

TeleMorph: A Fuzzy Logic Approach to Network-Aware Transmoding in Mobile Intelligent Multimedia Presentation Systems

Anthony J. Solon, Paul McKeivitt, and Kevin Curran

Abstract—Mobile intelligent multimedia presentation systems are limited by various constraints including mobile network characteristics, mobile device capabilities, and user preferences. Those presentation systems which incorporate remotely stored multimedia rely on the bandwidth which is available during actual content transmission on the connecting mobile network. One approach to deal with this is to transcode content, thus reducing its data rate requirement, although this technique is inherently limited by the lowest acceptable quality of that media element. Alternatively, content can be transmoded to different modalities with lower bandwidth requirements. TeleMorph, a cross-modality adaptation control platform is detailed in this paper. The main premise of TeleMorph is that cross-modality adaptation decisions in mobile presentation systems must occur with primary consideration for bandwidth fluctuations. TeleMorph has been implemented as a network-aware fuzzy inference system that controls cross-modality adaptations between multimodal audio-video, audio-images, and images-text presentations, as well as unimodal audio and text presentations. Initially, a brief introduction to Intelligent Multimedia and Mobile Intelligent Multimedia is given, and related systems discussed. TeleTuras, a tourist information application implemented as a testbed for TeleMorph, is utilized for objective and subjective evaluations through its integration of six disparate test scenarios that incorporate various media types and qualities. Positive results for each evaluation technique, based on specific metrics, are also presented. In addition, future work on TeleMorph is also detailed.

Index Terms—Bandwidth awareness, cross-modality adaptation, fuzzy logic, mobile intelligent multimedia presentation systems, mobile multimedia transmoding, mobile network constraints, multimodal output, TeleMorph, TeleTuras.

I. INTRODUCTION

MOBILE networks impose various constraints on connected mobile multimedia presentation systems including fluctuating bandwidth, packet loss, transmission latency, bit error rate, and unpredictable disconnections. These constraints vary on mobile networks as they do with all networks, but as a result of the intrinsic ability of mobile client

devices to maintain their mobility these constraints vary indeterminably, based on the device's proximity to the associated network access points and the resulting signal strengths. Mobile multimedia presentation systems which stream and download media content over such networks, from multimedia repositories such as streaming servers and web servers, rely on the availability of sufficient bandwidth to transmit this content. In the absence of adequate bandwidth, multimedia systems must adapt the composition of presentations in order to maintain an acceptable level of output quality for end users, and thus sustain the presentation's usability. Enabling ubiquitous access to rich multimedia content on any type of terminal connected across heterogeneous networks in this way is referred to as Universal Multimedia Access (UMA) [20]. UMA also considers user preferences and techniques for maintaining an acceptable quality of service for the user (QoS) [5], [6].

The main approach used in adapting multimedia presentations in response to varying presentation resources, such as network conditions, is to transcode or scale multimedia to multiple fidelities [8], [10], [11], [12] either before transmission or in real-time during transmission. These transcoding methods have storage and computing intensive implications, respectively. Transcoding allows a system to select an appropriate media fidelity based on monitored constraints during live transmission. However, as multimedia is distilled to meet constraints the quality of the content is degraded, as is the perceived usability of this content. Transcoding in this way is limited to the inherent lowest acceptable quality of any media element. Despite the content's data rate corresponding to that which is available on the host mobile network, this distilled fidelity may be below the media's lowest usable quality, which would result in an unacceptable presentation with insufficient usability.

An alternative to multimedia distillation or transcoding strictly within modalities is to transmode, or adapt multimedia content across modalities based on semantic similarities, thus portraying essentially the same information using alternative media. Transmoding enables presentation systems to replace media elements (e.g., video file), or fragments of media, with alternative media elements, or media fragments composed in different modalities (e.g., audio) [4]. When adapting across modalities the semantics of the presentation must be maintained seamlessly to ensure its usability, therefore suitable semantically equivalent media alternatives must be determined either dynamically or prior to adaptation. As the current goal of TeleMorph is not to create a multimedia document model

Manuscript received November 3, 2006; revised February 17, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Yucel Altunbasak.

