Comparison of Two Delayed Confirmation Strategies for Speech Recognition Interactive Voice Response Systems

TR 29.3675 August 18, 2003

Cheryl A. Kotan James R. Lewis

IBM Pervasive Computing

Boca Raton, Florida

Abstract

Guidelines for speech user interfaces generally promote the use of delayed confirmation in speech recognition interactive voice response (IVR) applications. A previous investigation of a simple delayed confirmation method revealed a significant design flaw, prompting the development of two alternative methods for delayed confirmation. To avoid the major design flaw of the previous method (replaying all collected information after each change, which forced users to repeatedly and unnecessarily hear a large amount of correct information when making multiple changes), both of the new methods required users to remember the items that needed to change. The data collection procedure was the same in each new method – collecting information one item at a time without confirming the entry, and then playing a review of all of the items. After the review, the two methods differed in their error-correction procedures. In one method (Serial Collection/Correction), the user named an item that needed correction, and then made that correction before naming the next item to change. In the other (Batch Collection/Correction), the user named all items that needed correction first, and then changed the named items in sequence. In an experiment we conducted to compare the methods, we required users to make two changes during the correction phase with each method. The data indicated that there was no significant difference in user preference between the methods, but significantly fewer memory errors occurred when using the Serial Collection/Correction method. Based on this finding, we recommend the use of the Serial Collection/Correction method for delayed confirmation of multiple items of information in speech recognition IVRs. We also recommend including an option to replay the information if a user has remembered the need to change an additional item, but does not remember which item to change.

ITIRC Keywords

Data correction
Data collection
Delayed confirmation
Batch confirmation
Confirmation strategy
Automated customer service
Voice systems
Speech systems
Interactive voice response (IVR)
Speech user interface

Contents

INTRODUCTION	1
METHOD	3
PARTICIPANTSAPPARATUS AND MATERIALSPROCEDUREEXPERIMENTAL DESIGN	3 4
RESULTS	
AFTER SCENARIO QUESTIONNAIRE (ASQ) OBSERVED TASK PERFORMANCE ACCURACYSPOKEN AND WRITTEN COMMENTS	5
DISCUSSION AND RECOMMENDATIONS	7
REFERENCES	9
APPENDIX A. ABSTRACTED TRANSCRIPTS OF TURNS REQUIRED TO COMPLETE TASKS	11
PROGRAM 1: PAYMENT BY ELECTRONIC CHECK, SERIAL COLLECTION/CORRECTION	12 13
PROGRAM 2: PAYMENT BY CREDIT CARD, BATCH COLLECTION/CORRECTION	

Introduction

Guidelines for speech user interfaces (Balentine & Morgan, 2002; IBM, 2001) generally promote the use of delayed confirmation in speech recognition interactive voice response (IVR) applications. A previous investigation of a simple delayed confirmation method (Lewis, 2003) revealed a significant design flaw, prompting the development of two alternative methods for delayed confirmation. To avoid the major design flaw of the previous method (replaying all collected information after each change, which forced users to repeatedly and unnecessarily hear a large amount of correct information when making multiple changes), both of the new methods required users to remember the items that needed to change following a review step. After the review, the two methods differed in their error-correction procedures. In one method (Serial Collection/Correction), the user named an item that needed correction, and then made that correction before naming the next item to change. In the other (Batch Collection/Correction), the user named all items that needed correction first, and then changed the named items in sequence. The following figures illustrate the two methods:

```
Figure 1. Serial Collection/Correction
```

Sys: Date? Hum: May 3rd. Sys: Time? Hum: 2:00 PM.

Sys: Let's review. Was that May 1st at 2:00 AM?

Hum: No

Sys: What needs to change? Select date or time.

Hum: Date.

Sys: New date?

Hum: May 3rd.

Sys: Was that May 3rd?

Hum: Yes.

Sys: Anything else need to change?

Hum: Time.

Sys: New time?

Hum: 2:00 PM.

Svs: Was that 2:00 PM?

Hum: Yes.

