



IDENTIFYING SALIENT USABILITY ATTRIBUTES FOR AUTOMATED TELEPHONE SERVICES

Stephen Love[†], R T Dutton[†], J C Foster[†], M A Jack[†] and F W M Stentiford[‡]

[†]Department of Electrical Engineering, The University of Edinburgh, 80 South Bridge, Edinburgh.

[‡]BT Laboratories, Martlesham Heath, Ipswich

ABSTRACT

This paper presents experimental results relating to the perceived usability of automated telephone services which incorporate a spoken language dialogue interface. Using a new Wizard of Oz experimental technique, users' attitudes towards automated telephone services are measured by a Likert attitude questionnaire. Factor Analysis performed on subjects' responses to the questionnaire produced a set of five underlying group factors which accounted for 74% of the variance in the questionnaire scores. On the basis of the factor analysis, a multiple regression model was constructed using the highest loading attributes on each of the five factors. The predictive power of this model was assessed.

I. INTRODUCTION

Users' attitudes are an important consideration in the design and implementation of automated telephone services which incorporate a spoken language dialogue interface. As a result, principles of usability engineering [1] are increasingly being employed in the design of new products and services. Fundamental to the application of these principles is the identification of salient usability attributes and user perceptions, work which forms part of an on-going programme of research at the University of Edinburgh into the usability evaluation of automated telephone services. This research programme uses a Wizard of Oz experimental methodology [2,3] which is distinguished from most previous versions of this technique by the degree of control the software provides over the experimental variables. In particular, speech recognition errors are handled automatically by the use of realistic simulations of both isolated and connected word recognisers [4] and do not as a result require the wizard's active intervention.

The facilities at Edinburgh allow experimentation with different tasks, different dialogue strategies and

recognition interfaces with different input characteristics. During experiments, objective and subjective measures of users' responses are made. *Objective measures* include the overall duration of the interaction and the number of speech recognition and user errors, where user errors include such features of the interaction as the number of times the user speaks over the service messages or tone prompts. *Subjective measures* include users' attitudes towards specified service attributes and evaluation of their own responses to the services. Attitude questionnaires specifically designed at the University for the evaluation of automated telephone services are used for the collection of subjective data.

This paper reports the results of a factor analysis carried out on the responses to an attitude questionnaire completed during one of the project experiments. The aim of this particular analysis was to identify the underlying factor structure of users' responses across a wide range of attributes and perceptions. The analysis yielded a small set of group usability factors from which a regression model was derived for the evaluation of user attitudes on the basis of a reduced set of salient attributes and responses. The paper also reports an evaluation of the predictive power of the model.

II. THE QUESTIONNAIRE

From pilot studies and literature reviews [5,6] carried out in the early stages of the research programme, a broad set of 18 service attributes and user responses important for the evaluation of automated telephone services were identified. Service attributes included *ease of use*, *perceived reliability* and *efficiency*. User responses included the *perceived friendliness* of a service, the *degree of control* over the service users felt they had during their interaction and the *degree of frustration* experienced while using a service. These service attributes and user responses

were initially selected to cover the largest number of usability dimensions relevant to dialogue interfaces generally. In addition, four features specific to voice activated telephone services were also identified. These included the perceived *quality of the spoken messages* and *attitudes to the tone prompts* used by the service. In total, 22 'core' usability features were identified and selected for inclusion in the initial attitude questionnaire.

The questionnaire was designed as a randomised sequence of stimulus statements, one for each of the 22 features, with associated 7-point Likert response scales as shown in Figure 1.

The service was easy to use

1 2 3 4 5 6 7

where

- 1 = strongly agree
- 2 = agree
- 3 = slightly agree
- 4 = neutral
- 5 = slightly disagree
- 6 = disagree
- 7 = strongly disagree

Figure 1
An Example Likert Attitude Statement and Scale

III. A USABILITY EXPERIMENT

One of the experiments carried out as part of the research programme used subjects who had no previous experience of voice activated automated telephone services. During the experiment, each subject completed one task from a simulated service viz. paying for goods or services over the telephone by reading out a credit card number one digit at a time. The speech interface used a dialogue scheme appropriate for isolated word input with the digit recognition accuracy for the simulated recogniser used in the Wizard of Oz workbench set at 95%. Following their use of the service, each subject completed the 22-statement attitude questionnaire described above.

