



The wizard of OZ: instilling a resilient heart into self-service business applications

Title	The wizard of OZ: instilling a resilient heart into self-service business applications
Author(s)	Costello, Gabriel
Publication Date	2006
Publisher	Springer



THE WIZARD OF OZ: Instilling a Resilient Heart into Self-Service Business Applications

Gabriel J. Costello

Galway-Mayo Institute of Technology
Galway, Ireland

Abstract *Speech enabled business applications are characterized by complex implementations that bring together language processing technologies, applications development, and end-user psychology. Resilience is critical to maintaining business to customer relationships when implementing these self-service solutions. The Wizard-of-Oz experiment is a valuable technique for simulating and building human-machine prototypes to ensure successful deployment of the completed service. This paper proposes the simplification and diffusion of the methodology to facilitate the growth in demand for automated e-business transactions.*

Keywords Wizard-of-Oz (WOZ) experiments, self-service technology (SST), automatic speech recognition (ASR)

1 INTRODUCTION

The growth of self-service technology (SST) in business to customer (B2C) transactions is being driven by the diffusion of information and communications technology (ICT) and the demand to move from high-cost manual transactions to low-cost automated self-service in enterprises and the public service. The Gartner Group have forecast that 70 percent of customer service contacts for information and remote transactions will be automated by the end of 2005 with an associated increase in investment in Web SST (Pujari 2004). These services are becoming increasingly critical for enterprises challenged with providing e-commerce solutions and building relationships in a world where customer and vendor do not meet face-to-face (Singh 2002). Among STT interfaces, the use of speech is regarded as ideal because it is the most “natural, flexible, efficient and economic form of human-machine communication”

(Koumpis 1998). However, creating conversational automated agents with responsibility for service levels and maintaining customer relationships is a complex challenge. Providing speech-enabled services requires capability in speech communication technologies, applications programming, and professional services developed in the environment of customer psychology and culture. Consequently, implementation of such solutions brings together many features crucial for resilience—cognitive, emotional, relational, and structural—as outlined in the conference theme. This paper takes a practitioner’s perspective based on the author’s experience of product lifecycle management of interactive voice response (IVR) and speech technologies in the e-business product portfolio of a telecommunications company. During this period, there was an industry realization that the main challenge was not the maturity of the underlying speech and language processing technologies but the development and deployment of robust customer solutions.

The first section of the paper presented an overview of the self-service business environment and introduced one innovation in the area: speech-enabled automated systems. The paper now proceeds as follows. Section 2 provides a technical description of speech and language processing, treated as a primary innovation in the area of SST. The subsequent section describes the testing of speech-enabled systems that are critical to the deployment of resilient applications. In particular, the focus is on a test methodology called Wizard-of-Oz experiments, viewed as a secondary innovation. Section 4 reviews the diffusion of the technique in areas outside of speech-enabled B2C solutions. The discussion in section 5 argues that the growth of business critical SST calls for the wider dissemination of knowledge and expertise in WOZ experiments and the need for research to address limitations with the technique. Finally, suggestions for future work and the conclusions are presented.

2 SPEECH AND LANGUAGE PROCESSING

Speech communication brings together a number of fields including language processing, computational linguistics, psycholinguistics, voice technologies, grammar checking, and information retrieval (Jurafsky and Martin 2000). Development of applications involve dialog design, integration testing, data collection, tuning, and performance analysis, which results in expensive implementations. Gartner’s researchers have included speech recognition as one of the top 10 technologies that will have the biggest impact on enterprises in the period 2002 to 2007. Speech solution providers are citing high profile implementations including a UK bank that is handling millions of calls per week and a U.S. healthcare service with two million customers that has patients and physicians using its speech-enabled appointments facility (Nortel 2005). However, it is worth noting that speech recognition has remained on the Gartner “Emerging Technologies Hype Cycle” for the last 10 years (Gartner 2005). Speech technologies include speech recognition, text-to-speech (TTS) synthesis, and speaker verification (authentication), a branch of biometrics. Recognition has evolved from initial basic discrete (isolated word) recognition to large vocabulary continuous speech recognition (LVCSR) and natural language understanding (NLU). These systems can be speaker dependent, as in the case of dictation products, or speaker independent, in the case of customer service applications with a large number of callers (Childers 2000).