Figure 2. Batch Collection/Correction

Sys: Date? Hum: May 3rd. Sys: Time? Hum: 2:00 PM.

Sys: Let's review. Was that May 1st at 2:00 AM?

Hum: No.

Sys: What needs to change? Select date or time.

Hum: Date

Sys: We'll make that change in a moment. Change anything else?

Hum: Yes, the time.

Sys: New date?

Hum: May 3rd.

Sys: Was that May 3rd?

Hum: Yes.
Sys: New time?
Hum: 2:00 PM.

Svs: Was that 2:00 PM?

Hum: Yes.

Note that both of these designs require nine dialog turns to complete in the given examples, but do not require users to listen repeatedly to the full confirmation message or to a complete menu of items to change for each correction, addressing the major usability problem observed for the simple delayed confirmation strategy (Lewis, 2003). Before deploying applications using these strategies, though, it is important to prototype and test them to understand their relative benefits and potentially unknown drawbacks.

For example, it isn't clear which approach has the least demand on working memory during task completion. The Serial design (Figure 1) requires users to make the first correction before selecting the second item to change. Performing the first correction action could disrupt the maintenance of additional items to change in working memory. The Batch design (Figure 2) lets users select all items to change before beginning the correction step, but does not conform to the correction steps that users seem to expect (to make a change immediately after identifying the element to change). This unexpected turn in the interaction could have a disruptive effect on the maintenance of additional items to change in working memory. The purpose of the current study was to conduct an experiment to investigate these two confirmation strategies.

Method

Participants

Eight users (five recruited from a temporary employment agency and three IBM employees) participated in this study. Four participants were female (3 < 40 years of age; 1 > 40 years of age) and four were male (2 < 40 years of age; 2 > 40 years of age). All participants had previously used credit cards or checks to make online purchases, had given credit card or check information over the phone, and had called or used a speech recognition system at least once in the past.

Apparatus and Materials

Participants used an Andrea¹ NC-61 microphone headset plugged into an IBM ThinkPad² running Microsoft Windows³ 2000 to provide spoken input to two VoiceXML programs. The VoiceXML programs were written and executed using the IBM WebSphere⁴ Voice Toolkit, Version 3.1. The programs performed the same function: collecting and confirming payment information for a fictional billing transaction. Participants received fictional checking account information on a sheet of paper, and got fictional credit card information in the form of a fake credit card. The checking information included a check number, checking account number, and an ABA routing number. The fake credit card depicted the credit card type, card account number, and expiration date on the front. A three-digit security code appeared on the back of the card. The credit card billing zip code was on a separate sheet of paper.

Both programs used the same introduction. The system welcomed the participant, told them their account balance, and asked what amount they would like to pay. It then asked whether the user was paying by check or credit card. From there, it proceeded to collect the appropriate payment information without giving the user any feedback. After collecting the data, the system read back each item the user entered in a batch review. From that point on, the two programs were different. Program 1 asked the user to name an item that needed changing, and would then go on to help the user make the change before asking if another item needed correction (Serial Collection/Correction). Program 2 kept asking the user to name items that needed changing until the user indicated that no other items were incorrect. After that, Program 2 prompted the user to change each of the items they had just listed (Batch Collection/Correction). Each program was intentionally "bugged" to read back two errors in the review. In the check condition, the check number and checking account number were incorrect. In the credit card condition, the account number and billing zip code were wrong. After users made the corrections, the programs both finished in the same way, asking users if they were ready to process the transaction, telling them the transaction had processed, and asking if they were ready to exit. If the user was ready, the system thanked them for calling and told them goodbye (see Appendix A for sample dialogs).

Participants also used a demographic/background sheet, a program comparison sheet, and two sheets with task information, fictional account information, and an After Scenario Questionnaire (ASQ – Lewis, 1995). The ASQ contains three questions, each scored from 1-7, with lower scores indicating greater satisfaction and ease of use (see Appendix B). The overall score is the arithmetic mean of the

3

¹ Andrea, Andrea Electronics and its products, logos and services are either trademarked or registered trademarks of Andrea Electronics Corporation or Lamar Signal Processing, Ltd.