A total of 61 subjects completed the experiment. For the purposes of analysis they were randomly divided into two groups:

- A total of 40 subjects whose responses were used for the factor analysis and the construction of a multiple regression model;
- A total of 21 subjects whose responses were used to test the predictive power of the regression model.

IV. FACTOR ANALYSIS AND A USABILITY MODEL

Factor Analysis [7] is a statistical method which aims to find hypothetical constructs to explain the relationship between subjects' scores on tests or subtests. The procedure investigates the underlying structure of a correlation matrix in order to identify "clusters" of variables which are statistically related. Factor analysis does, however, require the intuitive naming and interpretation of the factors which have been statistically identified. Employing factor analysis as a form of construct validity provides evidence which allows us to claim, in this particular instance, that the usability of automated telephone services can be organised in a particular way.

A factor analysis was performed on the responses of the first group of 40 subjects to the 22 attitude questions using the principal components method. The analysis produced a set of five factors which accounted for 74% of the variance in the 22 variables. The factor matrix was rotated using the standard Varimax rotation.

In order to assist in the interpretation of the factors, the Burt-Banks formula [8] was used to test the significance levels of the derived factor loadings. This method, which is one of several criteria available for testing the significance of factor loadings, is based on the assumption that as one moves from the first factor to higher factors there should be an increase in the value of the factor loading for it to be considered significant. By doing this, the authors claimed, one would be able to take account of the increase in unique variance (which is a combination of specific variance and error variance) in higher factors. In addition, this method is considered to be especially suited to small samples. Table 1 shows the salient loadings for each of the five factors based on this criterion.

Factor	Salient Loadings
1	Efficiency of the service Reliability of the service Improvement required to the service Willingness to use the service again User's degree of enjoyment User's degree of frustration Whether the service was confusing to use Degree of complication of the service Ease of use of the service Degree of control user had over the service
2	Speed of the service Stressfulness of the experience of using the service Concentration required to use the service Whether the user felt flustered while using the service Degree of control user had over the service Whether the service was felt to be confusing to use User's degree of frustration User's degree of enjoyment
3	User's attitude to the voice used by the service User's attitude to the tones used in the service The perceived friendliness of the service Whether user preferred the service or a human operator The perceived helpfulness of the service tones User's degree of enjoyment Degree of complication of the service
4	The perceived clarity of the voice used the perceived politeness of the service Whether users knew what they were expected to do Degree of complication of the service
5	Ease of use of the service The perceived helpfulness of the service tones Whether the user felt flustered while using the service Whether users knew what they were expected to do Degree of complication of the service User's attitude to the tones used in the service Whether the service was confusing to use

Table 1: Salient Factor Variables

The first factor, which accounted for 21% of the variance explained by this analysis, was interpreted as representing the *quality of interface performance*. Responses which loaded highly significantly on this factor included the perceived efficiency and reliability of the service, the amount of improvement judged by

the user to be required and willingness to use the service again.

A second factor, representing *cognitive effort and stress experienced by the user*, accounted for a further 17% of the sample variance and was associated with attitudes such as the perceived speed of service, the stressfulness of the experience, the amount of concentration required by the user, the degree to which users felt flustered and the degree of perceived control over the interaction.

A third factor, *user's conversational model*, was associated with attitude to the voice used for the service messages, attitude to the tone prompts and the perceived friendliness of the service. The factor relates to the extent to which the experience of human-computer conversation resembles the users' implicit model of human-human conversational behaviour which encompasses both task and social goals.

The fourth and fifth factors extracted by the analysis were closely related. They both had to do with the perceived ease of use of the service. They differed insofar as the fourth factor was associated with the clarity of the voice, the perceived politeness of the service and whether users felt they knew what was expected of them. This factor relates to the perceived *fluency of the experience*. The fifth factor, on the other hand, was associated with the ease of use of the service, the helpfulness of the tone prompts and the degree to which users felt they were flustered while using the service. This factor has been labelled the *transparency of the interface*.