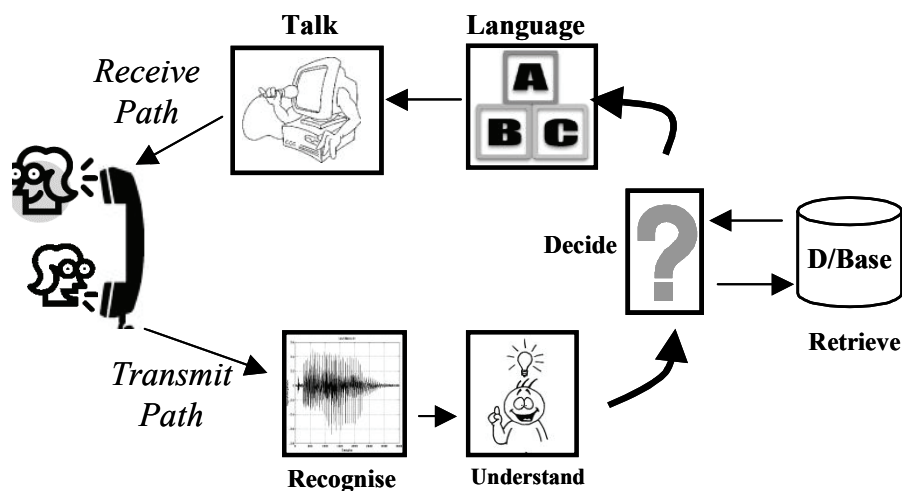


Figure 1. A Generalized Typology of a Speech-Enabled System

The move to provide speech services via the Web is being facilitated by VXML (Voice eXtensible Mark-up Language), a standard of the World Wide Web Consortium (W3C 2005), which allows conventional telephony interfaces such as IVR to evolve using voice gateways linked to a Web server.

Figure 1 shows the main components of a typical conversational system (Zue and Glass 2000) used for B2C information retrieval. First, the customer's spoken input is transmitted to a speech recognizer and the signal is processed. In the next step, the system produces a meaningful (semantic) representation of the signal. Then the system decides what information to retrieve from the database. This decision-making process may take a number of iterations if the dialogue manager seeks clarification from the customer when the input is deemed inadequate or ambiguous. The receive path generally delivers the required information to the customer in a spoken format using language generation and speech synthesis.

3 WIZARD-OF-OZ EXPERIMENTS

Three design principles have been proposed for designing dialogue systems and can be summarized as follows (Gould and Lewis [1985] in Jurafsky and Martin 2000):

- early focus on user and task
- simulations and prototypes
- iteratively test them on the user and fix the problems

