An Empirical Study on Voice-Enabled Web Applications

Studying pervasive computing's impact on consumer attitudes and the acceptance rate of new interfaces can help in the design and development of successful business applications employing voice-enabled Web systems.

Shuchih Ernest Chang
National Chung Hsing University

Michael S.H. Heng
Universitas 21 Global

In recent years, two great innovations have deeply influenced people's lifestyles—the Internet and mobile phones. The Internet offers a new way to communicate person to person and supports e-commerce. The mobile phone empowers people with its ease of use "anytime and anywhere." Moreover, mobile phones are emerging as a new medium that can connect to computers as well as the Web. The combination of these innovations has had a revolutionary effect on mobile commerce.[1] Because speech is the most basic and efficient form of communication, a voice interface for accessing the Internet by phone could further expand its utility and convenience.[2]

Self-service technologies have been used for years and are another area where voice interfaces have had an impact. Voice recognition technology was first applied in SST applications for hospital registration. Other sectors, such as the financial and airline industries, expanded their information services using voice recognition technology, before it finally expanded to other sectors. ACI-FIND (Advanced e-Commercial Institute-Focus on Internet News and Data), a research entity of the Institute for Information Industry,[3] reported that the market of voice recognition enjoyed a growth rate of 43 percent and the global total value would reach US$56 billion in 2006.

We built a self-service system integrated with multiple interfaces to evaluate pervasive computing. By identifying unmet user needs, the evaluation's results help suggest new applications. The technology acceptance model (TAM) has proved its worth in explaining and predicting user behavior in accepting or rejecting technical innovation.[4] We adapted TAM to evaluate users' perceptions of the proposed system.

A voice-enabled approach

The proposed voice-enabled Web system integrates multiple channels and provides users with choices for accessing information on the Internet. It gives users options to meet their needs, fit their preferences, or overcome environmental constraints. Although mobile phones have various options for accessing the Internet, small mobile devices generally have a hard time creating and displaying information. A typical mobile device's constraints—such as small screen size, slow speed, and inconvenient keyboard—make it cumbersome to access lengthy textual information.[5] However, a voice interface doesn't have these limitations. Recent speech recognition advances provide efficient voice capture and indexing mechanisms to make application systems with voice interfaces even more desirable. In addition, humans speak

Published by the IEEE CS and IEEE ComSoc ■ 1536-1268/06/$20.00 © 2006 IEEE

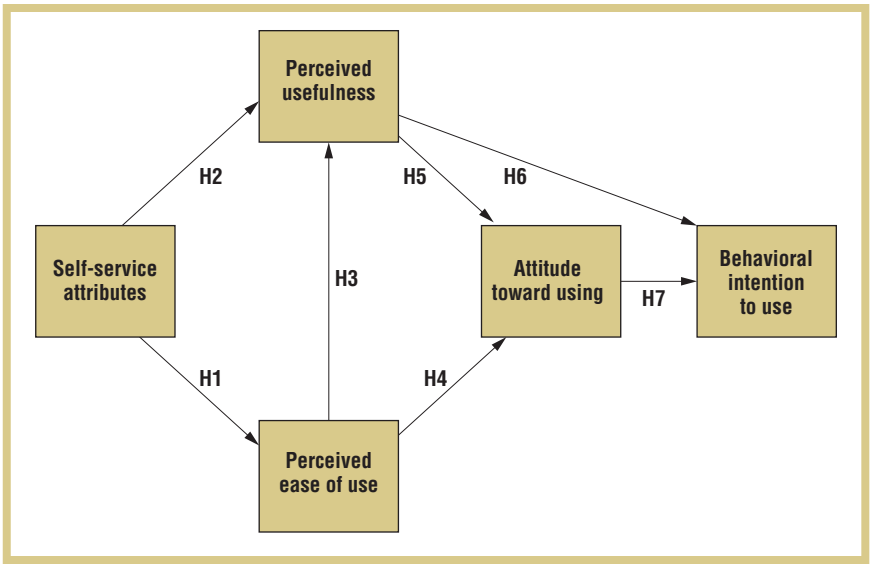


Figure 1. The research framework for evaluating the voice-enabled Web system.

faster than they type, and voice authoring can occur almost anywhere.

