ERGONOMICS IN DESIGN

The Magazine
of Human
Factors
Applications

Your Call Is Important to Us . . . Please Hold

BY CHRIS BOND & MARK CAMACK

Reprinted from *Ergonomics in Design,* Volume 7, Number 4, October 1999; pages 9-15. Copyright 1999 by the Human Factors and Ergonomics Society, P.O. Box 1369, Santa Monica, CA 90406-1369; 310/394-1811, FAX 310/394-2410, hfes@compuserve.com, http://hfes.org. All rights reserved.

Your Call Is Important to Us . . . Please Hold

BY CHRIS BOND & MARK CAMACK

Focus groups and usability testing help avoid the trap of unhelpful automated phone systems.

• ☎ •

HAVE YOU EVER CALLED A BUSINESS, ONLY TO FIND yourself trapped in its automated phone system? Ever listened indefinitely for a menu choice that wasn't there? Technological advancement brings with it a certain amount of aggravation, but let's not blame it on the technology – the designers need to do a better job. It's been said that a product's maturity is reflected in the complexity of its design. In most cases, products that have been around for awhile have simple designs and are easy to use. Unfortunately, automated telephone systems haven't been around that long, and it's clear they still have a long way to go.

In their 1996 *Ergonomics in Design* article, Meyer and Seagull ("Where Human Factors Meets Marketing," pp. 22–25) emphasized the importance of close collaboration between human factors and marketing professionals. We've taken that advice and have collaborated on several projects. Although we didn't know what to expect, we experienced a convergence rather than a divergence between the human factors and marketing disciplines. In this article we relate one recent, successful collaborative activity with improvements to an automated phone system.

PHOTOS COURTESY OF THE AUTHORS

Pitfalls of Automated Phone Systems

The goal of our project was to overhaul the menus and prompts for an electric utility. Our experience with customers of other utilities across the nation has shown that callers are interested in clarity, simplicity, and relevance. As a rule, the options they need most often should be presented first. But when I call my bank, the first option I hear is, "To open a new account, press 1." What? I (like most callers) already *have* an account. The new account customer calls once; the existing account callers are the repeat business. So what happened here? It's obvious the bank doesn't have a user-centered design process.

If you blow it, they'll never use your automated system again.

In our research we've discovered what really irritates people about telephone-based interfaces. Remember, you only get one chance to make a good first impression: If you blow it, they'll never use your automated system again. They may even tell their friends and family how terrible it is. The box below lists our top 10 ways to blow it with customers.

The fact is, these sources of irritation – all caused by interface problems – are easily fixed with improved wording, navigation and control, dialog structure, and so on. And although no single design flaw is a major problem, the cumulative effect of multiple flaws will result in failure. The best approach is to base the design on guidelines that contain the "blueprint" for your system and then thoroughly test the application before releasing it. Usability tests and focus groups will reveal any sources of irritation that remain.

Chris Bond (an employee of the electric utility) documented the design guidelines and performed usability tests. Mark Camack, a market research consultant, provided input to the system redesign and conducted focus groups.

Design Guidelines

Guidelines are an indispensable resource when making design decisions. They provide the mechanism for documenting your choices so that later, if anyone asks why a product or system was designed a certain way, you'll have a ready answer. As you learn about and test the interface, your guidelines can be used as a repository for documenting that knowledge. Most important, rules you establish for the application need to be documented before they can be effectively applied.

At Portland General Electric, we developed a set of guidelines for our interactive voice response (IVR) system. This document contains a system "philosophy" that describes the experience we want our customers to have. It also describes our operating protocol, our IVR Steering Committee, and a quality assurance program.

•☎• Pet Peeves of Customers •☎•

1. Put them on hold and tell them, "Your call is important to us." Repeat this message every 30 seconds.
2. Ask them to enter information, then transfer them to a representative who will ask them for the same information.
3. Don't include the standard "0" option to speak to a representative at any time. If they do press zero, tell them, "I'm sorry, that's an invalid choice. Please try again." Or send them to another menu of automated selections.
4. Overwhelm them with a multitude of menu choices and make the menu structure five or six layers deep. Or skip selection numbers or list selection numbers out of sequence.
5. Use the word "now" after every option.
6. Use run-on sentences and illogical groupings for each menu option. This way, a single option sounds like a whole bunch of options.
7. Organize the hierarchy of information to reflect internal politics and power instead of customer-based activity. Similarly, use internal jargon instead of plain English.
8. Put the most popular options last; force callers to listen to long lists of irrelevant options before giving them what they want.
9. Alarm them with the warning, "We've recently improved our system," followed by "Listen carefully to the following options," implying that the new version is complicated and they'd better pay attention.
10. Lead them down the garden path and, after they press a couple of invalid keys, disconnect the line.