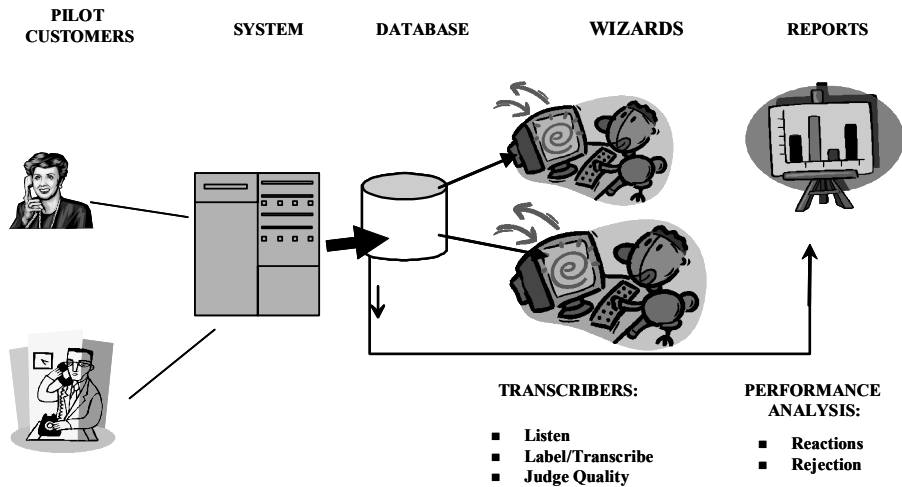


Figure 2. Typical Architecture of a WOZ Speech Application Experiment

A number of test methods are used to tune a speech-enabled system, including usability testing at various stages of the development process, focus groups, and piloting the service in a population of “friendly” users. One technique developed to simulate human-computer dialogue systems is called a Wizard-of-Oz (WOZ) experiment, where a hidden human operator replaces the automated agent in order to experimentally investigate the usability of the system before being deployed in the field (McInnes et al. 1997). The Wizard-of-Oz is also known as a PNAMBIC (Pay No Attention to the Man Behind the Curtain) system, where the human operator (the wizard) is disguised behind some interface software and the caller thinks that they are interacting with an automated application. The concept originated from the 1899 book by Frank Baum and the 1939 MGM film where Dorothy, the Tin Man, the Lion, and the Scarecrow follow the yellow brick road to Oz in order to make their requests to the awesome Wizard. In the end, the awe-inspiring ruler of Oz turns out to be just a simulation controlled by a very ordinary human (Biberman et al. 1999). In WOZ experiments, the function of the Wizard is played by a human but the user believes it is a computer. Figure 2 shows the typical architecture of a WOZ experiment designed to simulate a speech recognition application.

The technique can provide data on the interaction, acoustic performance, language model, and semantics. It has the advantage of being able to test the proposed solution at an early stage of the development cycle as only the interface software and the databases are required. Other advantages include rapid iterations and the ability to compare a number of design solutions. However, as it is very difficult to cover all of the errors, limitations and constraints of the live application, the conclusions from WOZ experiments can be rather idealised and even provide false positive results (Jurafsky and Martin 2000). In many cases, WOZ experiments are outsourced to specialist consultants.

4 DIFFUSION OF WOZ

In the previous sections, this paper has described the relationship between a primary innovation (speech-enabled self-service) and a secondary innovation (WOZ experiments) during the design and testing of B2C applications. However, there have been a number of research projects aimed at applying the methodology to other areas. Yang et al. (2000) applied WOZ to the development of a learning interface agent in order to make the system more natural, intelligent, and even emotional. In another context, a WOZ experiment has been extended to involve a robotic interface that is capable of simulating a number of different social behaviours (de Ruyter et al. 2005). The procedure has shown that the development of an artificial intelligence interface to provide “active help” has benefitted system users (Davis 1998). Wider applications of the method include collection of empirical data on mathematics tutorials in German (Benzmüller et al. 2003). One of the most novel implementations was the Neimo project, which extended the WOZ technique to the study of multimodal systems and provides evidence that the methodology can be broadened to designing, building, and evaluating services that allow the use of combined input media (Salber and Coutaz 1993). Another area of research has focused on simplifying both the data collection function and the model building that presently makes implementation of the methodology time consuming and expensive (Munteanu and Boldea 2000). A system called DiaWoZ has been designed to collect data in the complex domain of tutorial dialogues between university students and a mathematical tutoring system. The architecture places emphasis on modularity and clear interface requirements that allow for the progressive refinement of consecutive WOZ experiments (Fiedler and Gabsdil 2002). However, in spite of the implementation of the above applications, the evidence from the literature is that the technique is still very much in the realm of the innovator.

5 DISCUSSION

Business analysts are predicting the continued rapid growth in automated ICT applications and deployment of SST. These services bring together leading-edge technologies, the psychology of human-machine interaction, business processes, and the management of customer contacts. The resilience of the solution is critical in an environment where customer relationship management is being entrusted to a computer application and network. The focus on end-user driven development is posing questions on how to overcome traditional barriers between the user and the developer (Pettersson 2003). All of these challenges are presently being encountered in the deployment of speech-enabled business applications, which are currently limited to expensive implementations in verticals such as financial services, telecommunications and healthcare. Wizard-of-Oz experiments provide an effective early prototype environment to test interactions between humans and voice applications. The methodology requires a multi-disciplinary approach, bringing together psychology, culture and technology as well as fields such as synthetic computer characters that attempt to make humanoid agents lifelike (Thorisson and Cassell 1996). However, the technique needs to be diffused to a larger community especially through education, currently only being addressed on a