² IBM and ThinkPad are registered trademarks of International Business Machines Corp.

³ Microsoft and Windows are trademarks or registered trademarks of Microsoft Corp.

⁴ WebSphere is a registered trademark of International Business Machines Corp.

three item scores. In this study, we added a fourth question with the same format as the first three, pertaining to the ease of correcting the billing information. The overall score in this study was the arithmetic mean of the four item scores.

Procedure

Each participant completed a demographic/background form. Employees of the temporary employment agency also completed a nondisclosure form. Participants were the microphone headsets with the earphone on their preferred ear and the microphone positioned to minimize breath noise.

The participant then went through the IBM Voice Toolkit audio level setup procedure. They listened to a music clip to adjust the earphone volume, and then read a brief paragraph during which time the system adjusted the microphone input level to maximize the signal to noise ratio. After completing the setup procedure, participants received instructions to use two different programs to pay a bill over the phone, with the "phone" being the headset. They were to pay once with a check (with the account information on the task sheet), and once with a fake credit card. Each participant was aware that the fake credit card had information on the front and back, and that the billing zip code was available on the task sheet.

After the introduction, participants began their assigned tasks. They spoke the payment information when prompted for it and listened to the review. With the two errors already built in, the participant requested and made what they felt were the appropriate changes, then proceeded to process the transaction and exit the system. They answered the four ASQ questions about the first task and gave comments before going on to pay the bill again with the other program using the other payment method. They filled out the ASQ and gave comments for the second program, then filled out the final comparison sheet. They indicated which program they liked the best, and rated the degree to which they preferred their favorite on a 7-point scale. Finally, they commented about what they liked and disliked.

Experimental Design

As shown in Table 1, the experimental design counterbalanced the order of presentation for correction style (Serial vs. Batch), the order of presentation of the payment task (Check vs. Credit Card), and the pairing of correction style and payment task.

Table 1. The Experimental Design

Participant	Task 1	Task 2		
1	Batch-Check	Serial-Card		
2	Batch-Card	Serial-Check		
3	Serial-Check	Batch-Card		
4	Serial-Card	Batch-Check		
5	Batch-Check	Serial-Card		
6	Batch-Card	Serial-Check		
7	Serial-Check	Batch-Card		
8	Serial-Card	Batch-Check		

Results

After Scenario Questionnaire (ASQ)

Table 2 gives the means and standard deviations for the ASQ ratings for each style of correction and payment method. Participants rated all tasks better than the scale midpoint.

Table 2. ASQ Ratings for Correction Style and Payment Methods

Correction Style	Mean	SD		
Serial	2.41	0.98		
Batch	2.50	1.36		

Payment Method	Mean	SD		
Check	2.38	1.38		
Credit Card	2.53	0.94		

Two-tailed *t*-tests indicated no statistically significant difference (with the lowest value of *p* equal to .40) between the programs for any of the ASQ overall ratings or item ratings (ease of performing task, time task took to complete, support information available, and ease of correcting errors).

On the comparison form, 75% of participants indicated that Program 1 (Serial Collection/Correction) was their favorite (with the 90% binomial confidence interval for this preference ranging from 40% to 95.5%). Seventy-five percent also selected the program for which they used a check, probably because it required them to input fewer and shorter items. The payment method should not have influenced the overall ASQ ratings, however, because payment method and correction style were balanced across the eight subjects.

Observed Task Performance Accuracy

All participants successfully completed the bill payment task using Program 1 (Serial Collection/Correction). Three of the eight participants failed to complete the payment task with Program 2 (Batch Collection/Correction) due to non-technical causes. All three failures occurred when participants were trying to change erroneous items. A statistical test of this difference using a test for nonindependent samples in 2 x 2 tables (Steel & Torrie, 1980, p. 506-507) indicated that this difference in failure rates was marginally significant ($\chi^2(1) = 3$, p = .08).