System architecture

We used a voice server to enable the creation of voice applications using industry standards XML, VoiceXML, and Java.[6] XML facilitates application integration and data sharing and enables the exchange of self-describing information elements between nodes. The proposed system will create and transform XML-based data into two different types of information. The first type includes information in various data formats that HTTP servers support, such as text, graphics, audio, and so on. The other information type is VoiceXML speech.

We set up the voice server between the phone and the Web server to interpret the VoiceXML documents. The server acts as a middleware processor. The VoiceXML interpreter, a key component of the voice server, contains the voice recognition and the synthesis engines used to automate the conversation between the site and the caller. Any Web site can be a VoiceXML content server. This system can give subscribers access to content offered by different Internet applications and services through wired or wireless PSTN (public switched telephone network) telephones.

Applications

We illustrate how the proposed system works with an application in a meal order service. In the US, the notion of eating quickly and conveniently appeals to people on the go. Queuing and crowding are two major reasons of considerable dissatisfaction.[7] In Taiwan, counter and online ordering are two popular methods of ordering fast food. We built a voice-enabled system prototype to provide meal order service to help users understand and evaluate the voice-enabled Web system.

The voice-enabled Web system has two user interfaces. One is a two dimensional (plane) Web browser interface, which presents photographs of food, text descriptions, and so on. The other is a one dimensional (linear) voice channel that doesn't include much information. It's extremely important in designing voice user interfaces to provide effective interactions and dialogues. We designed the voice interface following guidelines such as using simple words or brief Chinese characters to increase customers' recognition rate and satisfaction level.

Methodology

The technology acceptance model is a powerful way to explain users' acceptance of an information system or technology.[4] It postulates two beliefs, *perceived usefulness* and *perceived ease of use*, which determine an individual's attitude and behavioral intention. Perceived usefulness refers to "the prospective user's subjective probability that using a specific application system will increase his or her job performance within an organizational context." Perceived ease of use refers to "the degree to which the prospective user expects the target system to be free of effort." There are direct relationships among the attitude toward using a system (user preference regarding voice-enabled Web applications), the behavioral intention to use a system (users' willingness to use a voice-enabled Web system, even for other services offered in the future), and actual system use. In our research, we used TAM to design the research framework (see figure 1) for evaluating the proposed voice-enabled Web system's perceived usefulness and ease of use.

SST attributes

For our empirical study, we adapted TAM to incorporate SST attributes. To create hypotheses (see figure 2), we used the following eight characteristics to describe the SST attributes.

Convenience indicates being able to use any device anytime and anywhere. Some users might have a difficult work schedule or other factors that cause them to need a service on demand. The SST lets them perform the service from off-site locations such as from home, from the office, on the road, and so on.

Fun consists of two dimensions: curiosity and enjoyment. Curiosity is the extent that someone's interest is aroused during the interaction. Enjoyment is the extent to which someone finds the interaction intrinsically pleasurable or satisfying.

Accuracy consists of an accurate

- **H1**. Expected SST attributes have significant effects on the *perceived ease of use* of a voice-enabled Web system.
- **H2**. Expected SST attributes have significant effects on the *perceived usefulness* of a voice-enabled Web system.
- **H3**. *Perceived ease of use* has a significant effect on the *perceived usefulness* of a voice-enabled Web system.
- **H4**. *Perceived ease of use* has a significant effect on the *attitude toward using* a voice-enabled Web system.
- **H5**. *Perceived usefulness* has a significant effect on the *attitude toward using* a voice-enabled Web system.
- **H6**. *Perceived usefulness* has a significant effect on the *behavioral intention to use* a Web-based voice system.
- **H7**. *Attitude toward using* has a significant effect on the *behavioral intention to use* a voice-enabled Web application.

Figure 2. Operational hypotheses for this study.

outcome and process. An accurate outcome is the extent to which the system is error-free, and an accurate process is the extent to which the system is delivering and processing without failures. The system's reliability and accuracy are important factors in this, and apparent process failures might cause significant complications. Individuals might have concerns about arithmetic precision and implementation especially for services such as e-banking.

