgot it before the list was over. Distinguishing a forgotten “that’s it” from a user in the wrong state on timeout is one of the tricky design problems that can lead to error amplification.

⁸ Reference Damasio (Descartes or Spinoza), page 84

⁹ This principle, of course, applies only to first-time or infrequent users. You must still allow barge-in for such a menu to support repeat users who have memorized the entire list—few though they are. And that fact really ekes a penalty on turn taking, as we’ve seen in turn-taking studies and discussed in other essays in this series.

¹⁰ See also Damasio, Descartes. pp 196-198 for more on attention and memory.