경영정보학연구
제9권 제2호
1999년 6월

Can Structured Feedback in a Restricted Natural Language Database Interface Improve Casual User Performance?*

Suh, Kil Soo**, William C. Perkins***

한정 자연어 방식의 데이터베이스 사용자 접속에 있어서 구조적인 피드백의 효과

서 길 수, 윌리엄 퍼킨스

A feedback echo is a structured restatement of the user's query and informs the user what the system intends to do for the query. This provides backtracking of the user's query so that the user knows how the system has interpreted the query. This study scrutinizes whether the feedback echo, as currently available in the INTELLECT commercial system, can improve the performance of a casual user using a restricted natural language database interface. This study concludes that the particular type of feedback echo available in this commercial system was not effective in terms of overall performance for casual users. It is worth mentioning, however, that the feedback echo was effective for the specific type of error: using wrong conditions for data retrieval.

---

* An earlier version of this paper appeared in the Proceedings of the Twenty-Seventh Hawail International Conference on System Sciences(IEEE Computer Society Press, 1994). This research was supported by the Suam Foundation of Korea and the Institute for Research on the Management of Information Systems (IRMIS) of Indiana University.

** College of Business and Economics, Yonsei University
*** The Kelly School of Business, Indiana University

I. Introduction

A database is the heart of computer-based information systems. A database query language (DBQL) is a special-purpose language by which a user is able to issue high-level commands or statements to retrieve information from a database (Date, 1995). A DBQL is primarily intended for use by end-users rather than professional programmers; thus, the functionality provided by database systems is determined mostly by user interfaces (Stonebraker, 1988). The field of DBQLs is approaching maturity, and the focus has begun to shift towards user interface (Owei and Higa, 1994).

One approach to enhance user interface of DBQLs is employing natural language. Natural language systems allow users to carry on a dialogue with the computer in their own native language, rather than a structured command language. Barbary (1987) argued that traditional structured DBQLs were not adequately serving the needs of the novice end-users. He suggested that natural language, by permitting users to communicate with the system in their own native language, was an attractive approach. A complete natural language system, however, is not currently available, nor expected in the near future. By contrast, restricted natural language (RNL) interfaces have become more common and easier to use since the first commercial system was introduced in 1981 (Stevens, 1986).

Results on usability of RNL database interfaces have been conflicting. Dekleva (1994) reported that an existing RNL system, INTELLECT, might be practical for novice users, and other studies (Suh and Jenkins, 1992; Napier et al., 1989) have shown that an RNL interface was superior to a structured language interface. Ein-dor and Spiegler (1995) proposed a model for natural language access to multiple databases to enhance the usability of data retrieval which was an extension of RNL interfaces for a single database. On the other hand, RNL systems are criticized by their verbosity, inability to handle ambiguities, and lack of guidance (Capindale and Crawford, 1990). For example, users of the RNL systems committed "over-confidence" errors nearly twice as often as structured language users (Suh, 1989). These errors occurred when there was miscommunication (a valid query but not the intended one) between the user and the system because of the ambiguity and imprecision of queries expressed in an RNL.

Many studies (Batra and Sein, 1994; Molich and Nielsen, 1990; Slator et al., 1986) has suggested that appropriate feedback would improve user performance by reducing miscommunication between the user and the system. The statement that feedback will improve user performance has strong face validity, but it has not yet been examined empirically. The objective of this study is to examine the effects of a structured feedback in an RNL database interface on a casual user's performance. This study scrutinizes whether the structured feedback, as currently available in a commercial system, INTELLECT, can improve the performance of a casual user using an RNL database interface.

Ⅱ. Background

This section examines the relevant literature, including that on the problems of RNL, the concept of feedback, and the use of a feedback echo.

2.1 Problems of Restricted Natural Language

A study (Suh, 1989) comparing an RNL database interface with SQL concluded that RNL provides a better user interface than SQL for novices performing data retrieval tasks. However, RNL subjects did experience problems of miscommunication between the user and the system. From the results of this study, the incorrect solutions (queries) can be classified as either:

Type 1: The solution was incorrect, and the subject knew that it was wrong.
Type 2: The solution was incorrect, but the subject did not know that it was wrong.

In case of the RNL interface, 86% of the incorrect solutions were of type 2. This percentage was almost twice as high as in SQL. The type 2 error arises largely from miscommunication between the user and the system. This result strongly suggests that methods of improving the communication ability of an RNL should be studied. A feedback echo, which is explained in detail below, can be a candidate method.

2.2 Concept of Feedback

Feedback has long been of central concern in psychological studies of group experience and decision making (Wing, 1990; Hogarth et al., 1991). According to Napier and Gershenfeld (1973), feedback is nothing more than a process to find out whether the message intended is the message actually received. Even though most research on feedback has considered feedback effects as a black box, Payne and Hauty (1955) suggested two possible mechanisms: motivational and directive properties. The motivational explanation asserts that feedback acts as a reward or reinforcement that leads to enhanced performance. Under the directive explanation, feedback contains information that allows the subjects to correct errors in performance.