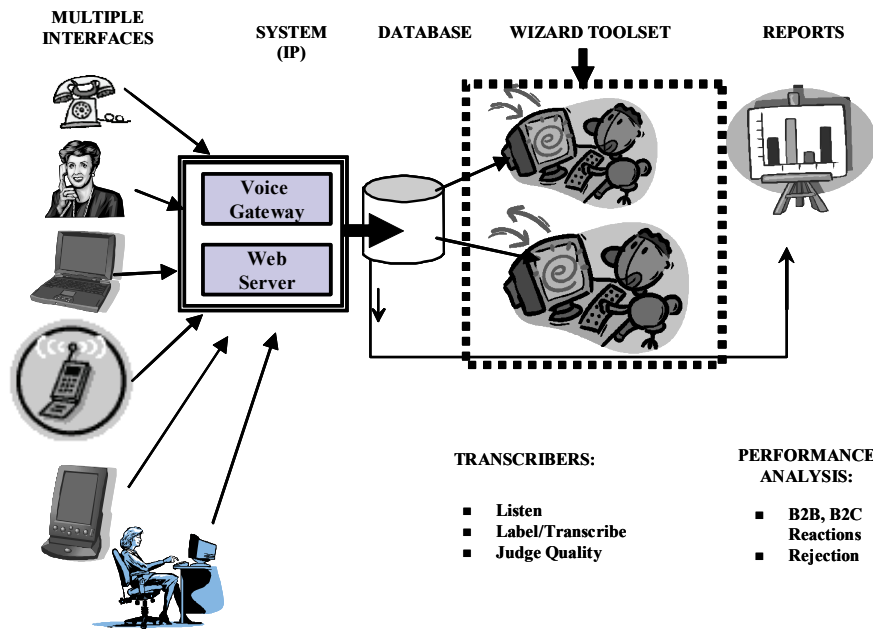


Figure 3. Future Work on Multiple Input WOZ Toolset

small scale, for example, by Henry Lieberman in the MIT Media Lab (2005). Future cross-disciplinary research is required to simplify WOZ experiments and address the weaknesses outlined above to enable the diffusion of the methodology to the wider area of SST and automated Web services. Challenges include ethical considerations of using people in usability testing, language and cultural localization, cost-effective implementation, and accessibility. Figure 3 presents the concept of multiple interfaces to a WOZ environment that is Internet protocol enabled and supported by a flexible development toolset—a “wizard” for wizards. The proposed topology also includes the generation of statistical data for performance analysis of the design.

6 CONCLUSIONS

This paper has provided a practitioner’s perspective on the requirement for resilient self-service speech-enabled applications. The business factors driving the growth in automated e-business solutions were described. An overview was given of the complexity of speech and language processing and the skills required to deploy voice-enabled B2C services while managing customer relationships. The important part played by Wizard-of-Oz experiments for the early simulation and testing of such systems was illustrated. Examples were provided of research into the dissemination of the methodology beyond the present, expensive, customer contact solutions. It was argued that the increasing demand for resilient automated e-business and the associated

capability to integrate end-user psychology with technology calls for the wider diffusion of WOZ techniques. Future work was then proposed to simplify and automate the method to bring it to a wider audience and address concerns with the validity and reliability of results. B2C self-service applications of the future will, like the Tin Man, need to have a heart.