The authors are with the School of Computing and Intelligent Systems, Faculty of Engineering, University of Ulster, Magee Derry, Northern Ireland, BT48 7JL (e-mail: A.J.Solon@ulster.ac.uk; P.McKeivitt@ulster.ac.uk; K.J.Curran@ulster.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2007.901521

for semantically representing content semantics, but to control cross-modality adaptations given these semantics, currently only manually generated media alternatives in various modalities and subsequent fidelities are utilized. It is, however, the eventual goal of TeleMorph to integrate a multimedia document model to represent content semantics. Currently within TeleMorph one high quality video file (audio-visual) acts as the root source of content from which alternatives are derived including: speech audio, image sequences from video key frames and a text transcript, as well as combinations of these including: audio-images, and images-text. This content facilitates controlled evaluations of TeleMorph’s primary objective: to efficiently and effectively control cross-modality adaptations based on fluctuating bandwidth. Multiple-fidelity multimedia content has been created manually in TeleMorph so that as the data rate for a media element increases, so too does the semantic quality/value of the media.

A major issue to consider with cross-modality adaptations is the effect it has on the end-user as regards maintaining a presentation’s semantics as well as the presentation’s seamless information flow. The problem with frequent cross modality adaptations within one presentation is that the end-user’s attention may be distracted from the actual information being portrayed to the actual modality modifications occurring. Thus, the challenge in developing a system to control blindfolded media adaptations lies in the ability of the system’s logic to determine *when* the system should perform cross modality adaptations and *what* quality of multimodal or unimodal presentation is selected as the most suitable output. The system must also maintain an awareness of how often it allows these adaptations to occur on-the-fly and what modifications these consist of specifically. Specifically, the primary challenges in controlling cross-modality adaptations include the following.

- 1) Determining whether to transcode media further or transmode across modalities. This issue relates to the quality of the current output presentation and whether usability of the presentation would better maintained by further distilling the current media type, thus risking the possibility of over-transcoding the content, or by adapting (possibly across-modality) to a presentation composed of alternative media types. It should be noted that reducing the number of cross-modality adaptations, by distilling content to a lower quality than alternative media types, may be preferred than adapting across-modality despite the resulting output presentation quality. Consideration of this tradeoff is something which is considered inherently as a result of TeleMorph’s implementation using fuzzy logic.
- 2) Determining the most effective and acceptable destination media types following adaptation, and the quality/fidelity of this new content. This issue analyzes the composition of the presentations which TeleMorph adapts to, and if these transitions maintain the information flow and continuity of the output presentation (e.g., maintaining continuous media types following adaptations if used before hand). Also this issue is concerned with how well TeleMorph adheres to Cognitive Load Theory and Dual Coding Theory (i.e., always adapting to the most cognitively effective output modality combinations) as discussed in [9],

[17], and [18], respectively. Evaluations using this metric will qualify whether adaptations incorporated substantial changes in the output presentation’s quality.

- 3) Cautiously controlling the timing and frequency of cross-modality adaptations. This challenge ensures that cross-modality adaptations maximize the potential quality of output presentations based on available bandwidth, whilst also minimizing the affect the frequency of these adaptations has on usability and user acceptance. The acceptable frequency of cross-modality adaptations varies based on the current and destination modalities involved in an adaptation.

The effects of cross-modality adaptation between both discrete and continuous (temporal) multimedia elements must be considered before morphing presentations across modalities, as should the effectiveness and user-acceptance of the output presentation. As discovered in [9], when a system presents data on a mobile device which must be remembered (e.g., medical surgery training) the most effective presentation mode should be used (picture and speech) which does not cognitively overload the user. When the system simply has to inform the user (e.g., a tourist information system) the most appealing/accepted presentation mode should be used (picture, text, and speech). These points which are based on the Cognitive Load Theory (CLT) [2], [22] should be incorporated into multimodal presentation systems to achieve ultimate usability.