2.3 Feedback Echo

A feedback echo is a restatement of the user's query and informs the user what the system intends to do for the query. This provides backtracking of the user's query so that the user knows how the system has interpreted the query. Some commercial RNL systems such as INTELLECT provide this function using a structured language. A feedback echo in an RNL system is expected to reduce miscommunication between the user and the system. Because the feedback echo contains information that allows the user to correct errors, its effects, if any, seems to be produced by the directive property rather than motivational.

Even in communication between humans, it is seldom true that one person completely understands everything the other person intended to convey. Guinan (1988) conducted a field experiment to study communication behaviors of highly-rated versus lowly-rated system developers. The results showed that highly-rated developers used specific communication behaviors more frequently than their lowly-rated counterparts while eliciting system requirements from a user. Among these specific communications behaviors the most important behavior was backtracking. Backtracking refers to statements made to verify what someone has said or experienced. Use of backtracking improved communication ability.

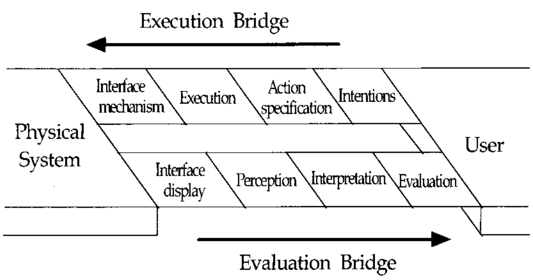
The feedback echo is a kind of backtracking provided by the system. In this way, ambiguities in communication between the user and the system can be reduced, and as a result user performance is expected to improve.

The feedback echo of an RNL system is also expected to educate a user subliminally such that the user mimics the echo when he or she issues a command. Slator et al. (1986) showed that appropriate mnemonic feedback built into a natural language interface could act as a teacher to help users acquire formal-language skills as they work. Another study (Capindale and Crawford, 1990) has shown that feedback helped users understand the language limitations and learn how to avoid or recover from errors.

In summary, a feedback echo of an RNL system should have positive effects both on user performance and on learning through its directive property if it is appropriately designed.

Ⅲ. Research Model and Hypotheses

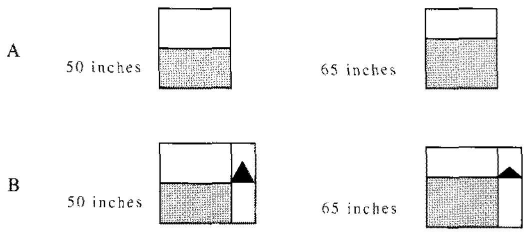
The user of a system expresses his goal in psychological terms, while the system presents its current state in physical terms. According to Norman (1986), the user's goals and the system state differ significantly in form and content, creating the gulfs that need to be bridged if the system can be used. The gulf of execution goes from the user to the system. This distance is bridged in four sequences: intention formation, specifying the action sequence, executing the action, and making contact with the input mechanisms of the interface <Figure 1>.



<Figure 1> Bridging the Gulfs of Execution and Evaluation

The gulf of evaluation, which is the focus of this study, goes from the system to the user. This gap is bridged in another four segments: starting with the output displays of the interface, moving to the perceptual processing of those displays, to its interpretation, and to the evaluation comparing its interpretation with the original goals and intentions. The amount of cognitive effort required to bridge these gaps can be attributed to the semantic distance and the the articulatory distance. As the focus of this research is on the evaluation side, the concept of semantic and articulatory distance in evaluation is examined in more detail.

Evaluation semantic distance refers to the mental process required for the user to determine whether the goal has been achieved. Hutchins et al. (1986) provides an excellent example of matching the user's intentions by appropriate output language in order to reduce evaluation semantic distance. When a user attempts to control the rate at which the water level in the tank is rising, the output in <Figure 2B> is more semantically direct than that in <Figure 2A>. Because <Figure 2A> does not provide the rate of change directly, the user has to observe the value over time and mentally compare the values at different times to calculate the rate of change. On the other hand, the rate of change is directly displayed by a little arrow in <Figure 2B>. This reduces the cognitive effort required by the user, making the evaluation semantic distance between user and system much shorter.



<Figure 2> Example of Two Output Languages

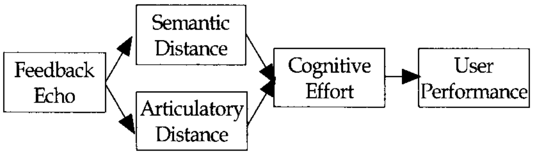
While the semantic distance is related to a decision about what to do, the articulatory distance is related to deciding how to do it. Evaluation articulatory distance concerns the relationship between the meanings of expressions and their physical output form. For example, if the user is interpreting the changes in net income over last 10 years, a graphical display can provide more articulatory directness than a table of numbers, even though both contain the same semantic information.

