~Mark Lucente

CONVERSATIONAL INTERFACES *for* E-COMMERCE APPLICATIONS

These e-commerce shopping agents are knowledgeable, patient, and affable during their conversational interaction with shoppers—while helping generate sales for site owners.

CONSUMERS SPEND BILLIONS ONLINE, THOUGH MOST ARE UNHAPPY WITH their shopping experiences [1]. My colleagues and I at Soliloquy, Inc., are creating conversational natural language interfaces for e-commerce applications that appear on Web sites as sales "Experts" conversing with shoppers, answering their questions, and helping find products they can buy. Each Soliloquy Expert is a software agent custom-developed for a particular e-commerce application. The metaphor reflected in an Expert-based e-commerce experience is natural conversation, as if the shopper were talking to a human salesperson. The shopper and the Expert co-produce knowledge and understanding through the conversation—a shopping experience more effective, fun, and human than its conventional counterpart of point-click-download.

By constraining an Expert's knowledge domain to a particular product line, say, laptop PCs, restaurants, or mutual funds, they can be made robust and helpful. Domain constraint facilitates several features of the personal shopping experience:

- Rich contextual understanding;
- Robust speech recognition;
- High-quality text-to-speech synthesis;
- Effective Web site navigation; and
- A business model based on spoken interfaces.

Amazon/VB Assets
Exhibit 1011
Page 1

The first two features benefit from the disambiguation, or derivation of precise contextual meaning, made possible by the limited size of a particular domain—its concepts and concomitant words—that is, of course, significantly smaller than all the concepts and all the words of any particular human language. For the company behind the Web site, profitability is emphasized; for spoken interfaces to flourish and ultimately meet popular expectations, they must support a "killer app" that is both compelling and profitable—in this case, conversational e-commerce.

Each Expert combines several technologies suitable for mass-market e-commerce applications: speech recognition, speech synthesis, natural-language understanding, and the Web. Early use over the past two years has yielded many insights from this confluence of technologies while prompting several important questions about the evolving online shopping experience. Especially notable are the multimodal nature of the interaction and the natural human perception built into each Expert's "personality."

Experts appear to the shopper as a small conversation window on an e-commerce Web page (see Figure 1). A human-like Expert (which can include a cartoon face) appears eager to help the shopper understand what is for sale and find the most appropriate items to buy. An Expert begins the conversation by greeting and prompting the shopper for questions. The shopper speaks to the Expert or types into the window. The Expert responds in several ways:

- Spoken (if audio output is enabled);
- Text; or
- Hyperlinked multimedia, including images, such as those of products available for purchase.

An Expert's main goal in a conversation is to help a shopper find the right item. Its response aims to satisfy several subgoals:

- Answer the shopper's questions;
- Proffer items the shopper may wish to buy;
- Ask for clarification; and
- Prompt the shopper for information, enabling itself to refine the shopper's search.

Each Expert uses a proprietary natural language understanding system developed by Soliloquy built atop a structured knowledge base. Natural-language subsystems convert the incoming text (from the keyboard or from a speech recognizer) into conceptual representations, add these concepts to the conversational context, determine subgoals, and finally motivate actions, such as constructing an answer, searching a product database, navigating a Web page, or prompting the shopper for more information. A built-in conversation manager orchestrates the input and output while updating the context representation. The structured knowledge base in each Expert contains the following components:

- Functional representations of general concepts (relevant to e-commerce shopping) and their relationships to the words and phrases of a particular language (English for now);
- Functional representations of domain knowledge (such as the often-specialized language of PC technology);
- A semantic representation of the essential structure and function of domain products and what they are used for (such as a laptop computer with a certain amount of RAM and at least one hard disk or a laptop that doesn't weigh much, necessitated, perhaps, by the user's traveling schedule);
- A database listing the products for sale and their features (such as price, speed, and size) on a given e-commerce Web site.

The knowledge-engineering approach facilitates a rich representation of context, allowing the shopper-computer conversation to be natural, nonlinear, and efficient [1]. Our object-oriented approach to conceptual representation makes it possible to quickly scale up the knowledge base of an Expert and create additional Experts in a multitude of domains.