In Section II, an overview of transmoding in mobile intelligent multimedia presentation systems is given. Section III describes TeleMorph and how it has been implemented as a network-aware fuzzy inference system, with subsections detailing inputs, rules, and outputs. TeleTuras, a tourist information application, which has been developed as a testbed for TeleMorph using specific test scenarios, is detailed in Section IV. Following this, Section V discusses objective and subjective evaluations of TeleMorph using TeleTuras and a disparate selection of media files with varying bandwidth requirements. Section VI concludes and proposes future work.

II. TRANSMODING IN MOBILE INTELLIGENT MULTIMEDIA PRESENTATION SYSTEMS

Substantial research has been carried out to address the specific issues which exist in the area of mobile intelligent multimedia presentation systems. Malaka [15, p. 5] states that “the main challenge for the success of mobile systems is the design of smart user interfaces and software that allows ubiquitous and easy access to personal information and that is flexible enough to handle changes in user context and availability of resources.” Some projects have already investigated mobile intelligent multimedia systems, using tourism in particular as an application domain including DEEP MAP [16], CRUMPET (Creation of User-friendly Mobile services Personalized for Tourism) [7], [29], SmartKom [26], and EMBASSI [13]. Though all of these systems focus on some aspect of mobile intelligent multimedia presentation, none have specifically addressed the issue of dynamically transmoding based on network constraints, and in particular available network bandwidth.

The majority of research that tackles the specific problem of cross modality adaptation has focused on multimedia document

models with inherent substitution or alternative semantics. Reference [4] proposed a cross-media adaptation and substitution model which focused on maintaining the semantics of a presentation's information flow using semantically equivalent media alternatives. In the application, Gallery of Cardiac Surgery (Cardio-OP) [14], multimedia document models are augmented with all the unimodal or multimodal presentation alternatives, thus facilitating adaptation. At run time alternatives are excluded based on presentation constraints. ZyX [3], Tiempo [27], and REVEAL THIS [19] are other examples of multimedia document models which provide some element of flexibility and adaptation. References [23]–[25] propose a content value model, called the Overlapped Content Value (OCV) model which is a systematic approach to determining the conversion boundaries between modalities based on resources available to the system, as apposed to actually controlling cross-modality adaptations. Their approach considers the fidelity of adapted content quality and modality, with respect to the original content. It has inherent temporal limitations in its ability to make intelligent dynamic adaptation decisions. The thresholds dictating the resource availability points at which adaptations should occur remain static once determined in the OCV model, thus preventing efficient and truly intelligent on-the-fly adaptations.

III. TELEMORPH

TeleMorph [21] performs network-aware content adaptation decisions for mobile intelligent multimedia presentation systems. More specifically it is a cross-modality adaptation control platform which bases adaptation decisions on constraints relevant in a mobile environment, with particular focus on bandwidth fluctuations. TeleMorph utilizes fuzzy logic to control these cross-modality adaptation decisions, which occur between multimodal combinations including audio-video, multimodal audio-images, multimodal images-text, unimodal audio, and unimodal text. TeleTuras, a tourist information application for the city of Derry in Northern Ireland, has also been implemented in order to provide an example testbed for TeleMorph. Through TeleTuras subjective and objective evaluations of TeleMorph's performance during dynamic cross-modality adaptations have been carried out. This was enabled using manually distilled multimedia of the multimodal combinations integrated into six disparate test scenarios. Using a Fuzzy Inference System (FIS) allows TeleMorph to provide rapid and intelligent responses to constantly fluctuating mobile network properties such as bandwidth and packet loss.

Essentially, TeleMorph is a black box decision engine allowing any intelligent multimedia presentation system to exploit its decision-making abilities. Thus, it is not dependent on any particular application domain or implementation, but instead, can provide cross-modality adaptation decisions to any presentation system given the required eighteen inputs. A mobile intelligent multimedia presentation system invokes the TeleMorph fuzzy inference decision system to determine what content to present in response to user queries and most importantly in response to bandwidth availability on the wireless/mobile network which connect the presentation system to its media repositories.