There are only two ways to bridge these distances between user and system: move the system closer to the user, or move the user closer to the system (Norman, 1986). The former means providing an interface that matches the user's psychological requirements, in a form that can be readily interpreted and manipulated. The latter means training the user to build up mental structures that are compatible with the system's.

3.1 Research Model

A feedback echo can be designed to reduce both the semantic and articulatory distances by moving the system closer to the user. The feedback echo should reduce the semantic distance by providing information that allows the user to correct errors; in addition, it should be presented in a manner easily understood by the user to reduce the articulatory distance. The research model of this study is illustrated in <Figure 3>. Use of a feedback echo should reduce the semantic and articulatory distances, and the resulting distances should reduce the cognitive effort required by the user. Finally cognitive effort inversely affects user performance. In this study, the semantic and articulatory distances and cognitive effort are treated as intervening variables.[1]

1) An intervening variable is an unobservable process or state associated with an organism that helps to explain linkages between an independent variable and a dependent variable (Stone, 1978, pp. 24-25)



<Figure 3> The Research Model

The present study focuses the effects of a feedback echo on the casual user. Casual users are those who access a database occasionally; thus, they may have difficulty remembering the rigid syntax if a structured language is employed for data retrieval. A restricted natural language database interface can benefit this type of user the most.

This study also focuses on the initial performance of the user, not on learning. Feedback can affect not only initial performance, but also learning, because a feedback echo can educate the user subliminally and thus move the user closer to the system. This is another important research issue, but it is not handled in this study. One of the claimed advantages of a natural language interface is that it requires little or no formal training; because of this claim, the authors chose to focus on initial performance of the casual user.

3.2 Hypotheses

Guinan (1988) found that backtracking was the most important tool for effective communication. In this study, a feedback echo is introduced as a backtracking tool which can reduce the gulf of evaluation between the user and the system. If the feedback echo can reduce the semantic and articulatory distances, the echo may enable subjects receiving the feedback to become more aware of their errors when their query does not retrieve the right data than will be the case for the subjects without the feedback echo. To test this hypothesis in the present study, subjects are allowed to correct their queries as many times as they wish after receiving results. The error recovery rate is defined as the percentage of the queries which are not correct at first, but modified once or more and finally are correct. The detailed description of the error recovery rate is provided in the Measurement section.

H1: The feedback echo (FE) group will perform better than the non-feedback echo (NE) group in terms of the error recovery rate.

The FE subjects receive a structured restatement of how the system interpreted their queries. If this restatement can reduce the feeling of miscommunication, then they will be more confident about the results than the NE subjects.

H2: The FE group will be more confident about their solutions than the NE group.

When a user has retrieved data using an RNL interface, two kinds of communication errors can occur. One is that the user considers wrong data as correct. In other words, because the user retrieved some data successfully, he or she automatically believes that these data are correct, even though they are not. This type of errors needs to be eliminated because the user may use the incorrect data to make a critical decision.

The other type of error is just the opposite.

In this case, even though the user retrieved the correct data, he or she believes that these data are incorrect. The problem in this case is that the user will not use the valid data in his or her decision-making. Either of these errors will have a negative impact on user performance. If it is well designed, the feedback echo is expected to reduce these two types of communication errors.

H3: The FE group will perform better than the NE group in terms of communication errors.

Ⅳ. Research Method

To examine the above hypotheses a posttest-only control group design was employed. Because subjects are randomly assigned to one of the two groups, with and without a feedback echo, it is a true experimental design (Campbell and Stanley, 1963).

4.1 Restricted Natural Language System

Several RNL interfaces for database systems are currently available in the market. From these interfaces, INTELLECT (1982) was selected as the RNL system for this experiment because it has obtained a notable success as an RNL system as well as it provides a feedback echo facility (Shneiderman, 1992). Typical examples of feedback echoes supplied by INTELLECT for user queries are listed in the Appendix. The basic structure of the feedback echo of INTELLECT is:

PRINT <items displayed>
OF ALL RECORDS WITH <selection conditions>.

4.2 Subjects

A total of 70 subjects were recruited for this study from the undergraduate student population enrolled in the introductory computing course in the School of Business at a large state university, and 67 subjects completed the experiment successfully. Demographic data for the subjects are provided in <Table 1>. There were no significant differences between the FE and NE groups on the basis of demographics, according to t-tests. Students received limited credit in the computing course by participating in the experiment. To enhance students' interest in the experiment, prizes of $20 were awarded to the subject in each group (FE and NE) with the highest number of correct queries (solutions).