The guidelines and examples of how to apply them are organized according to menus and prompts, general wording, navigation and control, dialog structure, and feedback and error handling. At the end is a section entitled "Design Decisions," which shows before-and-after examples and lists the reasons behind changes and additions.

Many of the guidelines are derived from our own test results and from studies in the literature (e.g., Schumacher, Hardzinski, & Schwartz, 1995; Fowler & Stanwick, 1995). Here are a few examples and how we applied them.

Limit the menu tree to three levels and menu options to four or fewer. Keep it simple; the menu hierarchy should be no more than three levels deep. We've found that as the number of levels (and options per level) increases, so does the number of errors. This applies to prompted options, not standard commands for previous menu, repeat, or 0 to speak to a representative.

Eliminate padding, repetition, and conversational pleasantries. Avoid saying "now" after each option and phrases such as, "Please select from the following options." Also, avoid error messages, like "We apologize for any trouble you may be having," when the wrong keys are pressed. Don't offer greetings such as "We hope you enjoy a safe and happy holiday." Empty phrases and repetition serve no purpose other than to increase listening time. Each additional word also increases the possibility of frustration, confusion, or error. To avoid these problems, we counted the number of words in our main menu, and then streamlined them from 118 words to 40.

Organize menus by importance, followed by frequency of use. In the electric utility business, the two most important things are outages and emergencies, such as downed power lines. To handle these kinds of problems, we created an outage branch and a customer service branch. The first words customers hear are, "If you are calling about a power outage, downed power line, streetlight, or other power problem, press 1. For all other business, press 2." In the customer branch main menu, options are listed in the order determined by the volume of calls received about each matter: The first option, to start or stop service, represents about 300,000 calls per year; the second option, regarding billing or payment matters, 250,000 calls/year; the third, requesting field service, 44,000 calls/year; and the last option, regarding marketing, 30,000 calls/year. This strategy has been an effective defense against the internal politics of various departments jockeying for a position on the menu. See the illustration on page 12 for an example of the menu hierarchy.

Provide "escape hatches." You should always provide the option to press 0 to speak to a representative at any time. But, more important, include alternatives at the end of a path. Don't leave callers in a dead end or a continuous loop. This is especially important for speed dialers, who sometimes "jump the gun" by pressing a key before hearing the entire menu. As they arrive at the end of a path and realize they goofed, don't force them to hang up and redial. Also, always give them the option to repeat a menu and return to a previous menu.

Avoid language that might be taken as patronizing. As we said earlier, telling customers how important their call is while they're stuck on hold is pretty patronizing. Another patronizing message is, "We are experiencing unusually high call volumes due to the holiday. Please try again later." What this does is tell your customers you don't know how to run a call center (you should staff up for increased calls during or after holidays). Let the caller decide if he or she wants to remain on hold by announcing queue times. Customers react very positively to announced queue times; knowing how long they will have to wait reduces tension and anxiety.