Zadeh [28] stated when speaking in reference to decision systems that "as complexity rises, precise statements lose meaning and meaningful statements lose precision." Fuzzy logic is used in the implementation of TeleMorph as it enables the development of a control system which relies on the meaningfulness, significance and consequence of decision outputs. When performing dynamic cross-modality adaptations in a live presentation system, priority lies in the control system's ability to maintain continuity of the output presentation and seamlessness during these adaptations whilst observing restrictions imposed by the constraints acting upon the system, as well as determining output fidelities for multimedia content. In accordance with this approach, it is more important that TeleMorph controls the frequency of adaptations and performs them in an intelligent and efficient manner than it is that the system chooses one precise output media-fidelity. It is determined that through the number of Membership Functions (MFs) describing the Universe of Discourse (UoD), or domain of possible values, for each input fuzzy set that a sufficiently meaningful level of granularity is enabled and that the degree of precision provided through the current implementation of fuzzy logic is sufficient to maintain a high performance of TeleMorph and its decision making ability.

TeleMorph's FIS was implemented using Matlab's Fuzzy Logic Toolbox [1]. This toolbox allows for the fast and efficient development of intelligent fuzzy inference systems. When developing a FIS there are various system properties which one must determine during design stages before implementation of the system can commence. In TeleMorph's case the FIS implements a Mamdani type fuzzy system which uses 'min' as the 'And' method, 'max' as the 'Or' method, 'min' for implication, 'max' for aggregation, and the 'Centroid' method for defuzzification. An overview of the inputs and outputs in TeleMorph are shown in Fig. 1. The implementation of TeleMorph described here incorporates network-aware multimedia adaptations between multimodal audio-video, multimodal audio-images, multimodal images-text, unimodal audio and unimodal text, and thus the inputs and outputs, as shown in Fig. 1, reflect this.

A. TeleMorph FIS Inputs

TeleMorph provides output variables including audio-video (v), audio-images (ai), images-text (it), audio (a), and text (t) detailing the quality at which each modality should be present in the output presentation given inputs including: currently available network bandwidth (bpsCurr), current network packet loss (pktLoss), adaptation frequency (adaptFreq), current video quality (vCurr), lowest available video quality (vLowest), highest available video quality (vHighest), current audio-images quality (aiCurr), lowest available audio-images quality (aiLowest), highest available audio-images quality (aiHighest), current images-text quality (itCurr), lowest available images-text quality (itLowest), highest available images-text quality (itHighest), current audio quality (aCurr), lowest available audio quality (aLowest), highest available audio quality (aHighest), current text quality (tCurr), lowest available text quality (tLowest), and highest available text quality (tHighest).

The decision system has 18 normalized inputs and five normalized outputs, as shown in Fig. 1. Using these 18 inputs

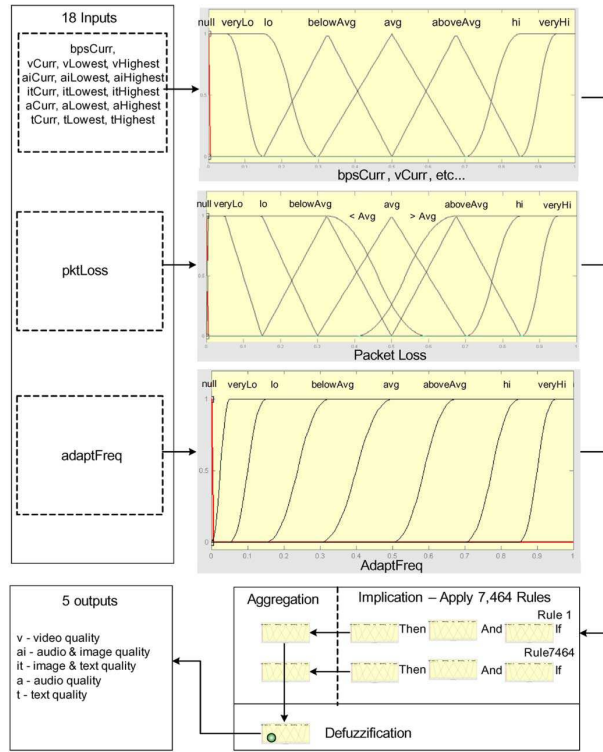


Fig. 1. Architecture of TeleMorph’s network-aware FIS.