References

- Benzmüller, C., Fiedler, A., Gabsdil, M., Horacek, H., Kruijff-Korbayová, I., Pinkal, M., Siekmann, J., Tsovaltzi, D., Vo, B. Q., and Wolska, M. "A Wizard-of-Oz Experiment for Tutorial Dialogues in Mathematics," in V. Aleven, U. Hoppe, J. Kay, R. Mizoguchi, H. Pain, F. Verdejo, and K. Yacef (eds.), *Proceedings of the AI in Education (AIED) Workshop on Advanced Technologies for Mathematics Education*, Sydney, Australia, 2003, pp. 471-481.
- Biberman, J., Whitty, M., and Robins, L. "Lessons from Oz: Balance and Wholeness in Organizations." *Journal of Organizational Change* (12:3), 1999, pp. 243-254.
- Childers, D. G. *Speech Processing and Synthesis Toolboxes*, New York: John Wiley & Sons, 2000.
- Davis, J. S. "Active Help Found Beneficial in Wizard of Oz Study," *Information and Software Technology* (40), 1998, pp. 93-103.
- deRuyter, B., Saini, P., Markopoulos, P., and Breemen, A. v. "Assessing the Effects of Building Social Intelligence in a Robotic Interface for the Home," *Interacting with Computers* (17:5), 2005, pp. 522-541.
- Fiedler, A., and Gabsdil, M. "Supporting Progressive Refinements of Wizard-of-Oz Experiments," in C. P. Rose and V. Aleven (eds.), *Proceedings of the ITS2002 — Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain, 2002, pp. 62-69.
- Gartner. "Gartner Highlights Key Emerging Technologies in 2005 Hype Cycle," Group, Boston, 2005 (available online through www.gartner.com).
- Gould, J. D., and Lewis, C. "Designing for Usability: Key Principles and What Designers Think," *Communications of the ACM* (28:3), March 1985, pp. 300-311.
- Jurafsky, D., and Martin, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Upper Saddle River, NJ: Prentice-Hall, Inc., 2000.
- Koumpis, K. "Corporate Technological Positioning in Automatic Speech Recognition and Natural Language Processing," Master of Science Dissertation, Science and Technology Policy Research (SPRU), University of Sussex, 1998.
- Lieberman, H. "The 'Wizard of Oz' Agent Experiment," unpublished paper, 2005 (available online at <http://www.media.mit.edu/~lieber/Teaching/Collaboration/Turvy/Turvy.html>).
- McInnes, F. E., Jack, M. A., Carraro, F., and Foster, J. C. "User Responses to Prompt Wording Styles in an Automated Banking Service with a Wizard of Oz Simulation of Word-Spotting," *IEE Colloquium on Advances in Interactive Response Technologies for Telecommunications Services* (Digest No: 1997/147), June 12, 1997, pp. 7/1-7/6.
- Munteanu, C., and Boldea, M. "MDWOZ: A Wizard of Oz Environment for Dialog Systems Development," in *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, Greece, 2000, pp. 1579-1582.
- Nortel. "Position Paper: The Power of Speech," 2005 (available online through www.nortel.com).
- Pettersson, J. S. "Prototyping Interactivity before Programming," in *Proceedings of the Workshop on End User Development, Conference on Human Factors in Computing Systems* (CHI 2003), Fort Lauderdale, FL, April 6, 2003, pp. 73-75.

- Pujari, D. "Self-Service with a Smile? Self-Service Technology (SST) Encounters among Canadian Business-to-Business," *International Journal of Service Industry Management* (15:2), 2004, pp. 202-209.
- Salber, D., and Coutaz, J. "Applying the Wizard of Oz Technique to the Study of Multimodal Systems," in L. J. Bass, J. Gornostaev, and C. Unger (eds.), *Human-Computer Interaction Selected Papers*, Berlin: Springer-Verlag, 1993, pp. 219-230.
- Singh, M. "E-Services and Their Role in B2C E-Commerce." *Managing Service Quality* (12:6), 2002, pp. 434-446.
- Thorisson, K., and Cassell, J. "Why Put an Agent in a Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue" (Abstract) in *Proceedings of Lifeline Computer Characters '96*, Snowbird, UT, October 5-9, 1996, pp. 44-45.
- W3C. World Wide Web Consortium (W3C): Voice Extensible Markup Language (VoiceXML) Version 2.0, 2005 (available online through <http://www.w3.org/>).
- Yang, Y., Okamoto, M., and Ishida, T. "Applying Wizard of Oz Method to Learning Interface Agent," *IEICE Workshop on Software Agent and its Applications (SAA2000) and Special Issue on Software Agent and its Applications Transactions of IEICE*, 2000, pp. 223-230.
- Zue, V. W., and Glass, J. R. "Conversational Interfaces: Advances and Challenges," in *Proceedings of the IEEE* (88:8), Special Issue on Spoken Language Processing, 2000, pp. 1166-1180.

About the Author

Gabriel J. Costello is a lecturer in Engineering at the Galway-Mayo Institute of Technology, Galway, Ireland. Prior to this he worked in the telecommunications industry with Nortel where he held new product introduction and product line management positions in the IVR, Self-Service and Customer Contact Center business units. Gabriel can be reached by e-mail at gabrielj.costello@gmit.ie.