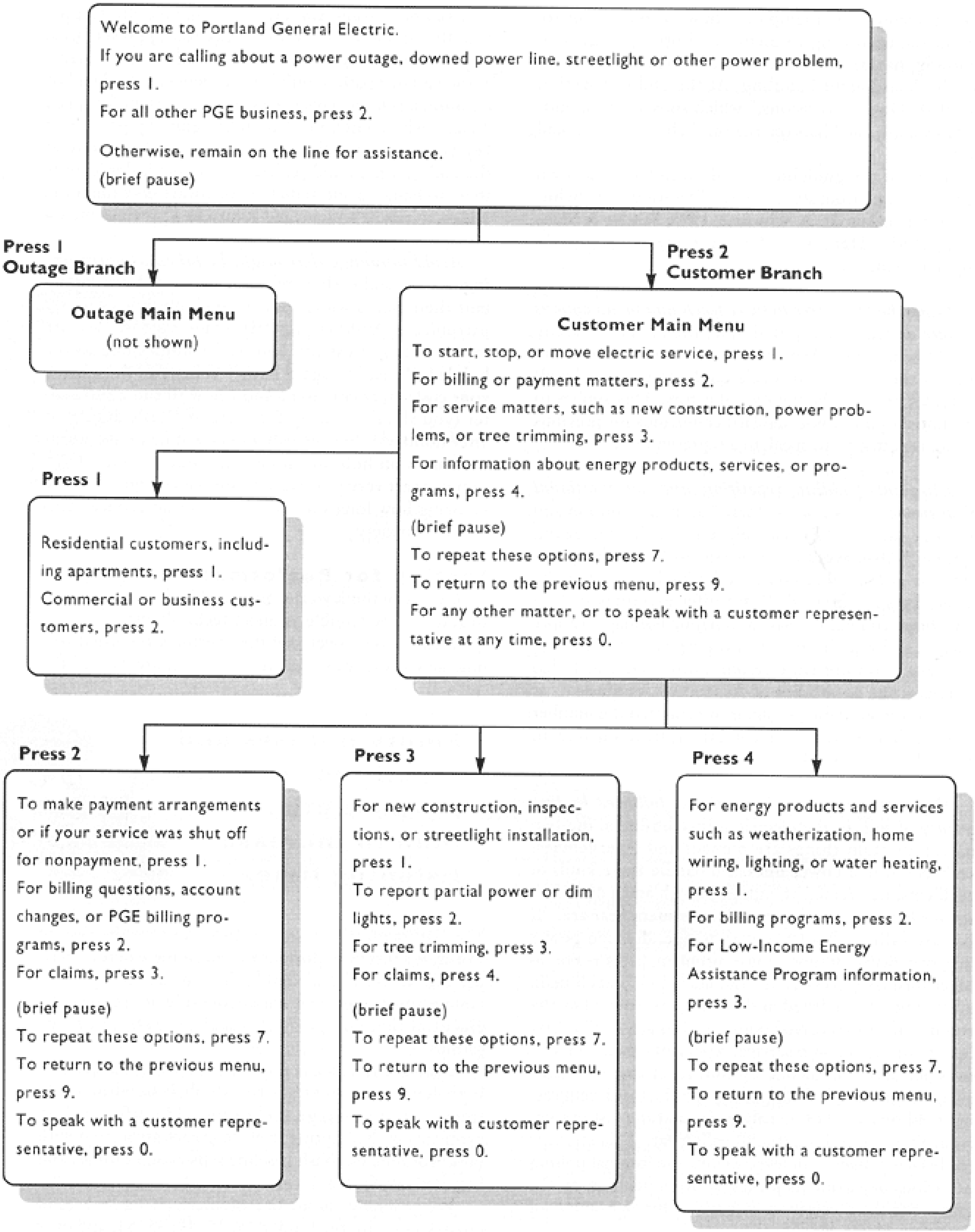
Testing for Performance

Once you think you've got a usable design, you need to test it. The trouble is, most technologists feel they have their bases covered if they perform the usual volume and stress tests, unit tests, function tests, and the like. Although it's important to make sure the system works, it's just as important to determine whether people find it useful and usable. To ensure that performance and preference are adequately addressed, it's essential that you perform usability tests and focus groups.

Empty phrases and repetition serve no purpose other than to increase listening time.

Usability tests focus primarily on performance. The high-level objectives are to establish baseline cycle times, reduce average cycle times, identify wording problems with scripting, reduce errors associated with poor wording, establish baseline satisfaction ratings, and increase satisfaction levels.

We generally recruit 12 customers, trying for a good cross-section by age (under 35, 35–44, 45–64, and over



Welcome to Portland General Electric.
If you are calling about a power outage, downed power line, streetlight or other power problem, press 1.
For all other PGE business, press 2.

Otherwise, remain on the line for assistance.
(brief pause)

Press 1
Outage Branch

Outage Main Menu
(not shown)

Press 2
Customer Branch

Customer Main Menu
To start, stop, or move electric service, press 1.
For billing or payment matters, press 2.
For service matters, such as new construction, power problems, or tree trimming, press 3.
For information about energy products, services, or programs, press 4.

(brief pause)
To repeat these options, press 7.
To return to the previous menu, press 9.
For any other matter, or to speak with a customer representative at any time, press 0.

Press 1

Residential customers, including apartments, press 1.
Commercial or business customers, press 2.

Press 2

To make payment arrangements or if your service was shut off for nonpayment, press 1.
For billing questions, account changes, or PGE billing programs, press 2.
For claims, press 3.