4.3 Task

To capture various aspects of ease-of-use of a query language, experimenters have developed a number of different tasks--query writing, query reading, query interpretation, question compre-

<Table 1> Demographic Data for Subjects

Demo / Group	Major			Age			Typing			Total
	Busines	Others	Missing	16-20	21-25	26-30	Novice	Casual	Good	
FE	27	6	1	32	2	0	1	18	15	34
NE	27	5	1	29	3	1	3	18	12	33

hension, memorization, and problem solving (Reisner, 1981). Query writing has been among the most frequently used tasks (see Greenblatt and Waxman, 1978; Reisner, 1977; Suh and Jenkins, 1992; Tuner et al., 1984), and it was employed for this experiment. Each subject was given ten questions and required to enter a query for each question into the terminal. After entering the initial query, INTELLECT responded with an error message or a solution (which may or may not have been correct). Subjects could modify the query as many times as they wished, until they were either satisfied with the solution or they had no idea how to proceed.

4.4 Procedures

Subjects were randomly assigned into one of the two groups--34 subjects for the FE group and 33 subjects for the NE group. Subjects studied a training manual at their own pace at the terminal. The training manual included a series of increasingly more difficult queries to formulate and enter into the terminal. It took an average of about 46 minutes for the FE group and 45 minutes for the NE group to complete the training session, with no significant difference between the times for the two groups. This short training session did not provide sufficient time for the FE subjects to be educated in any significant way by mimicking the feedback echo.

After subjects finished the manual, they received a task booklet containing ten questions. The average times taken for the test sessions were 34 minutes for FE and 33 minutes for NE, again with no significant difference between the time for the two groups. For the NE group, the feedback echo facility of INTELLECT was turned off during both the training and test sessions.

4.5 Measurement

A query was graded by comparing the data retrieved with the data requested in the question. Each correct query received a score of 1, and there was no partial credit because subjects were allowed to change their queries as many times as they wished. The correct queries were classified as either:

Type 1: The solution was correct at first.
Type 2: The solution was not correct at first, but it was modified once or more and finally was correct.

Correct queries were divided into two types to examine the effects of the feedback echo on the error recovery ability of subjects. Because each subject was given 10 questions, the error recovery rate was calculated according to the following formula:

$$\text{Error recovery rate} = \frac{\text{Number of Type 2 solutions}}{10 - \text{Number of Type 1 solutions}} \times 100$$

The confidence of subjects in their solutions was measured using a single question after they finished each problem. The question was in the form of a 5-point Likert-type scale: 1=very sure query is correct, 2=fairly sure query is correct, 3=50-50 chance query is correct, 4=fairly sure query is incorrect, 5=very sure query is incorrect.

Communication errors occurred when the actual results from a query (correct or incorrect) did not match the confidence in the result expressed by the subject (very sure query is correct or very sure query is incorrect). Two types of communication errors are considered:

Type 1 error: the subject is very sure query is correct (confidence=1), but the actual result was wrong.

Type 2 error: the subject is very sure query is incorrect (confidence=5), but the actual result was right.

To increase the FE subjects' attention on the feedback echo, the echo was highlighted on the screen. In addition, the FE subjects were asked to assess the usefulness of the feedback echo for each problem using a single question. The question was in the form of a 5-point Likert-type scale: 1=quite helpful, 2=somewhat helpful, 3=neither helpful nor unhelpful, 4=somewhat unhelpful, 5=quite unhelpful. The mean score of the FE group on the usefulness question was 1.92. This score can be interpreted that the FE subjects felt the feedback echo for each problem was somewhat helpful on average.

V. Data Analysis

The results are summarized in <Table 2> and <Table 3>. The independent samples t-test is employed to analyze differences between various means of the two groups and the significance level of .05 is used. The two groups were not significantly different in terms of the number of type 1 correct queries (those queries which are correct on the first try). The FE group had an average of 3.71 and the NE group 4.09 type 1 correct queries. The standard deviations were quite large in both groups, 2.30 for the FE group and 2.40 for the NE group, which means the individual differences within groups were large. In fact, the best scores on the first try were 9 in both groups and the worst scores were 1 in the FE group and 0 in the NE group. The two groups were not different in terms of average number of trials for a question. The FE subjects made 2.24 trials per question and the NE subjects tried 2.39 times per question.

The fact that the two groups were not different in terms of the number of type 1 correct queries and the number of trials per question provided a sound basis on which to test hypothesis H1. If they were different, it would have been difficult to compare their error recovery rates because the two groups would not been starting from the same base.

<Table 2> Comparison of Two Groups for Initial Conditions

Measurement	Mean(Standard deviation)		t	p
	FE	NE		
Number of Type 1 Correct Queries	3.71 (2.30)	4.09 (2.40)	-.67	.51
Number of Trials per Question	2.24 (0.68)	2.39 (0.95)	-.74	.46

<Table 3> Summary of Hypothesis Tests

Measurement	Mean(Standard deviation)		t	p	Related Hypothesis
	FE	NE			
Error Recovery Rate	43.2% (29.8%)	41.9% (23.5%)	.20	.84	H1
Confidence Level	2.09 (0.62)	2.04 (0.66)	.35	.72	H2
Communication Error Rate	13.8% (10.5%)	13.0% (12.3%)	-.16	.88	H3

H1: The feedback echo (FE) group will perform better than the non-feedback echo (NE) group in terms of the error recovery rate.

