

Modality Usage for Digit Entry in Telephone Speech User Interfaces

Josh Bers, Bernhard Suhm and Dan McCarthy

BBN Technologies

10 Moulton Street, Cambridge, MA 02138 USA

+1 617 873 3200

{jbers,bsuhm,dmccarthy}@bbn.com

ABSTRACT

Speech recognition is commonly thought to improve customer satisfaction, however, there is little evidence that it should replace touch-tones for all interactive voice response (IVR) applications. Initial results from a recent field trial indicate that callers prefer being given the option to say or key in their telephone numbers, and that the usage split is roughly 50/50 even when touch-tone is the more successful data entry mechanism. Our analysis indicates that the overall yield of correct telephone number entry is the same for the multi-modal speech/touch-tone data entry option as for touch-tone only baseline system, mainly because more callers respond to the multi-modal system. We also found that callers migrate to speech over time as the recognition performance increases, with tuning, and as they gain experience with the system.

Keywords

telephone voice user interfaces, speech recognition, multi-modal

INTRODUCTION

In deploying speech recognition in IVR systems, the challenge is to use modalities judiciously. Studies show that callers generally prefer speaking to keypad entry for some tasks [2]. While speech enables designers to depart radically from cumbersome, multi-layer, touch-tone menu systems, touch-tone input still needs to be considered, since some callers may prefer touch-tone input over speech on certain tasks [2] or under certain circumstances (e.g., in noisy conditions, to ensure privacy, or after several recognition errors). Some design guidelines have been introduced [1], but most have been anecdotal with few lab studies addressing the issues of multi-modal (speech and touch-tone) telephone user interfaces.

This paper presents results from a field trial of a speech-enabled IVR that offered both speech and touch-tone for the input of digits. We analyze performance and modality usage at both the beginning and end of the trial. The results indicate that offering the option of keypad entry in addition to speech increases the likelihood of success and allows users to gradually adapt to a new speech system.

THE FIELD STUDY

Method

The field trial was conducted at a call center of a large telecommunications provider. More than 20,000 calls were handled by the new speech-enabled IVR system during a period of eight weeks.

Callers were first presented with an open ended routing prompt that allowed them to describe their reason for calling in their own words. A study of natural language call routing appears elsewhere in these proceedings. The initial prompt was followed by a request for a telephone number to access account information, “Please say or key-in your area code and telephone number.” Whenever ten digits were not recognized (which could be due to malformed input, recognition errors, or no caller response), the caller was prompted again (*reprompt*). After two consecutive failures in entering the telephone number, or after interacting with the automated system, callers who did not hangup were transferred to a customer service representative (*agent*).

Data Capture

Data were collected in two ways: IVR logs and survey questionnaires. The speech-enabled IVR produced a log for every call. This log contained, for each call, the complete sequence of prompts that callers heard, until they were transferred to a live agent or they hung up in the IVR. For the analyses presented in this paper we examined the caller behavior at the telephone number prompts.

Data collected early in the trial was used to tune the speech recognition engine to maximize system performance before final evaluation during the last two weeks of the trial. For the final evaluation, we not only collected IVR logs, but callers were also asked by the agents, at the end of each call, to participate in a survey to assess their experience with the speech interface.

RESULTS

Objective Modality Usage

Table 1 shows the objective modality usage of callers at the initial telephone number prompt and the reprompt. The untuned data are based on a sample of 8737 callers at the first telephone number prompt, of which 2846 also experienced the reprompt; the sample sizes for the tuned system are 2876 and 908 respectively.

Table 1: Modality usage for telephone number entry

	Speech	Touch-Tone	No Response
<i>Initial Prompt</i>			
Untuned	38%	46%	16%
Tuned	45%	45%	10%
<i>Reprompt</i>			
Untuned	29%	43%	28%
Tuned	45%	39%	16%

Comparing the usage rates of the tuned with the untuned system shows that tuning significantly increased response rate, mostly due to an increase in spoken responses ($p < .01$ for each prompt).

We note that the significant increase in the usage of speech cannot be attributed solely to the tuning of the speech recognition system, some may be due to caller adaptation. We estimate that about 30-40% of the callers in the post-tuning phase had previously experienced the system (by observing repeated telephone numbers). Repeat callers may have grown more accustomed to speaking the number.

Roughly the same number of callers spoke as used the keypad at the initial prompt of the tuned system, with 10% not responding. Response rate at the reprompt is lower, but usage of speech remains steady at 45%. The rate of non-compliance, those who fail to respond at both prompts or hangup before responding, fell from 8.2% to 3.5% (of those hearing the 1st prompt) after tuning ($p < .01$).

Success Rates

Table 2 shows the success rates for entering a complete telephone number for callers requesting account information. Successful entry is determined by a match with a valid account in the database. Telephone number entry by touch-tone input is significantly more successful than by spoken input both before and after tuning ($p < .01$).

Table 2. Success rates for telephone number entry by input modality (sample size in parenthesis)

	Speech	Touch-tone	Used Both
Untuned	50% (307)	83% (438)	59% (49)
Tuned	62% (100)	83% (94)	47% (17)

As expected, tuning improved the success rate for the speech modality ($p < .05$). The success rates may appear low, but 35% of the spoken responses were malformed. Rates for touch-tone and for callers who used both speech and touch-tone (i.e., switched modality at the reprompt) did not change significantly ($p > .05$). Note that the modality switch indicates that the initial entry failed, which could have been either speech or touch-tone. After tuning, the

overall yield of telephone number matches equaled that of the touch-tone only baseline system ($p > .05$).

Caller Modality Preferences

Table 3 shows the responses of callers to the survey question. Surveyed callers heavily favored having the choice of both modalities, even if “no preference” is counted as a negative response ($p < .01$).

Table 3. Responses to caller survey

<i>“Did you like being given the choice to either say or key-in your telephone number?”</i>		
Yes	No	No preference
385 (78%)	31 (6%)	81 (16%)

CONCLUSIONS AND FUTURE WORK

Our field trial shows that callers strongly prefer being given the choice of speech or touch-tone input and that they continue to use both modalities, even after becoming familiar with speech. Usage of speech and touch-tone is about equal in a tuned system configuration. This study investigated only telephone number entry, but results presumably generalize to similar digit entry tasks.

Despite evidence for higher success for callers who fell back to touch-tone after an initial error with speech, only 1/3 did so in the tuned system. Future work should look at the effect of the reprompt wording on modality switching, e.g., “key-in or say the number” may increase touch-tone usage. Future work should also focus on separating the effects of tuning from those of caller adaptation on modality usage.

The fact that the overall yield for the multi-modal speech/touch-tone system is not higher than for the touch-tone only baseline suggests that other usability issues impose limitations: for example, access to and accuracy of the call center’s customer database, and the willingness and ability of customers to enter the requested information. Further improvements to the success rate for speech may be achieved through the use of confirmation dialogues and/or n-best recognition output. Our study showed that careful design requires consideration of both success rates as well as caller satisfaction.

Acknowledgements

The authors gratefully acknowledge the contributions of all members of the Call Director team at BBN Technologies.

REFERENCES:

- Balentine, B. and Morgan, D.P., *How to Build a Speech Recognition Application*, Enterprise Integration Group, San Ramon, CA, 1999.
- Basson, S., Springer, S., Fong, C., Leung, H., Man, E., Olson, M., Pitrelli, J., Singh, R., & Wong, S. User participation and compliance in speech automated telecommunications applications. *Proc. of the ICSLP* (Philadelphia, PA, 1996), 1680-1683.