the inference system’s rules determine whether an adaptation should occur, to what media types the system should adapt, and at what fidelity the new target media type should be used in the output presentation. The system can determine that no compositional transitions should be made to a presentation given certain parameters. This can be as a result of low packet loss or of an insufficient duration of time since the preceding cross-modality adaptation. The Membership Functions (MFs) in TeleMorph’s input fuzzy sets include ‘null’, ‘veryLo’, ‘lo’, ‘belowAvg’, ‘avg’, ‘aboveAvg’, ‘hi’, and ‘veryHi’. The fuzzy set representing packet loss (‘pktLoss’) has two additional membership functions (along with the aforementioned eight default MFs) including less than the average (‘<Avg’) and greater than average (‘>Avg’).

Membership Functions in TeleMorph are relative to the normalized Universe of Discourse (UoD) which is limited by the highest bandwidth (or data rate) required by the available multimedia files. The UoD adapts to represent the domain which includes multimedia qualities ranging from the highest bandwidth required (e.g., video requiring 309 Kbps or audio-images requiring 272 Kbps) to 0. This means that the membership functions always represent the maximum possible area within the UoD and thus TeleMorph always has optimum granularity in its decision making.

‘bpsCurr’ is the input fuzzy set, as shown in Fig. 1, that represents the normalized value for currently available bandwidth in the mobile network which is being used to transmit the output multimedia presentation, and represents 1 of the 2 inputs that facilitate network-awareness in TeleMorph. The ‘pktLoss’ input

fuzzy set encapsulates the percentage of data packets lost or dropped during transmission from the media repositories to the mobile client device for presentation. This value enables the FIS to adapt more reactively in response to increasing packet loss during a presentation, or more stringently as a result of 0% packet loss (no urgency to adapt if ‘pktLoss’ is 0). The shapes of the ten membership functions within this fuzzy set are depicted in Fig. 1. ‘pktLoss’ is used in various rules in TeleMorph in particular those that result in adaptation to a unimodal presentation. In these rules the percentage of packet loss must be very high before TeleMorph will adapt to a unimodal presentation, as this is considered to be a less efficient method of portraying information in accordance with the Cognitive Load [9] and Dual Coding theories [17], [18], respectively.

One of the principal limitations of cross modality adaptation is the effect these adaptations have on the flow and continuity of the output presentation, and thus its usability. TeleMorph was developed with this limitation in mind and consequently it controls adaptations in a prudent manner and maintains an awareness of their frequency to increase the efficiency of this control. The ‘adaptFreq’ input variable handles this aspect of adaptation decisions in the fuzzy inference system. Specifically, ‘adaptFreq’ is a combination of relevant factors affecting the frequency of cross modality adaptation: the time elapsed since the previous adaptation (‘adaptTimeElapsed’), and the number of adaptations carried out so far during the multimedia presentation (‘numAdapts’). The method used to combine ‘numAdapts’ and ‘adaptTimeElapsed’ to form ‘adaptFreq’ is based on expert knowledge about the application domain, as discussed in [21]. The resulting variable, ‘adaptFreq’, is an integrated value whose inclination towards adaptation increases as more time elapses following the preceding adaptation, and diminishes as more adaptations occur.

