

# USER PARTICIPATION AND COMPLIANCE IN SPEECH AUTOMATED TELECOMMUNICATIONS APPLICATIONS

Sara Basson, Stephen Springer, Cynthia Fong, Hong Leung, Ed Man,  
Michele Olson, John Pitrelli, Ranvir Singh, Suk Wong

NYNEX Science & Technology, Inc.  
500 Westchester Avenue  
White Plains, NY 10604, USA  
email: sara@nynexst.com

## ABSTRACT

This paper reviews results from a number of field trials assessing speech recognition feasibility for telecommunications services. Several applications incorporating speech automation are explored: Directory Assistance Call Completion (*DACC*), partial speech automation of Directory Assistance (*OSF* - Operator Store and Forward), banking over the telephone (*Money Talks*) and partial speech automation of a customer calling center (*PREVIU*). The experimental results presented here were collected through Wizard-of-Oz experiments as feasibility precursors to speech recognition automation. Speech interfaces were clearly superior to Touch-Tone in one experiment (*PREVIU*), with caller participation increasing by 30%. In another experiment (*Money Talks*), speech recognition interfaces did not improve caller participation and, in fact, provided no advantages over Touch-Tone automation. A number of prompting strategies have been identified as advantageous in increasing caller participation and compliance in automated services. Ultimately, success with speech automated services will rely on identifying the services most suitable for speech automation and then carefully crafting the user interface.

## 1. INTRODUCTION

Speech recognition technology is deployable for many applications, but constraints on users remain. The technology is often unable to intelligently handle out-of-domain requests, and also places limits on how and when users can speak. As a result, telephone users may not always view speech technology as superior to Touch-Tone menus. The burden falls on application developers to determine which applications are best suited for speech automation, which speech technologies are required, and how to design the service to maximize user participation and satisfaction. This paper will present interface recommendations which NYNEX Science & Technology has derived from multiple domains. It will also examine why dramatic differences emerge with respect to caller participation across different speech automated services. Several applications incorporating speech automation will be explored: Directory Assistance Call Completion (*DACC*) [1], partial speech automation of Directory Assistance (*OSF* - Operator Store and Forward)[2], banking over the telephone (*Money Talks*)[3] and partial speech automation of a customer calling center (*PREVIU*)[4].

For each of these services, the initial customer field trials were conducted as Wizard-of-Oz experiments. Callers were presented with what appeared to be a fully automated interface. In fact, human representatives or operators were listening at all times, and serving as speech recognizers according to a set of pre-defined criteria. This configuration facilitated rapid assessment of the feasibility of alternative speech automation strategies, and also determined optimal design features and speech recognition requirements prior to deployment.

## 2. FIELD TRIALS

The field trials referred to in this paper are described below. In all cases, callers were unsolicited participants calling the standard numbers for Directory Assistance, automated banking, or customer service.

### Directory Assistance Call Completion (*DACC*)

Directory Assistance Call Completion (*DACC*) offered callers the option to be automatically connected to the number requested in Directory Assistance calls. Callers were asked to say "yes" if they wanted to complete the call, for a surcharge of 35 cents, after they were given the telephone number they had requested. The *DACC* field trial collected approximately 19,000 calls over a three month period.

### Speech Automation for Directory Assistance (*OSF*)

In partial speech automation for Directory Assistance (*OSF*), callers were queried via digitized recordings to say first the city they were searching, and then the full name of the person or business. The caller's speech was then compressed and background noises and silences were deleted before passing the utterances to an operator. Our technical trial processed 174,000 calls through the *OSF* platform.

### Banking over the Telephone (*Money Talks*)

The *Money Talks* speech automated banking trial was superimposed over a bank's current Touch-Tone automated voice response system. The caller success rate for the bank's Touch-Tone service is approximately 70%. The trial was designed to determine whether customer participation could be improved, and to identify which range of speech technology was most suitable for this service. Several automation scenarios were explored, ranging from relatively crude digit-only recognition to more sophisticated configurations, incorporating continuous speech recognition and natural language. In addition, a range of prompting configurations were explored, such as comparing customer responses to terse vs. complete prompts. Over the four month trial, 48,000 callers to the automated Touch-Tone system were routed instead to what appeared to be a speech automated position.

