Gwizdka, J. (1999). You have pressed an invalid key, please try again… Perceptual and Cognitive Limitations of Interactive Voice-Response Systems (a.k.a. Phone-Based Interfaces). IML Technical Report. IMLTR-99-02. 11pp.

You have pressed an invalid key, please try again...

Perceptual and Cognitive Limitations of Interactive Voice-Response Systems (a.k.a. Phone-Based Interfaces)

Jacek Gwizdka

1.0 Introduction

Most of us have used automated phone information systems. There is a wide variety of such systems, for example: government information systems, transportation schedules, airline reservation systems, technical support systems, voice mail systems, telephone banking, on-line trading, and university course registration. Anecdotal evidence suggest that, these systems are generally not liked and that users are often dissatisfied with them. This study sought to find some of the reasons for this state of affairs.

2.0 Description of Interactive-Voice Response Systems

Interactive-Voice Response (IVR) systems belong to a group of phone-based interfaces[1], that is, interfaces in which interaction with a system takes places through the phone. IVRs employ as user-input the touch-tone™ technology. Wide availability of this technology in North America[2] and its relatively high reliability make it a good choice for simple phone input. IVR systems, as the name suggests, respond by using voice. Thus, in IVR applications, telephone's keypad is the input device, while telephone's earpiece is the output device.

IVR systems can employ two main interaction styles: *prompted* and *command-driven*. In prompted systems, a voice menu is presented to the user who responds by pressing a key (or a sequence of keys). In command-driven systems, user enters commands without being prompted for them, and thus user must know the commands in advance. This study was focused on prompted systems, as these systems are much more common and are used in situations where no prior user experience or training can be assumed.

Characteristics of media and devices used in IVRs as input and output, affect user interaction in an important way, and thus are specially significant to our analysis. Phone keypad is a limited, but well known input device. Therefore, we should not expect that users will have difficulties in using this device. On the output side, speech is serial, it takes time and is transient, that is, once uttered, sounds of speech are permanently gone. If needed later, information communicated by voice has to be somehow recorded. In the absence of external aids, the recording has to take place in memory. We can therefore expect, that user interaction employing speech as output will be affected by the limitations of human working memory.

1. Phone-based interfaces can also employ voice input, but they do not belong to IVR systems. Due to unreliable speech recognition systems using voice input are still not very popular. The situation is, however, starting to change recently (see also Footnote 2.).
2. In Europe the situation is quite different (Gardner-Bonneau, et al. 1998); touch-tone is not widely available. Telephone interfaces have not been wide spread, and only the recent improvements in automatic speech recognition (ASR) allow for a wider use of phone-based interfaces.

3.0 Study

Rather than investigating one IVR system, we chose to study several systems in order to find common problems that are representative for this type of human-machine interfaces. Information for the study has been gathered by interviewing three IVR users. Additionally, further systems were directly explored by the experimenter.

The studied systems included: Revenue Canada Customs Infoline, Stats Canada Automated Information System, Ontario Fines Collection, two voice mail systems: Mitel and Octel (Bell), and a telephone banking (Toronto Dominion Bank). With the exception of the voice mail systems, the studied systems are used on a one-time or on a sporadic basis. Users of such systems can, thus, be considered novices, since the effect of learning is negligible.

4.0 The Framework for Analysis

Analysis has been performed by using a model of a user interacting with an IVR. This model is based, in part, on Huguenard (1997) and on some ideas from Norman (1991). The model describes steps in interaction with an IVR and shows implications on working memory. The model is shown in Figure .

User is listening to menu options. Each option is being evaluated and matched against the current user goal, while user is listening to it. Operation of evaluation and matching may require some processing. According to a theory of working memory (henceforth denoted by WM) proposed by Just and Carpenter (1992)[1], both storage and processing demands determine the load on WM. Due to the required processing, the load of WM can temporarily increase (denoted in Figure by **WM+t**) during this step. This increase of load can in turn lead to a failure of working memory, if the load, is over its capacity limit. If this happens, the previously stored chunks of information (for example, a previous option) can be forgotten.