The ‘vCurr’ input fuzzy set represents the data rate of the video file which is currently part of the output presentation. ‘vCurr’ is normalized against the maximum bandwidth required by any potential presentation output in the system (e.g., 300 Kbps video or 450 Kbps combination of audio-images). Thus, ‘vCurr’ is mapped to a normalized value relative to the highest bandwidth data rate required by one of the multimodal/unimodal presentation types available to TeleMorph. If video is not currently included in the output presentation then ‘vCurr’ is set to ‘null’. ‘vLowest’ represents the data rate required by the lowest fidelity video file, whilst ‘vHighest’ is the bandwidth required by the highest fidelity/quality video file available to the presentation system. The values for the current, lowest and highest data rates required by audio-images, images-text, audio, and text presentation types, which are available to the presentation system, are represented by fuzzy sets where the phrases, ‘Curr’, ‘Lowest’ and ‘Highest’, are preceded by ‘ai’, ‘it’, ‘a’, and ‘t’, respectively. The design of these input fuzzy sets are shown in Fig. 1.

B. TeleMorph FIS rules

There are 7 464 rules within the TeleMorph fuzzy inference system, each of which is network-aware based on the ‘bpsCurr’ and ‘pktLoss’ fuzzy inputs. The high number of rules is as a result of the complex permutations of media qualities which must

be handled as prerequisites and conditions by the decision making system. Rules in TeleMorph are divided into ten main groups: rules controlling adaptation between video and audio-images, video and images-text, video and audio, video and text, audio-images and images-text, audio-images and audio, audio-images and text, images-text and audio, images-text and text, and adaptations between audio and text. Each of these ten main groups of rules is further broken down into four subgroups controlling increases and decreases in the output presentation quality between the two different multimedia presentation types that are present in that group (e.g., four subgroups consisting of: adaptations involving a quality decrease from video to audio and from audio to video, as well as those adaptations involving a quality increase from audio to video, and from video to audio). Rules have been designed to give greater weight to adaptations which not only maintain the highest output presentation quality without inducing further packet loss following the presentation transition, but also those rules that encapsulate adaptations which are as seamless and acceptable for the end user as possible. TeleMorph's rules succeed in complying with this general approach, but more specifically they adhere to the three key challenges outlined in Section 1 with respect to controlling cross-modality adaptations. The techniques that TeleMorph implements to accomplish this are detailed below in points 1) to 3).

- 1) The fuzzy rules use TeleMorph's 15 media-related inputs (e.g., vCurr, itLowest, tHighest) to consider the types of multimodal and unimodal media available to TeleMorph. Adaptations between continuous (e.g., video) and discrete media types (e.g., images-text), as well as linguistic (e.g., text) and nonlinguistic media types (e.g., images), are controlled stringently as they have the potential to distract and frustrate end users. For example, a transition between audio-images and images-text results in the linguistic element of the presentation changing between continuous and discrete media types. Despite the caution which must be taken before carrying out this type of cross-modality adaptation, the target presentation (e.g., images-text) may be considered to be highly effective in portraying information to the end user. This caution is manifested within the fuzzy rule base by having stricter prerequisite input parameters, such as a very low packet loss, for rules involving a transition between discrete and continuous media types. The reasoning behind these approaches is related to principles outlined within Dual Coding Theory and Cognitive Load Theory, as discussed in Section I. These theories encourage the use of multiple modalities, whilst avoiding redundancy of information across the modalities, to achieve greater levels of memory persistence and information retention.
- 2) Additionally, the fuzzy rules have been designed to only facilitate adaptations when the quality transition across the Universe of Discourse is substantial. For example, decreases in a presentation's quality from 'avg' quality video to 'veryLo' audio where bandwidth is also 'veryLo', or 'veryHi' video to 'avg' audio where bandwidth is also 'avg'. In instances where there is a large difference (e.g., two membership functions or more) between the values for current bandwidth and the quality of the current

media being presented (e.g., aiCurr, aiLowest, aiHighest) transmoding is preferred. Whereas, in cases where the values for current bandwidth and the current media being presented are close (e.g., less than two membership functions) further distillation or transcoding of a presentation's current content may be the optimal adaptation.