(brief pause)
To repeat these options, press 7.
To return to the previous menu, press 9.
To speak with a customer representative, press 0.

Press 3

For new construction, inspections, or streetlight installation, press 1.
To report partial power or dim lights, press 2.
For tree trimming, press 3.
For claims, press 4.

(brief pause)
To repeat these options, press 7.
To return to the previous menu, press 9.
To speak with a customer representative, press 0.

Press 4

For energy products and services such as weatherization, home wiring, lighting, or water heating, press 1.
For billing programs, press 2.
For Low-Income Energy Assistance Program information, press 3.

(brief pause)
To repeat these options, press 7.
To return to the previous menu, press 9.
To speak with a customer representative, press 0.

The menu hierarchy for Portland General Electric's automated telephone system is only three levels deep.

65), attitude toward automated phone systems (ranging from acceptable to not acceptable), and experience (ranging from rejecter to frequent user).

There are a few key factors to look for when conducting user tests on telephone interfaces, such as task failures, unrecoverable errors, task times, and navigational choices. With the main emphasis on throughput, you need to identify the path(s) for accomplishing various tasks. We created an organization chart to show the hierarchy of information and can see, at a glance, all potential call flows, the wording of options, error handling, and general business rules.

We discovered early on that it's impractical to have participants think out loud, especially when speaker phones are used. Participants in our tests would speak over the dialog and lose their train of thought. It's important to let them perform the task and then ask questions immediately after the task has been completed.

In a few instances, participants would say they were confident that they performed the task accurately when, in fact, they had failed to accomplish the stated goal. This is a case of having a pleasant experience but nonetheless ending up speaking to the wrong person and having to be transferred to the right department (another cardinal sin in phone system design). Though it appeared in these instances that preference outweighed performance, in a real-life situation, these users would have been frustrated.

Typically, the participant is given a simple scenario, such as "You've just received your bill and realize you won't be able to pay the full amount." This task is for a "payment arrangement." During the task we closely observe which options the participant selects on the touch-tone keypad. Data sheets are used for recording observations. Each sheet lists the participant's name, task number, elapsed time, and whether the task was successfully completed. The sheet is also used to document errors (such as wrong menu choices), facial expressions, and comments made during or after the task.

Table 1 lists task times for one of our usability tests, excluding incomplete (failed) tasks, as these would adversely affect successful task times. The empty cells make it easy to identify task failures. As in most usability tests, patterns begin to emerge; tasks that took the longest are the ones with the most errors and problems, and those that were unsuccessful were almost always more than three levels deep in the menu hierarchy.

As long as the correct outcome is reached, it doesn't matter how callers get there.

Providing multiple paths to the same goal is an effective way to accommodate different expectations (similar to the personal computer desktop environment, where there are often multiple ways of doing the same thing). As long as the correct outcome is reached, it doesn't matter how callers get there. In a couple of

TABLE 1. TASK TIMES FROM THE USABILITY TEST

	Participants												
Tasks	1	2	3	4	5	6	7	8	9	10	11	12	Average
1. Move	00:45	00:45	00:41	00:49	00:43	00:40	00:52	00:46	00.43	00:44	01:24	00:45	00:48
2. Payment arrangement	00:50	00:46	00:40	00:42	00:41	00:42	01:09	00:45		00:47	00:49	00:51	00:47
3. High bill inquiry		00:56	00:53	00:53		01:08	00:51	00:54	00:48	00:56	00:54		00:55
4. Meter reading	03:54	02:43	02:22			01:55	02:02		02:11	02:32		02:32	02:31
5. New construction			01:01						01:03	01:07	01:56	01:01	01:14
6. Disconnect/cutout	01:29	00:52	00:44	00:48	00:39	00:52	00:39		00:36	00:41	00:46	01:47	00:54
7. Business hours	02:23	02:26					01:52		01:55				02:09
8. Program information	01:00	00:58	00:58	00:58	00:55	00:55	00:51	00:49	00:43		01:58	00:51	01:0c
9. Account info change	02:27	00:57	01:06	01:26	00:54		00:50		00:47	01:13	00:57		01:11
10. Equipment problems	00:42	00:42	00:41	01:04	0:59	00:57		00:49	00:37	00:39	01:04	00:45	00:49
11. Budget pay	00:52	00:56	00:46	00:54		00:43	01:04	00:51	00:45	01:01	00:50		00:52
12. Claims		01.23		01:25		01:56			01:01	01:10	01:06	02:43	01:32
Average	**01:36**	**01:13**	**00:59**	**01:00**	**00:48**	**01:05**	**01:08**	**00:49**	**01:01**	**01:05**	**01:10**	**01:42**	**01:13**