The result of matching an option to the current goal can, in a general case, be placed on a scale between a perfect “Match” and “No Match”. In these border cases, user action is marked by transition (1) or (4), respectively. Load of the WM does not change (denoted by **WM0**) during these transitions, since the selected or discarded option does not have to be remembered. Probability of a “Weak Match” depends both on the presentation (wording, speech quality) of menu options, and on the formulation of the current user goal.

1. This work is cited here after Huguenard et al. (1997).

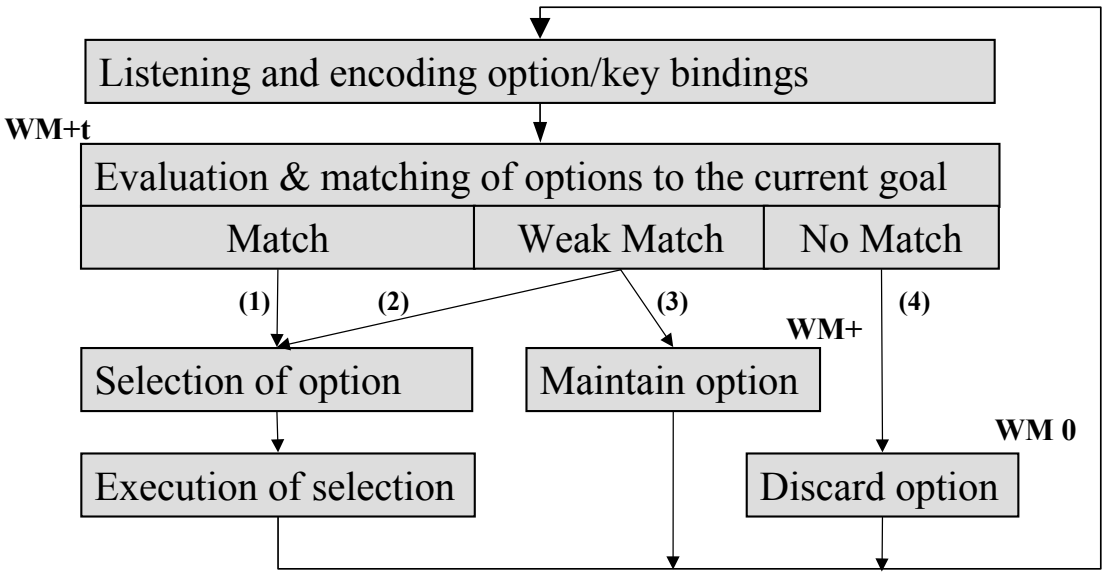

FIGURE 1. User model

In case of the "Weak Match" the user has two choices, take a risk and select the weakly matched option (2), or remember the option (3) and continue listening to further options. In the latter case, the weakly matched option increases load of the WM (denoted by **WM+**) because the information about the weak match has to be stored. Only weakly matched and not executed options need to be remembered by the user. The choice between (2) and (3) depends on how risk averse the user is and on the user's style of decision making.

This model allows us to predict situations in which interaction with IVRs can lead to user errors, and the type of problems that users potentially might have. Conversely, the model allows us to find explanations for problems observed during user interaction with IVRs.

5.0 Findings and Analysis

This section describes six commonly found problems during interaction with IVRs, and illustrates them with examples from the studied systems.

5.1 Long and complex prompts

Listening to complex options requires a lot of attention. Evaluation and matching of long and complex prompts is difficult and can lead to a temporary increase of WM load. In addition to the above problems, the very long prompt presented in Figure 2 is always preceded by an additional, not

directly relate message. Moreover, wording of this prompt does not match option names at the lower level. As a result of such a poor menu design, users are confused and have to repeatedly listen to this option.

- **For statistics on the consumer price index, population research, such as the cost of living, the inflation data, and COLA, as well as unemployment rates, population figures, or 1996 (pause in message!) census data press 1**
  - **For consumer price index data press 1**
    - 1 - 4
  - **For labour force data press 2**
    - 1 - 4
  - **For population estimates, births & marriages press 3**
  - **For 1996 census press 4**
- Various Stats Canada services press 2
- To order publication press 3
- Survey respondents press 4
- For listing of today's releases press 5
- For info on obtaining death, birth or marriage certificate press 6
- For hours of operation press 7

FIGURE 2. An example of a complex prompt

5.2 Semantically similar menu options

5.2.1 Relative position of similar menu options

An example of closely related menu options that do not immediately follow one another is show in Figure 3. If the first option is only weakly matched to the user goal, then it needs to remain stored in WM until the next related option is played. Since the other option is so far away, the user is likely to select the first encountered option after listening to a few more unrelated options. This design of the menu options potentially introduces choice errors, or the necessity of repeated listening to all options.