Spoken and Written Comments

Of the participants who failed using Program 2 (Batch Collection/Correction), one thought the account number may have been wrong, but assumed that his memory was incorrect and that the system would not let him proceed if the account number was wrong. The next participant changed the account number but not the billing zip code because she didn't remember it being wrong. The third didn't realize she should say "no" if there was nothing else to correct ("It never let me make the changes.") She kept cycling back and forth naming the two things she wanted to change over and over until she exceeded the allowable number of requests for change.

Some participants specifically said they liked the Serial Collection/Correction style better, describing it as "good error recovery for mistaken data entry." Participants also said they liked having the ability to review the information they entered and make any necessary changes. A few suggested that they would like to be able to request the review at any point to prevent losing track of their changes. Many

participants reported liking check pays	ment better because they didn't have to enter as many items and
the account numbers weren't as long.	Some wanted the speed of the review slowed down.

Discussion and Recommendations

Significantly fewer task failures occurred when using the Serial Collection/Correction method than with the Batch Collection/Correction method. There were no differences in the ASQ ratings for the two methods, but when asked directly, participants generally appeared to prefer the Serial method over the Batch method.

Recommendation 1: Use the Serial Collection/Correction method for delayed confirmation of multiple items of information in speech recognition IVRs.

Some participants indicated that it would be useful to be able to request a review of the entered data at any time during the process of correction. For example, it would be possible to include this option as the final option when listing the items available for change in the confirmation procedure.

Recommendation 2: Provide an option to replay the information if a user has remembered the need to change an additional item, but does not remember which item to change.

In this study, participants engaged in immediate confirmation of items after making a change, and never had to review the entire set of entered data. This might be appropriate in some applications, but in other applications (especially financial applications) it might be necessary to have users go through a final review before accepting the data for processing.

Recommendation 3: If an application requires a final review of all data before accepting the data for processing, then do not engage in immediate confirmation during changes. If there is no need for a final comprehensive review of entered data, then use immediate confirmation of changes.

References

- Balentine, B., & Morgan, D. M. (2001). How to build a speech recognition application: A style guide for telephony dialogs (2nd ed.). San Ramon, CA: Enterprise Integration Group.
- International Business Machines Corp. (2001). *IBM VoiceXML Programmer's Guide*. Boca Raton, FL: Author.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7, 57-78.
- Lewis, J. R. (2003). *Preliminary comparison of immediate and delayed confirmation strategies as a function of the number of corrections* (Tech. Report in press). Boca Raton, FL: International Business Machines Corp.
- Steel, R. G. D., & Torrie, J. H. (1980). *Principles and procedures of statistics*. New York, NY: McGraw-Hill.

Appendix A. Abstracted Transcripts of Turns Required to Complete Tasks

Program 1: Payment by Electronic Check, Serial Collection/Correction

Sys: Welcome to Maxtel Communication's automated Pay by Phone service.

You can say Repeat or Help at any time.

Sys: Your most recent payment was sixty seven dollars and sixteen cents on

April 23rd. That payment was made in the past 72 hours. Your account is current, with a balance of

sixty seven dollars and thirty four cents.

Hum: Pay it all. Svs: Pay \$67.34?

Hum: Yes.

Sys: Pay with credit card or check?

Hum: Check.

Sys: To pay by electronic check, please have your check number, checking account number, and ABA

routing number available. Say or enter your check number.

Hum: 5210

Sys: The ABA routing number is a nine-digit number located at the bottom of your check. Say or enter the

routing number.

Hum: 374190351

Sys: Say or enter your checking account number.

Hum: 644398

Sys: Let's review. Pay with check 9 2 1 4, using routing number 3 7 4 1 9 0 3 5 1, and checking account

number 6 4 1 3 5 6? Is all of this information correct?

Hum: No.

Sys: What would you like to change?

Hum: Check number.

Sys: Say or enter your check number.

Hum: 5210

Sys: Was that 5 2 1 0?

Hum: Yes.

Sys: Change anything else? Hum: Yes, account number.

Sys: Say or enter your checking account number.