Transaction cost is the total expenditures incurred while using the voice-enabled Web system, including the cost to buy a device to access the system and the cost to complete a transaction. The SST could help users save money by finding a better deal.

Security has three dimensions: authentication, the assurance that senders are who they claim to be; confidentiality, protection against eavesdroppers understanding intercepted messages; and integrity, the assurance that the message hasn't been changed en route.

Speed consists of speed of delivery and speed of voice navigation. Speed of delivery refers to a client-side device's speed in accessing the system. Speed of voice navigation refers to the time required to request voice information or a transaction; it depends on how fast the voice server can speak to the users.

Expected *time* consists of waiting time (the time spent waiting for a system response) and saving time (how much time users can save by using the voice-enabled system compared to conventional ways). Users would expect to use a SST especially in a hurry or in crowded conditions. Waiting time is a strong obstacle to the use of onsite SST.

Control includes self-control and avoidance of service personnel. Self-control is the amount of control users have over the process or service outcome. Avoidance of service personnel means humans can interact with SSTs instead of service personnel. It was expected that the benefit of using SSTs was that users didn't have to interact with service personnel, and if consumers had a low need for interaction with personnel, they would resort to SSTs.

Sample and procedure

We collected empirical data by conducting a field survey using online questionnaires. Subjects were expected to have experience in using mobile phones and accessing the Internet. We developed the questionnaires by referencing previous articles in related fields, and measured each questionnaire item on a seven-point Likert scale, ranging from "strongly disagree" (extremely important) to "strongly agree" (not at all important). We conducted pretests on a small group to ensure understanding of the questions and valid measurements. From the pretest feedback and a subsequent discussion with experts, we modified and refined the questionnaire.

The voice-enabled Web application provided a virtual food service scenario with a phone number and a Web site address. After accessing the meal-ordering system via their cellular phones, we asked respondents to answer the questionnaire.

Analysis method

We used the statistical analysis tools SPSS 11.0 and AMOS 5.0 to analyze the samples. We used descriptive statistics methods to describe respondent characteristics, reliability analysis to ensure measurement consistency, and validity analysis to assess the measurement validity. Afterwards, we used factor analysis to find out the SST attributes' correlated variables and display relevant correlated components. Finally, we used structural equation modeling (SEM) to verify the theoretical model, which had latent variables.[8]

Results

After we gathered 249 responses, we identified invalid survey results by using techniques such as reverse questions, which were inconsistent with other questionnaire items. Overall, we collected and used 196 valid questionnaires for analysis. Among the 196 respondents, 45.9 percent were male and 54.1 percent were female. Educational levels ranged from undergraduate to postgraduate. For mobile phone and Internet experience, 78.1 percent responded that their most frequent mobile phone use was solely for communication. 51 percent reported monthly average expenses of NT$201 to NT$600 in mobile phone use. 31.1 percent responded that they most frequently used the Internet for relaxation and rest, 30.6 percent for email, and 20.4 percent for online entertainment. Fifty percent spent on average between half an hour and three hours on the Internet every day.

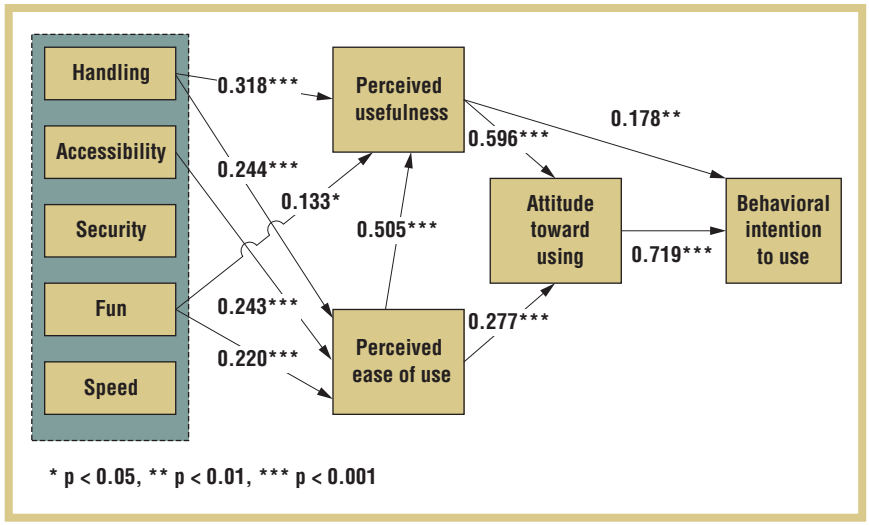


Figure 3. Results of the structural equation modeling analysis.