- 2 - info on payment
  - **1 - the due date passed, can I still pay?**
  - **2 - .....**
  - **3 - .....**
  - **4 - what if I don't pay by the due date?**
  - **5 - .....**
  - **6 - .....**
- Message: Press * to repeat *("*" does not work, menu repeats automatically!*)

FIGURE 3. An example position of related menu options

5.2.2 Options semantically including one another

Another problematic situation occurs when semantically related options that include one another are presented at the same menu level. This situation usually arises due to wording of prompts, and not

due to the actual relationship between options. Two examples are provided in Figure 4 (bill payment includes, or may include, post-dated bill payment), and in Figure 5 (postal importation includes one kind of personal importation). Evaluation and matching of such options is difficult without hearing all of them. If user makes an early choice, it is likely to be wrong. This type of menu design can also lead to repeated listening of the whole menu.

- 2 - Financial transactions
  - **1 - transfer**
  - **2 - bill payment**
  - **3 - post dated bill payment**

FIGURE 4. An example of semantically overlapping menu options

5.3 Menu structure

Complex menus, whether visual or auditory, always increase the difficulty of a task. Common guidelines and standards recommend broad visual menus and deep-narrow (3-4 items per menu level) auditory menus (Norman, 1991; Schwartz and Hardzinski, 1993; Schumacher, Hardzinski, and Schwartz, 1995; Marics. and Engelbeck, 1997). However, according to the previous studies (reported in Huguenard, et al, 1997) the choice of semantically different options, that are likely to appear in deep and narrow menus, takes more time, and hence, the optimal number of auditory menu options should depend on a particular application.

In an example presented in Figure 5, we observe both a lot of options and a deep menu structure. This menu design can lead to an increased load on WM due to disambiguation processing. Futhermore, big number of options leads to longer menu repeat times. As a result we observe choice errors and a long time to complete task.

FIGURE 5. An example of a complex menu (fragment) and of semantically overlapping menu options (1 and 4)

5.4 Prompting for actions

Prompts can be designed in an action-goal or in a goal-action format. Action tells users which key to press, goal tells what will happen as a result of pressing the key. According to the previous studies (Schumacher, Hardzinski, and Schwartz, 1995) goal-action format corresponds better to the cognitive makeup of the task, and, thus, it is recommended in the prompt design. In an example shown in Figure 6, action-goal sequence, rather than recommended goal-action, is used in wording of some of the options. Moreover, the style is not kept consistent across all prompts of this system, and some of them use the more proper goal-action format. As a result users make choice errors and repeatedly listen to the options. Additionally, this particular system did not offer return to the main menu. Therefore, repeating the main options required repeated calling the system.

- For Credit Card payment, **press 1**
  - **Press 1 if your document is entitled infraction notice**
  - **Press 2 if your document is entitled payment notice**

FIGURE 6. An example of Action-Goal versus Goal-Action prompt structures

5.5 Speech quality

Quality of speech is important for its correct perception. Speech quality has several dimensions. In a couple of the examined IVRs, two dimensions seemed to be particularly problematic: loudness and smoothness. In one of the systems (from example in Figure 5) loudness is not on the same level even within the same prompts. In another system, some voice responses are created based on user inputs or based on database contents. Such system responses cannot be prerecorded, but are created from small prerecorded pieces. If these pieces are not recorded at the same loudness level and if they are not smoothly put together, we observe an effect similar to hiccups (example in Figure 7). This makes perception more difficult and requires from users more additional attention. In the case of this particular system, the situation is made even worse by the lack of the possibility to repeat. As a result users are repeatedly listening and entering the data.