Hum: 644398

Sys: Was that 6 4 4 3 9 8?

Hum: Yes.

Sys: Change anything else?

Hum: No.

Sys: Ready to make electronic check payment of \$67.34. Say Process Transaction or Cancel.

Hum: Process transaction.

Sys: Transaction complete. The confirmation number is 8 2 1 5 7 9. Visit our web site to learn how to set up

a schedule for paying your bill automatically. To continue, select Repeat or Exit.

Hum: Exit.

Sys: Are you sure you want to end this call?

Hum: Yes.

Program 1: Payment by Credit Card, Serial Collection/Correction

Sys: Welcome to Maxtel Communication's automated Pay by Phone service.

You can say Repeat or Help at any time.

Sys: Your most recent payment was sixty seven dollars and sixteen cents on

April 23rd. That payment was made in the past 72 hours. Your account is current, with a balance of

sixty seven dollars and thirty four cents.

Hum: Pay it all. Sys: Pay \$67.34?

Hum: Yes.

Sys: Pay with credit card or check?

Hum: Credit card.

Sys: Which card?

Hum: Master Card.

Sys: For security and accuracy, we will need your credit card's number, expiration date, billing zip code and

the security code located on the back of the card.

Sys: Say or enter the sixteen digit credit card number.

Hum: 5424379186249535

Sys: Expiration date? Hum: 0 6 0 1 0 4

Sys: Credit card billing zip code?

Hum: 63851

Sys: Check the back of your credit card. Do you have at least three digits to

the right of the credit card number located on the signature panel?

Hum: Yes.

Sys: Are the last three digits legible?

Hum: Yes.

Sys: Say or enter the last three digits.

Hum: 573

Sys: Let's review. Charge to Master Card, account number 5 4 0 4 6 7 5 1 8 6 2 0 9 5 3 9, expiring 06 01 04,

with security code 5 7 3, and billing zip 6 3 4 9 1. Is this correct?

Hum: No.

Sys: What would you like to change?

Hum: Card number.

Sys: Say or enter the credit card number.

Hum: 5424379186249535

Sys: Was that 5 4 2 4 3 7 9 1 8 6 2 4 9 5 3 5?

Hum: Yes.

Sys: Change anything else?

Hum: Yes, billing zip.

Sys: Credit card billing zip code?

Hum: 63851

Svs: Was that 6 3 8 5 1?

Hum: Yes.

Sys: Change anything else?

Hum: No.

Sys: Ready to make Master Card payment of \$67.34. Say Process Transaction or Cancel.

Hum: Process transaction.

Sys: Transaction complete. The confirmation number is 8 2 1 5 7 9. Visit our web site to learn how to set up

a schedule for paying your bill automatically. To continue, select Repeat or Exit.

Hum: Exit.

Sys: Are you sure you want to end this call?

Hum: Yes.

Program 2: Payment by Electronic Check, Batch Collection/Correction

Sys: Welcome to Maxtel Communication's automated Pay by Phone service.

You can say Repeat or Help at any time.

Sys: Your most recent payment was sixty seven dollars and sixteen cents on

April 23rd. That payment was made in the past 72 hours. Your account is current, with a balance of

sixty seven dollars and thirty four cents.

Hum: Pay it all. Sys: Pay \$67.34?

Hum: Yes.

Sys: Pay with credit card or check?

Hum: Check.

Sys: To pay by electronic check, please have your check number, checking account number, and ABA

routing number available. Say or enter your check number.

Hum: 5210

Sys: The ABA routing number is a nine-digit number located at the bottom of your check. Say or enter the

routing number.

Hum: 374190351

Sys: Say or enter your checking account number.

Hum: 644398

Sys: Let's review. Pay with check 9 2 1 4, using routing number 3 7 4 1 9 0 3 5 1, and checking account

number 6 4 1 3 5 6? Is all of this information correct?

Hum: No.

Sys: What item would you like to change?

Hum: Check number.

Sys: We'll make that change in a moment. Change anything else?

Hum: Yes, account number.