The factor analysis therefore demonstrated the presence of a clear underlying structure to users' responses in which perceived quality of interface performance, cognitive loading, the user's conversational model, the perceived fluency and transparency of the interface were salient characteristics of users' responses. This factor structure is summarised in Figure 2.

A Multiple Regression analysis was carried out taking the highest loading variable on each of the five derived factors as the independent predictor variables and using the overall mean scores on the original 22-set questionnaire as the dependent criterion measure. The estimated model accounted for 86% of the variance in the responses to the full questionnaire ($R^2 = .859$).

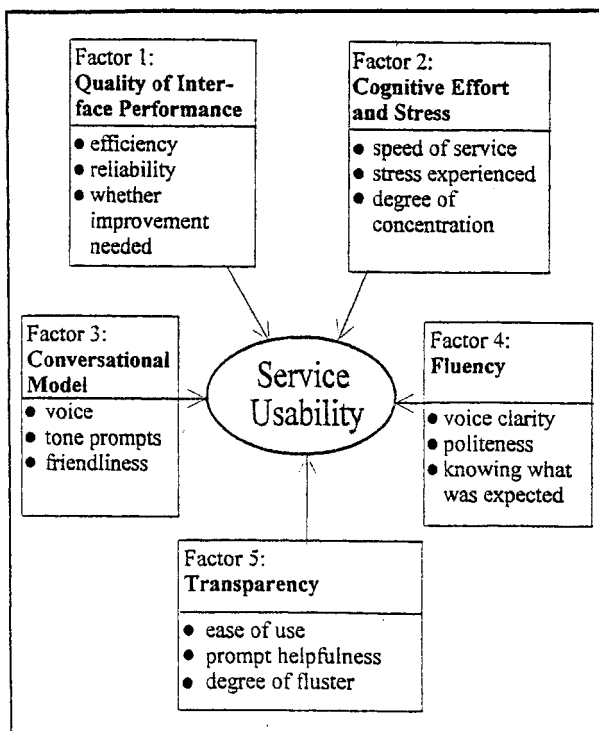


Figure 2: Service Usability Factor Structure

In order to obtain an estimate of the predictive power of this model, the regression coefficients were used to write a standard regression equation to obtain an overall measure of service usability using the set of values obtained from the five predictor variables for the 21 subjects in the second group. A comparison, using the Pearson Product-Moment Correlation, indicated a high correlation ($r = .874$) between the predicted scores using the model and the actual scores obtained from the full Likert questionnaire. This result confirms the predictive power of the model which has emerged as a result of the factor analysis.

V. CONCLUSIONS

The results from this preliminary experimental work indicate that it is possible by using factor analysis to decompose the overall concept of usability in relation to automated telephone services into significant sub-categories. In addition, the results of the multiple regression in conjunction with factor analysis have provided information about the possibility of reducing the number of questions contained in the Likert questionnaire while retaining a valid measure of usability.

VI. REFERENCES

- [1] Nelson, J. *Usability Engineering*, Academic Press, Boston, MA, 1993.
- [2] Foster, J. C., R. Dutton, S. Love, I. Nairn, N. Vergeynst and F. W. Stentiford, "Intelligent Dialogues in Automated Telephone Services". In Baber, C. And Janet M. Noyes (eds), *Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers*, Taylor and Francis, 1993, pp. 167 - 175.
- [3] Fraser N. M. and G. N. Gilbert 'Simulating Speech Systems'. In *Computer, Speech and Language*, 5, pp. 81 - 99 (1991)
- [4] Jack, M. A., J. C. Foster and F. W. Stentiford, "Intelligent Dialogues in Automated Telephone Services". In Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta, Canada, October 1992.
- [5] Poulson, D. 'Towards Simple Indices of the Perceived Quality of Software Interfaces'. In IEE Colloquium - Evaluation Techniques for Interactive System Design. IEE, Savoy Place, London (1987)
- [6] International Standards Organisation, 'Ergonomic Requirements for Office Work and Visual Display Terminals (VDTs)', ISO CD 9241-11 (1990)
- [7] D Child, "The Essentials of Factor Analysis", Cassell Education Ltd, (1990)
- [8] C Burt, "Tests of significance in factor studies", *British Journal of Psychology: Statistics Section*, 5, 109-133, (1952)