- You wish to pay ^ ***Bell Canada*** ^ the amount of ^ **63** ^ $ and ^ **93** ^ cents from your ^ **checking** ^ account and forward it on ^ December ^ **15** ^ **1998**

FIGURE 7. An example of a system response composed of separate pieces of information

5.6 Consistency of user input

Consistency can be considered at two levels: within one system and across several applications. As illustrated by three examples from different systems in Figure 8, input keys for some operations, are not consistent within one system. Different keys are used to perform the same command (examples 2 and 3 in Figure 8), or the same key is assigned a different, or, even worse, sometimes only a slightly different, meaning (example 1 in Figure 8). Placing of missing character keys (“Q” and “Z” on phone

keypads), and naming of “#” key is often different in different systems (example 5 and 6 in Figure 8). Lack of consistency results in choice errors, leads to confusion and requires increased attention.

1. “#” called “return to main menu” brings back either to the very top level or to the next level "main” menu
2. Saving messages “9”, but sometimes “2”
3. Recording messages “R” (7), but “A” (2) to re-record
4. “#” often end of entry, one system used it as return to the main menu
5. “Q”,”Z” missing - assignment not standard
6. Naming of “#” not always standard: “number sign key” vs. “pound key”

FIGURE 8. Examples of lack of consistency in entering system commands

5.7 Summary of findings

In summary, the most often observed were the choice errors, that is, selecting by users “wrong” options. Mistakes were due to: inconsistencies in systems, failure of WM, misunderstood menu options (users could not relate their goal to the current set of choices). The same problems caused repeated listening to menu options and resulted in longer task completion times. As a result, users of the IVR systems are often dissatisfied. It should be noted, that, as predicted, the data entry errors are practically not observed.

6.0 Recommendations

6.1 Consistency

First recommendation, and the technologically simplest one to implement, is to keep design consistent. Consistency within one system should be obligatory in all its respects. For example, in respect of: user input, menu navigation, prompt format and wording. Consistency also applies to speech loudness and voices used for recording of prompts. Also important is Consistency across different systems, especially in the case of systems that are used only sporadically, whereby one user is more likely to access several different IVR systems than one system repeatedly. This type of consistency is more difficult to enforce, since IVR systems are built and installed by various companies. Nevertheless, IVR manufacturers should follow industry and ISO standards (for example design guidelines published by Ameritech in Schwartz and Hardzinski, 1993).

6.2 Menu structure

Menus should be carefully designed and always brought up to date. Complex menu structures should be avoided. The typical recommendation (Schwartz and Hardzinski, 1993; Schumacher, Hardzinski,

and Schwartz, 1995; Marics. and Engelbeck, 1997) for auditory menus is to design for depth allowing up to 3-4 options at one menu level. This guideline should not be followed mechanically and the trade-off between breadth and depth of menu structure should be considered from the perspective of each IVR application. As argued by Huguenard (Huguenard, et al, 1997), users of some[1] IVR applications may benefit from broad rather than deep and narrow menu structures. Huguenard et al. shown that broad menus may effectively decrease the WM load.

Users of IVR systems with complex menus may support themselves by using external aids for temporary recording of information. These aids can help users to off-load their WM.

6.3 Speech

A uniform quality of speech should be ensured in IVR systems. Transitions between prompts should be smooth and “hiccup-less”.

Time is often an important factor for IVR users. It is thus very surprising, that only one of the examined systems offered optional faster playback of systems messages. Speech can be sped up to 1.5-2.0 times without too high demands for perception and attention (Arons 1992). On the other hand, IVR system designers should also take into account users for whom English is the second language. This user population may benefit from the possibility of slowing down system responses.

6.4 Navigation control

Although this study did not focus on problems with user control of navigation, these issues have been indirectly mentioned while discussing other findings. In general, users should be given more control, for example, skipping options ahead and returning to the previous and the top menu level, should be available at any time.

7.0 Conclusions

The problems observed in IVRs rarely lead to “fatal” errors, where “fatal” is used in a sense of total task failure. However, frequent choice errors and unnecessarily long time to complete tasks often lead to user dissatisfaction.