Data analysis and findings

Reliability analysis helps ensure the stability and consistency of measurements. Jum C. Nunnally suggested that 0.7 was the cut-off value for acceptable reliability coefficients.[9] Our test results, ranging from 0.8589 to 0.9091, for all constructs are good in this respect.

Construct validity shows that results are obtained from the use of the measurement and that the measurement doesn't violate our theories about the constructs.[10] In this study, we used factor analysis to measure construct validity and to determine the sets of correlated variables. We had some rules for measuring whether the data sufficed for factor analysis. First, the greater the KMO (Kaiser-Meyer-Olkin criteria) value, the more communal the factors are and the more suitable the data would be for factor analysis. A KMO value of 0.8 to 0.9 is meritorious,[11] so our study's KMO value of 0.815 is reasonable. The significance of Bartlett's test of sphericity means that the correlation among variables is suitable for factor analysis.

We used principal component analysis to extract factors. The communalities of each item ranged from 0.482 to 0.802. Our average communality of 0.662 is higher than the standard of 0.60 suggested by Henry Kaiser.[11]

We extracted five components using the rotation method of varimax with Kaiser normalization. The percentage of variance explained was 17.563, 15.781, 14.326, 11.785, and 11.031. The total percentage of variance explained was 70.486 percent. Karl Joreskog and Dag Sorbom suggested each component's eigenvalues should be higher than 1,[12] and all five components in this study achieved this value. After extracting the components, we individually named each one:

1. *Handling*. This component contained four variables: available anytime, waiting (for personnel) time, avoiding personnel, and accurate outcome. Anytime means users can freely choose when to use the system. Waiting time and avoiding personnel both indicate that users could control the process.
2. *Accessibility*. This component involved four variables: accurate outcome, accurate process, device cost, and device compatibility. Accuracy of outcome and process led to conformity. Being able to use any device, even an older device, complements peoples' lifestyles.
3. *Security*. It included three variables: authentication, confidentiality, and integrity.
4. *Fun*. It included two variables: curiosity and enjoyment.
5. *Speed*. It included two variables: speed of voice navigation and speed of delivery.

Cronbach's alpha for all extracted components were higher than 0.7 and had good reliability.

Nonsignificant correlations exist between Security and Fun, Security and attitude toward using, Security and behavioral intention to use, Speed and perceived usefulness, as well as Speed and attitude toward using.

SEM is a regression-based technology. It helps verify the research model on the basis of theories we have and tests the relationship between the cause and effect. Randall Schumacker and Richard Lomax indicated that a Chi-square (χ^2) with a nonsignificant p-value demonstrates a good fit between the data and the proposed measurement model.[8] Based on this criterion, our research model fits the data adequately.

Figure 3 presents the results of the research model with significant paths as straight lines and the standardized path coefficients between constructs. The Handling component has significant effects on perceived usefulness and perceived ease of use (β = 0.244 and 0.318, $p < 0.001$). Accessibility has significant effects only on perceived ease of use ($\beta = 0.243$, $p < 0.001$). Security has no significant direct effect on perceived ease of use and perceived usefulness ($p > 0.05$). Fun has significant effects on perceived ease of use ($\beta = 0.220$, $p < 0.001$) and perceived usefulness ($\beta = 0.133$, $p < 0.05$). As for Speed, it has no significant effect on perceived ease of use and perceived usefulness. Perceived ease of use has significant effects on perceived usefulness ($\beta = 0.505$, $p < 0.001$) and attitude toward using ($\beta = 0.277$, $p < 0.001$). Perceived usefulness has significant effects on attitude toward using ($\beta = 0.596$, $p < 0.001$) and behavioral intention to use ($\beta = 0.178$, $p < 0.01$). Attitude toward using has significant effects on behavioral intention to use ($\beta = 0.719$, $p < 0.001$).