### Automated Customer Service Center (*PREVIU*)

The *PREVIU* Wizard-of-Oz trial assessed speech automation potential in a customer service center. This service had recently

been automated via Touch-Tone, but suffered from a disappointing caller participation rate. An interface was designed to allow callers to say, rather than key in, their routing choices. Callers were presented with queries such as: "Please select one of the following: sales, service, or billing information." The menus were typically two layers deep before callers were passed to the appropriate representative or to another automated service. The *PREVIU* trial sampled 9000 calls over a period of three weeks.

### 3. IDENTIFYING SUITABLE SERVICES FOR SPEECH AUTOMATION

Speech recognition technology typically performs better for smaller vocabularies and constrained customer responses than for larger vocabularies and less predictable, more verbose responses. This tends to drive speech technology application developers towards selecting particular services for automation. These services, however, may not improve customer participation or customer satisfaction over standard Touch-Tone interfaces.

Telecommunications services can be automated via either speech technology or Touch-Tone. Speech technology is preferable in several instances, such as areas of low Touch-Tone penetration, or if there is predetermined customer aversion to using the Touch-Tone keypad to access service. Speech recognition technology is necessary to automate services where the number of service options is large. For example, a restaurant selector service that asks callers which cuisine they would like would be manageable as a speech automated service ("What kind of cuisine would you like?") but unwieldy as a Touch-Tone service ("For Chinese food, press 11; for Italian food, press 12...")

Sections 3.1 and 3.2 compare *Money Talks* and *PREVIU* in terms of their viability as speech automated services, since Touch-Tone automation data are available for both of these services.

#### 3.1. Results: *Money Talks*

The *Money Talks* system offered callers the opportunity to say their banking choices rather than keying them in on their Touch-Tone keypads. The bank involved in the study has provided an automated Touch-Tone customer service center since 1989. Approximately 70% of the bank's customers use the Touch-Tone system; and the remaining callers typically default to customer service representatives. The goal of the *Money Talks* trial was to evaluate the effect of a speech recognition front end on caller participation and caller satisfaction.

The *Money Talks* trial evaluated caller participation rates as a function of speech technology sophistication. Some customers were presented with a scenario that required them to say digits only, as in "For account information, say one; for funds transfer, say two..." Other callers were presented with a keyword interface, with prompts such as: "Please say one of the following: checking, savings, money market..." In the most sophisticated technology modeled, callers were allowed to respond using the syntax and semantics of their choice, with open-ended prompts such as "You can get account information, or transfer funds between accounts. What would you like to do?"

Results of the *Money Talks* trial indicated that fewer callers successfully traversed the speech automated system as compared to the Touch-Tone version. On average, 55% of the callers engaged the system, 16.8% abandoned the call, and 28.2% transferred to customer service representatives.<sup>1</sup>

<sup>1</sup> Customers transferred to representatives when they explicitly asked for an operator or when they repeatedly failed to comply with instructions on how to speak.

The most suitable speech recognition technology allowed callers to say embedded keywords, such as "I'd like *checking*." The discrepancies across the technology levels, however, were surprisingly small: Customers performed nearly as well with digit-only recognition as they did with the more sophisticated speech/language scenarios. (Table 1)

Overall, 85.5% of callers who engaged the system at all managed to complete at least one service. In *Money Talks*, speech interfaces were grouped into five broad "scenarios" for recognition: (1) isolated digit; (2) digit-spotting; (3) isolated keyword; (4) keyword spotting; and (5) continuous speech recognition with some natural language (NL). Results for the five scenarios are shown in Table 1.