Even though the FE group had a slightly higher error recovery rate (43.2%) than the NE group (41.9%), it was not a significant difference. The feedback echo facility provided by INTELLECT was not able to improve a casual user's error recovery ability. The FE subject felt that the feedback echo was somewhat helpful, but it did not actually help them.

H2: The FE group will be more confident about their solutions than the NE group.

There was no significant difference between the two groups in terms of confidence. The mean scores on confidence level were relatively high in both groups, 2.09 for the FE group and 2.04 for the NE group. In other words, subjects were fairly sure about the correctness of teir queries on average. Different from what was expected, the feedback echo of INTELLECT did not increase the confidence of casual users.

H3: The FE group will perform better than the NE group in terms of communication errors.

The communication error rates of the two groups were not significantly different, either. The average communication error rate of the FE group was 13.8% (12.6% believed that incorrect data were in fact correct, and 1.2% believed that correct data were incorrect), and that of the NE group was 13.0% (10.6% and 2.4%, respectively).

Ⅵ. Error Analysis

The feedback echo of INTELLECT did not improve the performance of casual users in terms of error recovery rate, confidence, and communication errors. This result is rather surprising, considering previous research about software feedback. Molich and Nielsen (1990) found that feedback is one of the most important features in terms of improving human-computer dialogue. Slator et al. (1986) have suggested that a feedback echo of a natural language interface could reduce the ambiguities inherent in a user's command. In addition, the INTELLECT manual emphasizes the importance of the feedback echo and recommends its use every time the user is querying.

To investigate this surprising result, every error and the subsequent trials made by the subjects were carefully scrutinized. The results of this detailed error analysis are reported in this section. Errors can be classified into two general categories according to whether a subject retrieved some data or not for the specific query. One type of errors occurs when a subject receives an error message and never corrects the error in the subsequent tries. In this case, the INTELLECT system will not provide a feedback echo for the query because it cannot interpret the query.

Another type of errors takes place when a subject retrieves some data but the data are wrong. In this case, the FE subject will receive a feedback echo with the data retrieved while the NE subject will not. Because of our interest in the effects of the feedback echo, we concentrated on these wrong data errors. These errors were divided into five categories

<Table 4> Summary of Errors

Error Type	FE Group			NE Group			p
	Number of Errors Occured	Number of Errors Recovered	Error Recovery Rate	Number of Errors Occured	Number of Errors Recovered	Error Recovery Rate	
Leaving out data items	52	19	36.54%	33	14	42.42%	.54
Wrong conditions	31	15	48.39%	33	6	18.18%	.04
And/or confusion	20	8	40%	19	9	47.37%	.65
Leaving out conditions	9	1	11.11%	8	1	12.5%	-
Others	4	0	0%	2	0	0%	-
Total	116	43	37.07%	95	30	31.58%	.82

based on the cause of retrieving wrong data. The frequencies of each error type and the recovery rates are summarized in <Table 4>.

The most common error found in this experiment was leaving out a data item required by the question. For example, some subjects retrieved product numbers when product names were requested, and others retrieved individual quantities when the sum of the individual quantities was required. The FE group made 52 errors of this type, and 19 of them were recovered in subsequent tries, while the NE group recovered 14 of 33 errors committed. There was no significant difference between the two groups in terms of the error recovery rate. In interpreting this result, it might be argued that the NE group could recognize the missing data items from the output data as easily as the FE group.

A second type of error was using a wrong condition for data retrieval. For example, some subjects retrieved products with a retail price of $19 when the question requested products with a retail price equal to or greater than $19. The FE group was significantly better in recovering this type of error than the NE group (48.4% vs. 18.2%, p=.04). The feedback echo seemed to reduce miscommunication between the system and the user for this type of error. This was especially true when the subject used a pronoun in a query. For example, when a subject followed a query relating to a customer named James Lee (e.g. SHOW ME THE ITEMS PURCHASED BY THE CUSTOMER JAMES LEE) with the query

> SHOW ME THE DEPARTMENT NAME
> AND ITS QUOTA FOR ITEMS WITH
> RETAIL PRICE=99.95

the INTELLECT system responded with the feedback echo

> PRINT THE ITEM NUMBER, ITEM DESCRIPTION, DEPARTMENT NAME
> AND QUOTA OF ALL RECORDS WITH
> CUSTOMER NAME='JAMES LEE'
> AND RETAIL PRICE=99.95.