H1 and H2 are not supported because not all SST attributes have direct effects on perceived ease of use and perceived usefulness. Among perceived ease of use, perceived usefulness, attitude toward using, and behavioral intention to use, our research results are consistent with the TAM findings proposed by Fred Davis.[4]

Discussion

By assessing the most important self-service characteristics extracted by factor analysis—Handling, Accessibility, Security, Fun, and Speed—we derived the following findings:

- When users handle voice-enabled Web systems at their pace and comfort, they perceive the system as friendly. An example is when they don't have a restriction on the time to use it.
- You can introduce this voice-enabled Web system into people's lives as long as they can access the system through popular devices. Assuming that users can get the desired outcome using the system, the system is more acceptable to users if they can use familiar, existing, and available devices to access the system. Users appreciate the ease of use this allows.
- Security is one of the most important factors. Nevertheless, Security isn't perceived as having an effect on ease of use and usefulness. Although some security alerts occur when surfing the Web, Security is still perceived as an invisible mechanism. Users can't see how it works, so it might be why they feel its Security is important without relating it to the system's ease of use and usefulness.
- Fun has significant effects on both ease of use and usefulness of the proposed voice-enabled Web system. This could be because Fun psychologically encourages users to approach and use the system. Additionally, if users enjoy using the system, it can increase their perception of its practicability.
- Speed is also an important factor that has no significant effect on ease of use and usefulness. Even though the users' preferences for Speed might differ from slow to fast, the system's speed can't be adjusted by using different channels, so Speed doesn't affect their perception of ease of use and usefulness significantly.

In summary, Handling was the greatest factor in using this voice-enabled Web system. Accessibility and Fun were other major factors in users' perception of the system. Security and Speed were also important factors but they didn't have significant effects on perceived ease of use and perceived usefulness.

Our findings suggest that there are positive directions that can be taken with perceived ease of use, perceived usefulness, attitude toward using, and behavioral intention to use. From the result, we found the strong effects of perceived ease of use on perceived usefulness, perceived usefulness on attitude toward using, and attitude toward using on behavioral intention to use.

Implications

In addition to our direct findings, we provide here some tactics and practical implications for leveraging the business opportunity offered by voice-enabled systems.

Applications of voice-enabled systems

According to the survey's result with respect to the food service application, more people have experience using mobile phones than using the Internet. Although a voice channel offers less information, it might make sense to replace an Internet channel with a voice channel (using mobile phones), especially in certain situations, such as in an emergency, when you're on the go, or when outdoors. So we suggest that you could initially use voice-enabled Web systems in such situations to increase the odds of success. Although users could browse the information on the Web first, they won't do that all the time. The system should probably take advantage of what people are familiar with by adapting its access channels to popular devices and by providing popular products and services. If users know about the content and the access procedures in advance, accessing the information will go more smoothly.

Perceptions of attributes of voice-enabled Web systems

With respect to the five component factors, we propose some ways to improve the user's attitude toward and acceptance of voice-enabled Web systems.

Improvement in the system's Handling could affect perceptions of ease of use and usefulness. However, the degree of expected Handling can greatly differ from how users actually respond. Future applications should pay more attention to facilitate system manipulation and operation for users.

Because the voice-enabled Web system has already provided multiple channels, users could choose how to access the system comfortably. Accuracy affects expected accessibility, and the system's stability also seems to play a significant role. Software and hardware affect the system's stability as well. While setting up a voice-enabled Web system, you should address the quality and compatibility of computer and phone I/O interfaces.

Because hackers emerge in an endless stream, Security is a big concern. For example, voice transmission requires

using encryption and decryption. The system should contain security mechanisms for authentication, confidentiality, and integrity. And, the system should inform users of the related security practices it implements. Clearly displayed security mechanisms and enforcement will increase users' trust in the system.

When arranging a voice-enabled Web application's structure, you should consider the level of fun in using the application, focusing on the content and navigation design. If users are having fun, they'll be more relaxed when experiencing a new technology-based tool.