	Scen 1	Scen 2	Scen 3	Scen 4	Scen 5
callers initiating a service	56.4	55.9	52.8	54.3	48.1
success (of callers initiating service)	84.4	84.8	85.6	88.8	83.8
calls with no re-prompts	63.3	62.4	46.2	57.3	45.4

**Table 1:** Measures of success, in percentages of all calls within each scenario.

The best prompting styles were terse, but clearly enumerated the service choices. Overall, however, speech automation did not increase caller participation over the current Touch-Tone interface, with a high percentage of callers abandoning or defaulting to a representative. Clearly, the limitation or rigidities of a particular speech technology were not the critical features accounting for these results, since caller participation was no higher with the more sophisticated speech recognition scenarios.

#### 3.2. Results: *PREVIU*

*PREVIU* presented callers with a speech automated front end to a customer service center. The speech technology modeled in *PREVIU* was keyword spotting (Scenario 4 in the *Money Talks* trial). In *PREVIU*, however, speech automation dramatically increased caller participation.

We have found advantages in offering callers the option of using Touch-Tone or speech as their automation medium. The menu offerings in the customer calling bureau were well-suited to Touch-Tone, with prompts such as "For billing, press 1...for orders, press 2..." Nonetheless, approximately 50% of the customers simply did not respond to the Touch-Tone queries. During the *PREVIU* trial, a subset of callers was asked as the first query to "press 1" if they opted to use the Touch-Tone service. Caller success rates for this version, as well as the version that offered speech automation only, appear in Table 2.

Prompts offered	Speech:% response	T-T:% response	Total Success
speech or T-T	31.4	52.7	84.1
speech only	78.8	NA	78.8

**Table 2:** Caller participation rates for speech and/or Touch-Tone

The customer compliance rates were highest (84.1%) for the condition where callers themselves selected Touch-Tone or speech as the mode of interaction. When callers were only given the option

to speak their choices, 78.8% of the callers participated. This exceeds both Touch-Tone participation for this service (50%) as well as speech participation for *Money Talks* (55%).

### 3.3. Money Talks vs. PREVIU

The large differences found between these two field trials warrant further scrutiny.<sup>2</sup>

Several characteristics account for the low success rates of speech technology with *Money Talks*. Callers were already familiar and successful with an automated Touch-Tone system. The speech activated system remained novel and unfamiliar for the duration of the field trial. As a result, callers who typically “key through” prompts on their Touch-Tone keypads were now obliged to listen. The interaction for these callers inevitably took longer than they were accustomed to. Customers exposed to the “digit only” scenarios were, ironically, in a better position to “barge in”, since they were told to say a digit that corresponded to the familiar Touch-Tone keypresses. Also, the services offered in the *Money Talks* trial were designed to be manageable with a Touch-Tone interface. As such, no more than five choices were presented at any one menu. The mnemonic advantages of speech over Touch-Tone are not so strong under these circumstances. Finally, the acceptance rate for the automated Touch-Tone interface (70%) is already relatively high for an automated service. The remaining 30% non-participants may have had queries that required human intervention, or they may have been strongly averse to automation of any kind. The results suggest, however, that speech automation may not be well-suited to capture the small number of rejectors in an already successful Touch-Tone service.

In contrast, customers calling the *PREVIU* system were not already trained and familiar with the Touch-Tone interface. The low participation rates with the Touch-Tone interface suggest that callers are not enthusiastic about using Touch-Tone, at least for this service. The circumstances that are most conducive for speech recognition automation are cases where Touch-Tone is not in place, not suitable, or not extremely popular.

## 4. EFFECT OF PROMPTS

Speech recognition interfaces require callers to speak in particular ways in order to be understood. Ideally, callers will say words that are in the recognizer’s vocabulary, and not embed the target words in lengthy and extraneous speech. Several field trials conducted at NYNEX have explored the effect of pre-recorded prompts on customers’ responses. Some of the findings about successful prompting are specific to a particular application; others are generic and carry across multiple application environments. A set of prompting recommendations is presented below.