Sys: We'll make that change in a moment. Change anything else?

Hum: No.

Sys: Say or enter your check number.

Hum: 5210

Sys: Was that 5 2 1 0?

Hum: Yes.

Sys: Say or enter your checking account number.

Hum: 644398

Sys: Was that 6 4 4 3 9 8?

Hum: Yes.

Sys: Ready to make electronic check payment of \$67.34. Say Process Transaction or Cancel.

Hum: Process transaction.

Sys: Transaction complete. The confirmation number is 1 7 2 3 6 8. Visit our web site to learn how to set up

a schedule for paying your bill automatically. To continue, select Repeat or Exit.

Hum: Exit.

Sys: Are you sure you want to end this call?

Hum: Yes.

Program 2: Payment by Credit Card, Batch Collection/Correction

Sys: Welcome to Maxtel Communication's automated Pay by Phone service.

You can say Repeat or Help at any time.

Sys: Your most recent payment was sixty seven dollars and sixteen cents on

April 23rd. That payment was made in the past 72 hours. Your account is current, with a balance of

sixty seven dollars and thirty four cents.

Hum: Pay it all. Sys: Pay \$67.34?

Hum: Yes.

Sys: Pay with credit card or check?

Hum: Credit card.
Sys: Which card?
Hum: Master Card.

Sys: For security and accuracy, we will need your credit card's number, expiration date, billing zip code and

the security code located on the back of the card.

Sys: Say or enter the sixteen digit credit card number.

Hum: 5424379186249535

Sys: Expiration date? Hum: 0 6 0 1 0 4

Sys: Credit card billing zip code?

Hum: 63851

Sys: Check the back of your credit card. Do you have at least three digits to

the right of the credit card number located on the signature panel?

Hum: Yes.

Sys: Are the last three digits legible?

Hum: Yes.

Sys: Say or enter the last three digits.

Hum: 573

Sys: Let's review. Charge to Master Card, account number 5 4 0 4 6 7 5 1 8 6 2 0 9 5 3 9, expiring 06 01 04,

with security code 5 7 3, and billing zip 6 3 4 9 1. Is this correct?

Hum: No.

Sys: What would you like to change?

Hum: Card number.

Sys: We'll make that change in a moment. Change anything else?

Hum: Yes, billing zip.

Sys: We'll make that change in a moment. Change anything else?

Hum: No.

Sys: Say or enter the credit card number.

Hum: 5424379186249535

Sys: Was that 5 4 2 4 3 7 9 1 8 6 2 4 9 5 3 5?

Hum: Yes.

Sys: Credit card billing zip code?

Hum: 63851

Sys: Was that 6 3 8 5 1?

Hum: Yes.

Sys: Ready to make Master Card payment of \$67.34. Say Process Transaction or Cancel.

Hum: Process transaction.

Sys: Transaction complete. The confirmation number is 1 7 2 3 6 8. Visit our web site to learn how to set up

a schedule for paying your bill automatically. To continue, select Repeat or Exit.

Hum: Exit.

Sys: Are you sure you want to end this call?

Hum: Yes.

Appendix B. After Scenario Questionnaire (ASQ)

TELL THE TEST MONITOR WHEN YOU FINISH THIS TASK. Then complete the following short questionnaire.

For each of the statements below, circle the rating of your choice.									
1.	Overall, I am satisfied with the ease of completing this task.								
	STRONGLY AGREE	1	2	3	4	5	6	7	STRONGLY DISAGREE
2.	Overall, I am satisfied with the amount of time it took to complete this task.							ask.	
	STRONGLY AGREE	1	2	3	4	5	6	7	STRONGLY DISAGREE
3.	Overall, I am satisfied with the support information (on-line help, messages, documentation) when completing this task.								
	STRONGLY AGREE	1	2	3	4	5	6	7	STRONGLY DISAGREE
4.	Overall, I am satisfied with the ease of correcting the billing information.						1.		
	STRONGLY AGREE	1	2	3	4	5	6	7	STRONGLY DISAGREE