The INTELLECT system interpreted ITS QUOTA in user's query as QUOTA OF ALL

RECORDS WITH CUSTOMER NAME='JAMES LEE' because it interpreted the pronoun ITS as the search condition of the previous query issued by the same user. Because the above feedback echo was not provided for the NE group subjects, they had more difficulty interpreting the result.

As another example of misinterpretation, INTELLECT processed the query

SHOW ME DEPARTMENT NAMES, NUMBERS SOLD BY DEPARTMENT,
AND THE QUOTA ASSIGNED FOR ITEM

as

PRINT THE DEPARTMENT NAME, QUOTA AND ITEM NUMBER OF ALL
RECORDS WITH DEPARTMENT NAME= NUMBERS.

It was much easier for subjects to detect this kind of miscommunication when the feedback echo was provided.

Other common errors occurred in writing AND/OR queries. For example, when the question requested the data about customers whose bill was above $1,000 and the customers whose bill was below $200, many subjects formulated the search condition as

BILL > 1000 AND < 200.

This was not correct because customer's bill cannot be greater than 1000 and, at the same time, less than 200. The condition should be stated as

BILL > 1000 OR < 200.

The feedback echo was not effective in facilitating recovery from this type of error. In fact, this type of error can be prevented by employing a simple artificial intelligence program to detect the error and convert an inappropriate AND to an OR automatically. At least one RNL system, CLOUT (1986), provides an automatic AND/OR conversion function.

Another type of error found in this experiment was leaving out a necessary condition. Both the FE and NE groups were ineffective in recovering from this type of error (11.1% vs. 12.5%). Most likely, the subjects in both groups who made this type of error did not fully comprehend the question, and the feedback echo did not facilitate better comprehension of questions for the FE group.

Finally, miscellaneous errors such as mistakes in an operator or a calculation were found, but the feedback echo was not helpful here either. Overall, there was no significant difference between the FE group (37.1%) and the NE group (31.6%) in terms of the error recovery rate. However, the feedback echo was effective in recovering errors caused by using the wrong conditions for data retrieval.

Ⅶ. Discussion and Conclusions

This study examined the effects of feedback echo of an RNL system on casual user performance. This section provides a general discussion of findings and suggestions for future research, as well as limitations of the study.

7.1 Discussion of Findings

From the results of this study, it can be concluded that the particular type of feedback echo available in this commercial system was not effective in terms of improving the overall performance of casual users. It is worth to mention, however, that the feedback echo was effective for one specific type of error: using the wrong conditions for data retrieval. These errors often occurred when the RNL system misinterpreted a pronoun in the query or made a seemingly strange interpretation of the query. In these cases, the feedback echo reduced the gulf of evaluation by permitting the user to compare the system's interpretation with the user's original goals and intentions.

For these errors using the wrong conditions for data retrieval, the difference of 30% in the error recovery rate between the FE and NE groups might be explained by the evaluation semantic distance. The FE subjects were provided information about how the system interpreted the users' query, so the user did not have to speculate why he/she received unexpected results, making the evaluation semantic distance between user and system much shorter.

While the feedback echo contributed to reducing semantic distance, the echo did not seem to have much impact on the articulatory distance. Evaluation articulatory distance concerns the relationship between the meanings of expressions and their physical output form. About half of the errors using the wrong conditions were not recovered even though the FE subjects received the feedback echo containing helpful information about the system's interpretation. These subjects may have ignored the feedback because they could not understand it; thus, a better method of presenting a feedback echo should be developed. Owei and Higa (1994) have suggested an interesting paradigm for natural language explanation of database queries employing a semantic data model approach. Another possibility is the development of a graphical or form-based display which can provide more articulatory directness than a structure query, even though both contain the same semantic information. For example, Adam et al. (1994) have described a methodology for processing data retrieval and update queries using a form-based natural language interface.

For other types of errors, the feedback echo did not improve the performance of casual users, at least statistically. However, most of these errors (e.g., leaving out a data item required by the question or leaving out a necessary condition) seemed to occur because the users did not comprehend the question, not because users had problems with the system. Those errors that occurred in writing AND/OR queries can be prevented by using a simple artificial intelligence program, as explained above. Considering all these factors, the feedback echo does seem to improve casual users' performance to some extent, although there is certainly room for further improvement in terms of reducing the evaluation articulatory distance.

7.2 Suggestions for Future Research

The results of this study suggest that more focused research into the types of errors related to the feedback echo is needed. Prior to this, however, a formal classification of error types in query wring should be developed for RNL. Previous studies (Thomas and Gould, 1975; Reisner et al, 1975) classified errors for the structured query language, but these classifications can not be directly used because the errors occurring in RNL systems are quite different from those in structured query language systems.

Further research is also required to develop a more effective feedback echo for an RNL database interface. What kinds of information should be presented in a feedback echo to reduce semantic distance? In what form should this information be presented to shorten articulatory distance? For example, the concept suggested by Owei and Higa (1994) should be empirically tested.