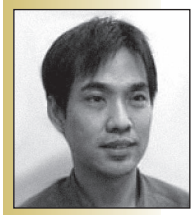
What speed will users consider reasonable and acceptable? Different users have different needs. Although speed doesn't significantly affect perceived ease of use and usefulness, you shouldn't overlook it during system design and development.

Limitations

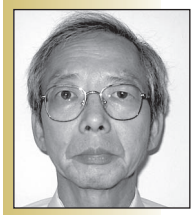
We conducted this study through an online survey, so the respondents had some computer experience. Because of this, we haven't heard from people without computer experience. For people who don't typically access computers, a voice-enabled system might provide their only access to the Web system. So, the acceptance rate of voice-enabled Web systems by inexperienced users might be higher, and the statistics result derived from our adapted TAM model might also differ.

We adapted field methods to a real-life environment to generalize the research. We couldn't control many variables, such as the surrounding noise, which could influence the perceived quality of speech. The system's recognition rate was beyond our research scope. If the system couldn't recognize a respondent's speech very well, the user would stop the conversation with the Web system. So, only those understood by the system completed the survey.

the AUTHORS



Shuchih Ernest Chang is an assistant professor at the Institute of Electronic Commerce, National Chung Hsing University, Taiwan. His research interests include Internet technologies, e-Commerce, enterprise application architecture, information security management, and voice-enabled Web systems. He received his PhD in electrical engineering from the University of Texas at Austin. He is a member of the IEEE. Contact him at eschang@dragon.nchu.edu.tw.



Michael S.H. Heng is a visiting professor at Fudan University, China, and an associate professor with the online university Universitas 21 Global. His research interests include e-business, globalization, IT strategy, and implementation. He received his PhD in information systems from the Vrije Universiteit Amsterdam. Contact him at michael.heng@u21global.com.

Due to space limitations, this article doesn't give our survey's full results. Please feel free to contact us for the data.

Many products and services could take advantage of voice-enabled Web systems. Applying a system to a different area might raise or extract different and important attributes regarding how users perceive the system. A future study might investigate individual traits to identify different user groups' attitudes and acceptance levels.

System hardware factors also affect users' perceptions of voice-enabled Web systems. Future studies could conduct laboratory experiments to manipulate variables, such as environment, device, recognition rate, and so on. P

REFERENCES

1. T. Teo and S. Pok, "Adoption of WAP-Enabled Mobile Phones among Internet Users," *Omega: The Int'l J. Management Science*, vol. 31, no. 6, 2003, pp. 483–498.
2. M. Lucente, "Conversational Interfaces for E-Commerce Applications," *Comm. ACM*, vol. 43, no. 9, 2000, pp. 59–61.
3. "ACI-FIND, Focus on Internet News and Data," *Inst. for Information Industry*, 2004; www.find.org.tw.
4. F. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, vol. 13, no. 3, 1989, pp. 319–340.
5. N. Anerousis and E. Panagos, "Making Voice Knowledge Pervasive," *IEEE Pervasive Computing*, vol. 1, no. 2, 2002, pp. 42–48.
6. J.A. Larson, "VoiceXML and the W3C Speech Interface Framework," *IEEE Multimedia*, vol. 10, no. 4, 2003, pp. 91–93.
7. I. Church and A. Newman, "Using Simulations in the Optimisation of Fast Food Service Delivery," *British Food J.*, vol. 102, nos. 5–6, 2000, pp. 398–403.
8. R. Schumacker and R. Lomax, *A Beginner's Guide to Structural Equation Modeling*, 2nd ed., Lawence Erlbaum Associates, 2004.
9. J. Nunnally, *Psychometric Theory*, 2nd ed., McGraw-Hill, 1978.
10. U. Sekaran, *Research Methods for Business: A Skill-Building Approach*, 4th ed., John Wiley & Sons, 2003.
11. H. Kaiser, "An Index of Factorial Simplicity," *Psychometrika*, vol. 39, no. 1, 1974, pp. 31–36.
12. K. Joreskog and D. Sorbom, *SPSS LISREL VII*, 2nd ed., McGraw-Hill, 1989.