IVRs are often introduced in the first place to the advantage of the service providers. IVR systems off-load human resources by moving part of interaction with customers to the automated systems. Oftentimes, users may have no other choice, but to use these systems. Obvious deficiencies in IVRs, and especially lack of certain navigation features, can make us wondering, whether these are design features introduces with the purpose of effectively limiting user time on-line.

1. Huguenard used in his experiment a university course entry system.

Recent rapid development of information services provided through the Web may decrease with time the significance of IVRs. However, phones with touch-tone still remain much more ubiquitous than web access devices. Moreover, there are projects underway that aim to integrate web access with telephone access and to offer web browsing through the phone interface (Atkins, et.al, 1997). Thus, IVRs are not disappearing due to the fast growth of the Web, they can be seen rather as complementary way of accessing information. In the near future, the role of IVRs is likely to increase rather than to diminish. It becomes therefore crucial to improve usability of IVR systems, to introduce user-centered design practices and to follow IVR design standards.

8.0 Acknowledgments

I would like to thank Anna Hart, Leonardo Ruppenthal, and Calum Tsang for taking time to answer my questions and for sharing their experiences with IVRs.

9.0 References

Arons, B. (1992). Techniques and Applications of Time-Compressed Speech. *In AVIOS '92: Conference of the American Voice Input/Output Society*, Minneapolis, MN.

Atkins, D.L. Ball, T., Baran, T.R., Benedikt, M.R., Cox, K.C., Ladd, D.A., Mataga, P.A., Puchol, P., Ramming, J.C., Rehor, K.G., and Tuckey, C. (1997). Integrated Web and Telephone Service Creation. *Bell Labs Technical Journal*. 2. (1). Winter 1997.

Gardner-Bonneau, D. Delogu C., Green C., Volaitis L., Lindeman M., Laux. L. (1998). INTERACTIVE VOICE RESPONSE (IVR) SYSTEMS AS AUTOMATIC SPEECH RECOGNITION (ASR) COMES OF AGE. Panel. *In Proceedings of the Human Factors Society 42nd Annual Meeting*. 444-447

Huguenard, Brian R. Lerch, F. Javier, Junker, Brian W., Patz, Richard J.and Kass, Robert E. (1997). Working-memory failure in phone-based interaction. *ACM Transactions on Computer-Human Interaction*. 4(2). 67 - 102.

Just, M.A. and Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*. 99 (1). 122-149.

Kroemer, K.H.E, Grandjean, E. (1997). Fitting the Task to the Human. Fifth Edition. London, UK: Taylor & Francis.

Marics. M.A., Engelbeck, G. (1997). Designing Voice Menu Applications for Telephones. *In Helander, M.G., Landauer, T.K., Prabhu, P.V. (Eds.) Handbook of Human-Computer Interaction, 2nd Edition*. North-Holland ELsevier.

Norman, K. L. (1991). The Psychology of Menu Selection: Designing Cognitive Control of the Human/Computer Interface. Ablex, Norwood, NJ.

Resnick P. and Virzi, R.A. (1995). Relief from the Audio Interface Blues: Expanding the Spectrum of Menu, List, and Form Styles. *Center for Coordination Science Technical Report*, #184, January 1995. Cambdirge, MA: MIT CCS

Resnick, P. and Virzi, R. A. (1992). Skip and Scan: Cleaning Up Telephone Interfaces. *In Proceeding from CHI '92: Conference on Human Factors in Computing Systems*, Monterey, CA, 419-426.

Schmandt, C. (1994). Voice Communication with Computers. Van Nostrand Rheinhold, New York.

Schumacher, R. M., Hardzinski, M., and Schwartz, A. (1995). Increasing usability of interactive voice response systems: Research and guidelines for phone-based interaces. *Human Factors*, 37 (2), pp. 251-264.

Schwartz, A. L., and Hardzinski, M. L. (1993). Ameritech phone-based user interface standards and design guidelines. [On-line]. Available: http://www.ameritech.com/corporate/testtown/library/standard/std-pbix.html

Schwartz, A. L., and Schwab, E. C. (1993). Improving menu design for the Rapid Order audiotext system using cluster analysis. *In Proceedings of the Human Factors Society 37th Annual Meeting*.