Another research issue relates to individual differences. There have been significant debates on the usefulness of individual difference studies (Huber, 1983). However, the performance differences among individuals in this study were too large to ignore. What attributes affect casual users' performance in querying RNL systems? Also this study can be replicated with more sophisticated users conducting more advanced tasks.

Even though this experiment provided strong internal validity, its external validity is limited and the results should not be overgeneralized. Among the factors that may influence external validity are the use of student subjects in a laboratory setting, the adoption of only the query writing task, and the use of only the INTELLECT feedback echo.

〈참 고 문 헌〉

[1] Adam, N.R., Gangopadhyay, A. and Clifford, J. "A Form-Based Approach to Natural Language Query Processing," *Journal of MIS*, Vol. 11, No.2, Fall 1994, pp. 109-135.

[2] Artificial Intelligence Co. *INTELLECT Query System Reference Manual*, Artificial Intelligence Co., Boston, Mass., 1982.

[3] Barbary, C. "A Database Primer on Natural Language," *Journal of Systems Management*, Vol. 38, No. 4, April 1987, pp. 20-25.

[4] Batra, D. and Sein, M.K. "Improving Conceptual Database Design Through Feedback," *International Journal of Human-Computer Studies*, Vol. 40, 1994, pp. 653-676.

[5] Campbell D.T. and Stanley J.C. *Experimental and Quasi-Experimental Designs for Research*, Houghton Mifflin Co., Boston, Mass., 1963.

[6] Capindale, R.A. and Crawford, R.G. "Using a Natural Language Interface with Casual Users," International Journal of Man-Machine Studies, Vol. 32, 1990, pp. 341-362.

[7] Date, C.J. *An Introduction to Database Systems* (6th ed.), Addison-Wesley, Reading, Mass., 1995.

[8] Dekleva, S.M. "Is Natural Language Querying Practical?" Data Base, Vol. 25, No. 2, May 1994, pp. 24-36.

[9] Ein-dor, P. and Spiegler, I. "Natural Language Access to Multiple Databases: A Model and a Prototype," Journal of MIS, Vol. 12, No. 1, Summer 1995, pp. 171-197.

[10] Greenblatt, D. and Waxman, J. "A Study of Three Database Query Languages," in

Databases: Improving Usability and Responsiveness, B. Schneiderman (ed.), Academic Press, New York, N.Y., 1978.

[11] Guinan, P.J. *Patterns of Excellence for IS Professionals: An Analysis of Communication Behavior*, ICIT Press, Washington, D.C., 1988.

[12] Hogarth, R.M., Gibbs, B.J., McKenzie, C.R.M., and Marquis, M.A. "Learning From Feedback: Exactingness and Incentives," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 17, No. 4, 1991, pp. 734-752.

[13] Huber, G.P. "Cognitive Style As a Basis for MIS and DSS Designs: Much Ado About Nothing?" *Management Science*, Vol. 29, No. 5, May 1983, pp. 567-582.

[14] Hutchins, E.L., Hollan, J.D., and Norman, D.A. "Direct Manipulation Interfaces," *Human-Computer Interaction*, Vol. 1, 1985, pp. 311-338.

[15] Hutchins, E.L., Hollan, J.D., and Norman, D.A. "Direct Manipulation Interfaces," in *User Centered System Design: New Perspectives on Human-Computer Interaction*, D.A. Norman and S.W. Draper (ed.), Lawrence Erlbaum Associates, Publishers, Hillsdale, N.J., 1986.

[16] Microrim Inc. *R:BASE CLOUT User's Manual*, Redmond, Wash., 1986.

[17] Molich, R. and Nielsen, J. Improving a Human-Computer Dialogue, *Communications of the ACM*, Vol. 33, No. 3, March 1990, pp. 338-348.

[18] Napier, H.A., Lane, D.M., Batsell, R.R., and Guadango, N.S. "Impact of a Restricted Natural Language Interface on Ease of Learning and Productivity," *Communications of the ACM*, Vol. 32, No. 10, October 1989, pp. 1190-1198.

[19] Napier, R.W. and Gershenfeld, M.K. *Groups: Theory and Experience*, Houghton Mifflin Co., Boston, Mass., 1973.

[20] Norman, D.A. "Cognitive Engineering," in *User Centered System Design: New Perspectives on Human-Computer Interaction*, D.A. Norman and S.W. Draper (ed.), Lawrence Erlbaum Associates, Publishers, Hillsdale, N.J., 1986.

[21] Owei, V. and Higa, K. "A Paradigm for Natural Language Explanation of Database Queries: A Semantic Data Model Approach," *Journal of Database Management*, Vol. 5, No. 1, Winter 1994, pp. 18-30.

[22] Payne, R.B. and Hauty, G.T. "Effect of Psychological Feedback upon Work Decrement," Journal of Experimental Psychology, Vol. 50, No. 6, December 1955, pp. 343-351.