### 4.1. Less is more

Lengthy, explanatory prompts tend to elicit higher abandon rates than short, directive prompts. The *PREVIU* trial compared prompts modeled after Touch-Tone prompts (as in 1) to prompts designed more directly for speech applications (as in 2):

1. For billing or account information, say, billing; if you are calling to place an order or check on the status of an order, say, order.
2. Please say: billing, or order.

<sup>2</sup> Note, however, that carefully controlled comparisons are not possible in field trials designed around actual services. Customers in different services differ with respect to their prior experience and expectations.

Caller participation showed a small but significant improvement with the shorter prompts displayed in 2. There was no significant effect on the number of in-vocabulary, isolated phrase responses as a function of these two prompting styles. These results were replicated in the *Money Talks* trial as well.

### 4.2. Machines should not pose as humans if they cannot understand as well as humans

As long as machines require more constrained and carefully articulated speech input than humans, they need to identify themselves as machines. In the *DACC* trial, callers were presented with multiple prompting variations. The goal of the prompt was to elicit isolated *yes* responses. Prompt 3 is a pre-recorded version of what operators said to callers. Prompt 4 is more clearly designed for automation.

3. Would you like us to complete this call for an additional charge of 35 cents?
4. Please say *yes* if you would like us to complete this call for an additional charge of 35 cents; otherwise, hang up.

The more natural sounding prompt 3 elicited a higher level of service acceptance, but a much lower level of isolated *yes* responses. Overall, prompt 4 was the more successful prompt, given the technology of that time.

In the *OSF* trial, several prompts were compared with respect to their success at eliciting city name responses from customers. Two sample prompts are presented below:

5. After the beep, please say just the name of the city.
6. What city, please? (beep)

Responses were analyzed with respect to what proportion of callers responded by stating their full listing request (that is, a name and address, instead of just the city name) in response to queries 5 and 6. For prompt 5, none of the customers responded with a full listing request. For prompt 6, 14% offered the complete listing request in response to the prompt. Prompts that provide marks of machine-like rigidity are clearly preferred for speech recognition applications, until automated speech systems can adequately handle any speech response that callers provide.

### 4.3. How do users speak to automated services?

Speech recognition development is advancing to provide more sophisticated caller interfaces, allowing users to speak more naturally to machines. Results from a number of studies, however, suggest that keyword recognition may be adequate for many of the services now targeted. [3,4,6].

In *Money Talks*, interfaces modeling keyword recognition outperformed the interfaces modeling NL capability. In scenario 5, callers were presented with open queries from the outset, such as:

7. You can get account information, or transfer funds between accounts. What would you like to do?

In response to prompt 7, callers frequently requested “*main menu*”. We believe that callers had a mental model that one cannot speak to machines in the same way we speak to humans, but they did not have a model of how, precisely, to customize their speech.<sup>3</sup>

In *PREVIU*, callers were asked to say keywords, but they were not explicitly asked to speak in isolation nor were they penalized with a reprompt if they embedded the required vocabulary in extra verbiage. Nonetheless, approximately 84% of the compliant keyword responses were spoken in isolation.

## 4.4. Audible Quoting

Callers participating in speech automated services must be made aware of the suitable vocabulary choices. One method that has been explored is "audible quoting." [5] Audible quoting presents two voices for a single prompt, with one voice presenting the carrier phrase and the second voice stating the possible vocabulary choices:

*female voice:* Please say one of the following:

*male voice:* checking, savings, or money market.

Audible quoting provides a subtle method for highlighting key words. We have incorporated this as an experimental dimension in both the *Money Talks* trial and the *PREVIU* trial. In *PREVIU*, audible quoting was not significantly better than standard prompts at eliciting caller participation or enhancing the quality and correctness of callers' responses. In *Money Talks*, however, callers presented with audible quoting required fewer reprompts. The value of audible quoting in eliciting in-vocabulary speech, based on these two field trials, remains inconclusive.