- 3) By designing rules based on the two aforementioned guidelines cross-modality adaptation frequency is inherently restricted within TeleMorph, which has the knock-on effect of reducing the impact cross-modality adaptations have on the flow and continuity of presentations, and thus usability. The third guideline which has directed the design of TeleMorph's rules details the ideal modality-specific timing and frequency of cross-modality adaptations, using the fuzzy input 'adaptFreq'. Explicitly maintaining an awareness of the timing and frequency of adaptations through 'adaptFreq' consequently enables TeleMorph to adhere to this guideline when performing transitions, based on the composition of the current and target presentations.

C. TeleMorph FIS outputs

The five output fuzzy sets represent the various combinations of modalities that are available to a presentation system (e.g., TeleTuras) employing TeleMorph's decision making capabilities. As TeleMorph deals with adaptations between video, audio-images, images-text, unimodal audio and unimodal text, five output fuzzy sets named 'v', 'ai', 'it', 'a', and 't' are used to communicate decisions regarding any transitions. Following the implication of the 7464 fuzzy rules in the FIS, as depicted in Fig. 1, output sets are created for each rule which have varying degrees of support based on the antecedents (inputs). Since decisions in any FIS are based on the testing of all the rules in the system, all rules must be combined in some manner in order to enable this. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable (i.e., 'v', 'ai', 'it', 'a', and 't').

The input for the defuzzification process in TeleMorph consists of the five aggregated output fuzzy sets, and the final output is five single numbers. The aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from each output fuzzy set. In TeleMorph the most popular defuzzification method, the centroid calculation which returns the centre of area under the curve, is employed to defuzzify the five output sets to give five crisp values for 'v', 'ai', 'it', 'a', and 't'. However, only one of TeleMorph's outputs is a whole number whilst the remaining outputs are always 0, e.g., 'v' = 0.812, 'ai' = 0, 'it' = 0, 'a' = 0, and 't' = 0. These output values represent the quality at which the chosen modality (e.g., video, audio-images, images-text, audio or text) should be incorporated into the output presentation. For example an output value of 0.812 for 'v' implies that the video file which has a data rate (required bandwidth) closest to 0.812 of the maximum bandwidth will be used within the presentation.

All outputs can also be 0 meaning that no cross modality adaptation should occur based on the given input parameters. The membership functions within TeleMorph’s five output fuzzy sets are identical to those used for the ‘bpsCurr’ and ‘vCurr’ fuzzy inputs, as shown in Fig. 1. In subsequent invocations of the fuzzy inference system from a mobile intelligent multimedia presentation system, the outputs values, ‘v’, ‘ai’, ‘it’, ‘a’ and ‘t’, will form the input values ‘vCurr’, ‘aiCurr’, ‘itCurr’, ‘aCurr’, and ‘tCurr’, respectively. However the one whole output value from the previous invocation of the FIS, such as 0.812 for ‘v’, may not have mapped directly to an existing video quality. As stated above, the video file which has a data rate closest to 0.812 of the maximum bandwidth required by any available media type will be used within the presentation, such as a multimodal video presentation requiring a bandwidth of 0.75, in which case 0.75 will be the value used for the ‘vCurr’ fuzzy input.

IV. TELETURAS

The domain chosen as a testbed for TeleMorph is *e* Tourism. The application that has been developed, *TeleTuras*, is now detailed. *TeleTuras*, an interactive information aid for tourists in the city of Derry, which has been implemented as a mobile intelligent multimedia presentation system, is an example testbed for TeleMorph’s network-aware fuzzy inference decision system. *TeleTuras* communicates TeleMorph-adapted presentations to tourists, focusing on the output modalities used to communicate information and also the effectiveness of this communication. In order to facilitate the comprehensive testing, evaluation, and demonstration of TeleMorph’s ability to control network-aware cross-modality adaptations *TeleTuras* integrates a wide range of manually generated multimedia and bandwidth specific test scenarios. *TeleTuras* enables input queries such as “Tell me about Derry”, “What is the Guildhall?”, and “Where is the University of Ulster?”. In this way *TeleTuras* provides tourists with instantaneous access to relevant information from a multimedia presentation system on their mobile computing devices. *TeleTuras*’ graphical user interface, as shown in Fig. 2, serves as an emulated and analogous mobile device interface for end-user evaluations.