TABLE 2. USER PERCEPTIONS FROM THE USABILITY TEST

	Participants												
Topics	1	2	3	4	5	6	7	8	9	10	11	12	Average
1. Overall ease of use	9	10	10	10	7	9	7	6	9	8	8	5	8.17
2. Number of options	8	9	10	10	8	10	9	5	10	10	7	5	8.42
3. Consistent menus	10	10	10	9	7	10	7	3	9	10	10	7	8.50
4. Overall impression	10	10	10	10	8	8	8	5	10	10	9	5	8.58
5. Logical organization of info	7	10	8	9	9	8	6	5	9	8	7	5	7.58
9. Return to previous option	10	10	9	10	7	10	5	7		10	8	8	8.55
10. Instructions and prompts	9	8	10	10	8	8	10	5	10	10	8	6	8.50
11. Terminology	9	9	10	10	9	9	7	4	7	10	10	4	8.17
12. Ease of finding info	8	10	10	10	7	8	7	8	9	8	8	5	8.17
13. Ease to access a person	10		10	10		7	3	3	10	10	9	8	8.00
14. Confidence level	9	9	9	9	8	8	9	4	9	8	9	8	8.25
15. Voice for instructions/prompts	10	10	10	10	9	3	6	9	10	10	5	9	8.42
16. Ease of error recovery	10	9	8	10	5	8	9	4	8	9	10	7	8.08
17. Use of zip code	10	10	10	10	10	10	9	4	10	10	10	10	9.42
18. Use of voice form	10	10	5	10	7	4	7	3	6	10		4	6.91
Average	**9.27**	**9.57**	**9.27**	**9.80**	**7.79**	**8.00**	**7.27**	**5.00**	**9.00**	**9.40**	**8.43**	**6.40**	**8.25**

instances, we thought we designed the menu perfectly, only to find that half the participants looked for the option down a path other than the one we intended them to use. Our solution was to build in redundancy and put the option on both branches; this minimized task failures and ultimately helped people get where they needed to go. In one test, Task 10 (power problems) was successfully completed by 11 participants; 6 selected the outage branch path and 5 selected the "all other business" path. Although task times varied for the different paths (the outage path is slightly faster), the end result is the same – mission accomplished by the customer.

Some of our guidelines were validated in the usability test. As always, in the real world, there are exceptions to the rules, because of either political pressures or group consensus. In one or two cases, the menu hierarchy was more than three layers deep (remember that the literature indicated a limit of four). To no one's surprise, usability tests showed a sharp degradation in task completion for any of these tasks. In one scenario, only 4 of 12 were successful (the path was five levels deep). In another, only 6 of 12 were successful (four levels).

We used a posttest questionnaire to gather satisfaction ratings from participants. Reactions were based on a sliding scale (1–10, with 1 being the least favorable). Meaningful descriptors were used at each end of the scale; for example, difficult to easy, inconsistent to consistent, frustrating to satisfying, and confusing to clear. Table 2 shows preference ratings of our usability test participants. The highest rating (9.42) was for entering a zip code to identify a service territory. Participants' overall impression of the system was favorable (8.58), as was the rating of ability to return to a previous option (8.55). The lowest ratings were for the use of a voice form, requiring the caller to leave a message (6.91), and logical organization of information (7.58). We suspect that poor performance (high failure rate), combined with a negative experience on tasks with menu paths four and five levels deep, explains the low rating on the logical organization of information. This demonstrates how one or two negative experiences can adversely affect customers' perceptions.

We included space for written responses on the posttest questionnaire. Participants could tell us what they liked best and least about the system and the one thing they would tell designers to change. The combination of sliding scale and free-form comments provides valuable insight on users' reactions to the design.

Testing for Preference

Speed of throughput and accuracy are only part of the design equation – it's important not to stop with usability testing. The customer experience (their reactions to the design) should be tested as well. Focus groups (typically a group of 10 research participants and a moderator) reveal users' preferences. They're especially helpful in diagnosing specific areas of concern to target constituencies, and in identifying trouble spots in scripting, pace, or tone. Viewed through a two-way mirror, focus group participants provide developers with firsthand reactions to proposed system menus and applications. Participants go through a series of role-play exercises, giving developers the chance to observe reactions to experience during and after a call. When customer reactions are negative, developers are quick to correct any potential scripting ambiguity before a particular application is launched.