## 4.5. Barge-in: Do customers talk over prompts?

Barge-in is a familiar feature with Touch-Tone automated services. It is not *a priori* clear how frequently callers will barge-in on *speech* automated systems. First, protocols of turn-taking may inhibit callers' willingness to talk while the prompt is still playing. Second, speech prompts can be considerably shorter than Touch-Tone prompts, as in the following examples:

8. For repairs, press 1; for billing information, press 2; for sales, press 3...

9. Please say repairs, billing, or sales.

Data on barge-in from multiple experiments suggest that callers are not averse to talking while the prompt is playing. Barge-in rates of a number of experiments are presented in Table 3:

		utterances with barge-in
<i>PREVIU</i>	terse prompts	10.8%
	wordy prompts	46.8%
<i>Money Talks</i>	terse prompts	37.0%
	wordy prompts	64.0%
<i>OSF</i>	all prompts	15.0%

**Table 3:** Barge-in rates for *PREVIU*, *Money Talks*, and *OSF*

Barge-in rates differ for different trials. It is apparent, however, that barge-in rates are extremely high when prompts are lengthy; the barge-in percentages drop when prompts are brief. Many of the callers in all of these field trials were also new to the speech-automated interface, and were therefore obliged to listen to the entire prompt in order to know what options were available. We

<sup>3</sup> Callers are familiar with menu-driven systems, and expect to be told explicitly how to advance from one menu to the next. We anticipate that preference for a natural language interface will increase after customers have become more familiar and experienced with speech automated services, and realize that they no longer need the crutch of an explicit menu.

expect these barge-in rates to rise with increased familiarity with these services.

It appears, therefore, that barge-in capability should be incorporated in speech applications as well as Touch-Tone applications. Echo cancellation technology must be robust enough so that unintended vocalizations do not erroneously cut off the prompt, resulting in caller confusion. Other methods of prompt cancellation, such as asking callers to press the pound key (#) before they begin to speak, are currently under exploration.

## 5. CONCLUSIONS

The field trials described have been conducted under real user conditions. As such, they do not easily lend themselves to controlled comparisons. Nonetheless, some generalities have emerged that are applicable to other speech automated services as well. For example, terse prompts are, overall, more acceptable than wordy prompts. If barge-in is to be discouraged, terse prompts followed by a beep tone elicit the best effect. Callers do well if they are given a choice to use Touch-Tone or speech; application designers should not assume that speech input will be the automatic preference.

Service dimensions also impact speech technology success. In the *Money Talks* system, callers expected a Touch-Tone interface, which had been in place for six years. Caller participation with the Touch-Tone system is relatively high. The introduction of a speech interface supplanting the popular Touch-Tone interface was not well received. The *PREVIU* customer calling center application, however, has only recently been automated via Touch-Tone, and callers to the Touch-Tone system are not enthusiastic users. Many of the customer calling center participants called with no expectation about whether the call would be automated at all. These results suggest that speech recognition applications will be most successful when deployed with services that are new, or with Touch-Tone automated services that are achieving limited success.

## 6. REFERENCES

1. Basson, S., Christie, O, Levas, S., Spitz, J. *Evaluating Speech Recognition Potential in Automating Directory Assistance Call Completion*, AVIOS Proceedings, 1989.
2. NYNEX Science & Technology Technical Report, *Operator Store and Forward: Operator Performance and Customer Interaction*, 1995.
3. Springer, S., Basson, S., Kalyanswamy, A., Man, E. and Yashchin, D. *The Money Talks Interactive Speech Technology Assessment: A Report from the Field*, EURO-SPEECH, 1995.
4. NYNEX Science & Technology Technical Report TM95-0045, *PREVIU Trial Assessment*, 1995.
5. Bellcore Technical Report SR-NPL-000681, *Results of the Automated Alternate Billing Service (AABS) Trial*, June, 1987.
6. Deffner, G. & Melder, K. *User Acceptance and Preference for Advanced Voice Services Features and Dialogue Styles*. Proceedings of the Human Factors Society 34th Annual Meeting, Human Factors Society, 1990.