Inherently Multimodal

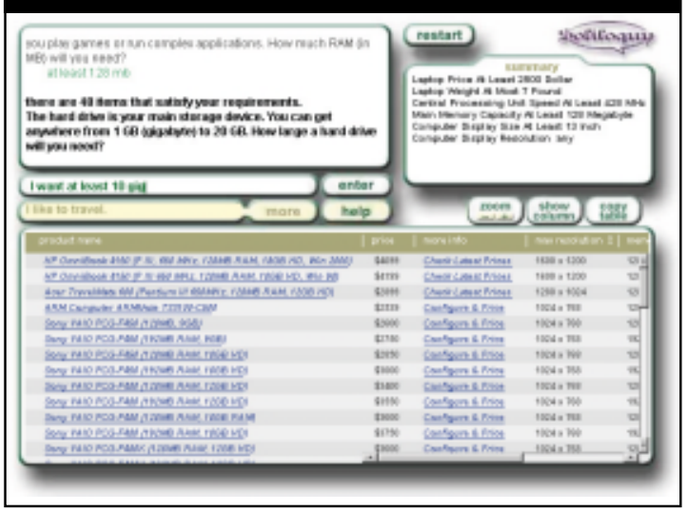
The interaction between a human user and a shopping Expert differs from human-human conversation in important ways that we have now begun to understand and exploit. Because each Expert is designed to appear on a Web page, the shopper-Expert interaction is inherently multimodal. The Expert senses the shopper's mouse and keyboard activity. The shopper sees images and hypertext generated on-the-fly. Moreover, each Expert can offer graphical interface elements (such as buttons or hypertext words) to augment the conversation, like a human-human conversation involving physical objects the participants might gesture toward or use iconically.

Multimodal interaction allows for rich conversation and contextual development often exceeding that of human-human speech-only conversation. Although each Expert's natural-language understanding is not, of course, as advanced as that of a typical human, the presentation of multimedia responses (such as pictures of the items for sale) facilitates complex conversation and delights users. For example, if a

Amazon/VB Assets
Exhibit 1011
Page 2

shopper tells the system he or she wants to spend less than $2,000, this constraint is illustrated in Figure 1 as a summary table, visible at a quick glance to the user, helping limit redundant or divergent exchanges. This simple on-screen table is not possible in a speech-only conversation, such as one over the telephone. Sensing user selections via mouse position further supports an Expert's multimodal interaction, which is often much faster than speech-only interactions. For example, when an Expert responds with

Figure 1. Expert-based conversation window on an e-commerce Web page.



several pictures of products matching a shopper's criteria, the shopper can point to one of them while saying, "How much does this one cost?"

Affable Personality

User testing has shown us that shoppers perceive Experts as having a personality [3]. Verbal output contributes to this personality in a way that is similar to human-human interaction (see Boyce's "Natural Spoken Dialogue Systems for Telephony Applications" in this section). (The prosody of the speech output contributes too, though I don't address this subject here; see Shneiderman's "The Limits of Speech Recognition" in this section.) Soliloquy's software developers give each Expert a personality suited to its specific shopping function, teaching it to use an adaptive blend of vernacular, humor, and down-to-business tone. The resulting personality is pleasant for most users to interact with while also projecting authority—rather like an affable college professor. Each Expert's personality adjusts dynamically and is designed to be adjustable by the e-commerce site owner.

Consider the Notebook Expert now operating on a notebook PC Web site that conversationally helps shoppers find the most appropriate machine for them to buy (see notebook-expert.com). Notebook computers are expensive and involve some measure of personal adaptation. Shopping for one is fraught with anxiety, indecision, soul-searching, doubt, and confusion over the technology's specialized jargon and features. The Expert's personality helps shoppers through this process, increasing the rate at which they become buyers, as found by Soliloquy's statistical dialogue mining analysis. Moreover, it exhibits patience by remaining attentive, no matter how long the conversation lasts.

This Expert is perceived to be honest and forthcoming, according to Soliloquy's research, as it shows all the information requested and answers questions quickly and precisely. It even adds a personal touch—remembering the names of individual shoppers, as well as their special requirements while engaging them on topics central to their needs.

All of this personality (and perceived intelligence) adds up to an improved shopping experience and increased sales, according to Soliloquy's dialogue mining of Experts in use. On the shopAcer.com notebook PC Web site, 30% of shoppers who conversed with the Expert went on to purchase, according to Soliloquy's dialogue mining, compared with less than 2% for shoppers who did not use the Expert.

Soliloquy's goal is to educate all its Experts, extending their personalities and intelligence. Educating while engineering represents a wonderful trend in software development, as it allows teachers, psychologists, linguists, consultants, and human experts from many fields to contribute to creating useful software that is a pleasure to engage. ■

References

1. Cullingford, R. *Natural Language Processing: A Knowledge Engineering Approach.* Rowman & Allanheld Computer Science, New York, 1986.
2. Manning, H. The right way to test ease-of-use. In *The Forrester Brief.* Forrester Research, Inc., Cambridge, Mass., 1999.
3. Nass, C. and Lee, K. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *Proceedings of CHI'00* (The Hague, The Netherlands, Apr. 1–6). ACM Press, New York, 2000.

Mark Lucente (lucente@soliloquy.com) is the chief technology officer of Soliloquy, Inc. in New York.

© 2000 ACM 0002-0782/00/0900 $5.00

Amazon/VB Assets
Exhibit 1011
Page 3