Residential focus groups used for gauging phone system interaction tend to be recruited from among recent callers to the utility, with group constituency based on homogeneous lines of age or income (the two strongest connections to ability and willingness to use automated phone systems). This usually results in discrete groups of "Generation X'ers" (ages 18–29), "Boomers" (ages 30–50), and mature customers (51 and older). Commercial groups are typically conducted with those who frequently call the utility company (e.g., apartment managers, builders/developers). This helps to establish important baseline measures of predisposition, expectations, and, ultimately, success or failure.

In focus groups, as in usability tests, participants aren't given any special instructions about how to use the automated system or prototype. They're given a simple scenario, such as the following: "Your power just went out. Call the electric company." They're also told that the scenario may contain false information in order to test the system's default messages. (This encourages participants' willingness to discuss potential "incorrect" paths they may take, even if the scenario does not contain an intentional error.) Finally, scenarios contain no key words or clues that may affect test results.

Participants call the system individually and then write down their overall reaction, likes and dislikes *prior to* discussing the experience with anyone else in the group. This enables the moderator to ask each participant to read what he or she wrote and avoids the potentially misleading "bandwagon effect" of focus groups, in which a dominant participant sways the opinion of others. The written exercises are then collected and evaluated afterwards.

When consulting for the project at Portland General Electric, Market Strategies, Inc., evaluated a dozen call scenarios (about three or four in any single group is the limit, with discussion). Most of the call scenarios survived intact, minor changes were suggested to a couple of situations, and one or two went back to the drawing board. One latter case with a multiple-path error was later corrected when we created redundant paths to the same end.

In cases when a prototype application is not available, calling another company's automated phone system during a focus group session can illuminate issues similar to your own, such as how the company asks for or gives account information. The moderator uses a speaker phone which enables the entire group to repeat, isolate or compare particular words, phrases or tempo of the system in question. In this way focus groups can be totally complementary with the one-on-one approach of usability tests.

An important principle of human factors is that the mind seeks patterns and rules and will follow them if they are apparent. Showing customers the text of the menu hierarchy on large flipcharts sometimes led to interesting observations. If more than four options were listed, some would say, "You can't do that – there's only supposed to be four options per level. That menu has five." This was fascinating, considering that we hadn't discussed our design strategy with the participants.

Conclusion

Performing usability tests and conducting focus groups at Portland General Electric have reinforced what we already know and taught us what we need to learn. Perhaps the best indicator of performance/preference alignment is when a participant confidently breezes through a task and then hangs up, smiles, and says, "That was easy." We're getting better at it. In a recent usability test of our automated outage system, there were no task failures or unrecoverable errors, and task times were faster than predicted. This is good news, especially considering that half the participants felt that telephone-based interfaces were unacceptable. Posttest questionnaires showed that users' overall impression of the system averaged 9.08. Not bad, not bad at all.

Chris and Mark have collaborated on projects since this one and taught a workshop earlier this year on how to conduct focus groups and usability tests for telephone systems. The key to success is to work together. Our successful collaboration on automated phone systems is based on a mutual respect for two disciplines that share the same objective – customer satisfaction.

References

Bond, C. J. (1998, February). *Portland General Electric interactive voice response (IVR) design guidelines.* Internal publication, Portland, Oregon.

Fowler, S. L., & Stanwick, V. R. (1995) *The GUI style guide.* Cambridge, MA: AP Professional.

Meyer, J., & Seagull, F. J. (1996, July). Where human factors meets marketing. *Ergonomics in Design*, pp. 22–25.

Schumacher, R. M., Hardzinski, M. L., & Schwartz, A. L. (1995). Increasing the usability of interactive voice response systems: Research and guidelines for phone-based interfaces. *Human Factors, 37*, 251–264.

Chris Bond is a user interface consultant at Portland General Electric, One World Trade Center, 121 SW Salmon St., Portland, OR 97204. He has written several Windows-based interface design guidelines and is responsible for implementing human factors principles and techniques across the corporation. Mark Camack is a vice president of Market Strategies, Inc., in the Energy Research & Consulting Division. MSI serves the energy, telecommunications, high-technology, and health care industries.