[23] Reisner, P. "Use of Psychological Experimentation As an Aid to Development of a Query Language," *IEEE Transactions on Software Engineering*, SE-3, May 1977, pp. 218-229.

[24] Reisner, P. "Human Factor Studies of Database Query Languages," *Computing Surveys*, Vol. 13, No. 1, March 1981, pp. 13-31.

[25] Reisner, P., Boyce, R.F., and Chamberlin, D.D. "Human Factors Evaluation of Two Database Query Language--SQUARE and SEQUEL," in *Proceedings of the National Computer Conference*, Vol. 44, AFIPS Press, Reston, Va., 1975, pp. 447-452.

[26] Slator, B.M., Anderson, M.P., and Conley, W. "Pygmalion at the Interface," *Communications of the ACM*, Vol. 29, No. 7, July 1986, pp. 599-604.

[27] Stevens, L., "Getting Data in Plain English," *Computer Decisions*, Vol. 18, No. 9, April

22, 1986, pp. 42-47.
Stone, E. *Research Methods in Organizational Behavior*, Scott, Foresman and Company, Glenview, Ill., 1978.

[28] Stonebraker, M. "User Interfaces," in *Readings in Database Systems*, M. Stonebraker (ed.), Morgan Kaufmann, San Moateo, Calif., 1988.

[29] Suh, K.S. "A Comparison of Linear Keyword and Restricted Natural Language for Data Retrieval," Ph.D. Dissertation, Indiana University, 1989.

[30] Suh, K.S. and Jenkins, A.M. "A Comparison of Linear Keyword and Restricted Natural Language Data Base Interfaces for Novice Users," *Information System Research*, Vol. 3, No. 3, September 1992, pp. 252-272.

[31] Thomas, J.C. and Gould, J.D. "A Psychological Study of Query by Example," in *Proceedings of the National Computer Conference*, Vol. 44, AFIPS Press, Reston, Va., 1975, pp. 439-445.

[32] Turner, J.A., Jarke, M., Stohr, E.A., Vassiliou, Y., and White, N.H. "Using Restricted Natural for Data Retrieval--A Plan for Field Evaluation," in Human Factors and Language Interactive Computer Systems. Y. Vassiliou (ed.), Ablex, Norwood, N.J., 1984.

[33] Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (2nd ed.), Addison-Wesley, Reading, Mass., 1992.

[34] Wing, K.T. "Implications of Feedback Research for Group Facilitation and the Design of Experimental Learning," *Small Group Research*, Vol. 21, No. 1, February 1990, pp. 113-127.

Appendix: Typical Examples of Feedback Echoes

User's query: show me the department names and quota for the departments where retail price=99.95.

Feedback echo: PRINT THE DEPARTMENT NAME AND QUOTA OF ALL RECORDS WITH RETAIL PRICE=99.95.

User's query: give me the sum of onhand quantity whose retail price > =$19.

Feedback echo: PRINT THE TOTAL ONHAND QUANTITY OF ALL ITEMS WITH RETAIL PRICE > =19.00.

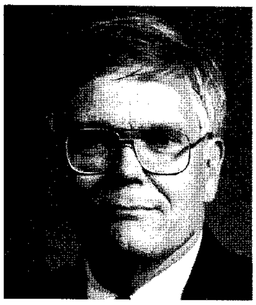
User's query: show me the description of the item and (retail price×1.3) where onhand quantity < .5×reorder quantity.

Feedback echo: PRINT THE ITEM DESCRIPTION, ITEM NUMBER, (RETAIL PRICE×1.3), ONHAND QUANTITY AND (.5×REORDER QUANITTY) OF ALL ITEMS WITH ONHAND QUANITTY < (.5×REORDER QUANITTY).

◆ 이 논문은 1998년 8월 24일 접수하여 1999년 1월 18일 게재확정되었습니다.

◆ 저자소개 ◆

서길수(Kil Soo Suh)
Kil Soo Suh is Associate Professor of Information Systems at Yonsei University, Seoul, Korea. Professor Suh holds a Ph.D. in Management Information Systems from Indiana University where he also earned his M.B.A. He holds a B.A. in Business Administration from Yonsei University. His research interests are in the areas of man-machine interfaces, communication media, internet marketing and conceptual modeling and implementation of information.



William C. Perkins
William C. Perkins is Professor and Coordinator of Information Systems in the Kelley School of Business at Indiana University-Bloomington. He received the B.S.C.E. degree from Rose-Hulman Institute of Technology and the M.B.A. and D.B.A. degrees (Quantitative Business Analysis) from Indiana University. Professor Perkins has published extensively in the areas of systems simulation, assembly line design, decision support systems, and negotiation support systems. He is a Fellow and Past President of the Decision Sciences Institute as well as Council Member of the Association for Information Systems.