TeleTuras communicates pre-distilled canned multimedia content related to tourism in the city of Derry. The repository of content incorporates manually created multiple fidelities for all media types with ranges for video, audio-images, images-text, audio and text ranging from 44-887 Kbps, 24-344 Kbps, 24-288 Kbps, 3-320 Kbps, and 8 Kbps, respectively.

V. EVALUATION OF TELEMORPH

The evaluation of TeleMorph has been performed through both subjective and objective testing. Both evaluation types have utilized six disparate test scenarios which are integrated into and accessible from *TeleTuras*’ GUIs. The objective evaluation and 30 subjective evaluations of TeleMorph have been carried out using *TeleTuras* and the six disparate test scenarios which it incorporates. Both evaluation techniques assess how TeleMorph performs in relation to a number of metrics which arise from the challenges associated with cross-modality adaptations based on mobile network bandwidth fluctuations in mobile intelligent multimedia presentation systems. These challenges include 1)



Fig. 2. *TeleTuras*’ GUI with mobile device emulator.

determining whether to transcode media further or transcode across modalities, 2) determining the most effective and acceptable destination media types following adaptation (and the quality/fidelity of this new content), and 3) cautiously controlling the timing and frequency of cross-modality adaptations.

In TeleMorph’s fuzzy logic implementation the outputs represent the fidelity or quality at which each new modality should be incorporated into the presentation, which addresses 2) above. The fuzzy implementation also integrates a degree of flexibility when dealing with the ‘adaptFreq’ fuzzy input, which control the maximum acceptable frequency of cross-modality adaptations and this varies based on the current and destination modalities involved in adaptations. The direct affect that type and frequency of cross-modality adaptations (through the use of ‘adaptFreq’) have on a presentation’s effectiveness and end-user acceptance are assessed based on an empirical figure of 11 s for the amount of elapsed time which ideally should have elapsed between cross-modality adaptations. This aspect of TeleMorph, referenced in 3) above, is further assessed in the subjective evaluation discussed below in Section V-B.

A. Objective Evaluation

TeleMorph’s objective evaluation includes a detailed description of what adaptations occur during six indicative sample presentations. These six sample test scenarios, as depicted in Fig. 3, incorporate all available media types in various adaptation types and thus represent appropriate evaluations of TeleMorph. Each test scenario incorporates different qualities of the same media elements, e.g., images-text at a quality of 0.1 in the first scenario and 0.3 in test scenario 3. To facilitate this objective analysis, *TeleTuras*’ GUI, as described in Section IV, was used to create trace data for each test scenario including details of the timing and frequency of adaptations, the packet loss prevented as a result of each adaptation, as well as details of the input parameters that brought these adaptations about. Specifically the trace

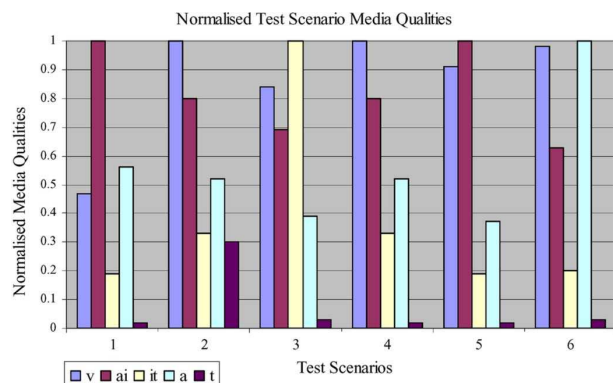


Fig. 3. Available media qualities in the six test scenarios.

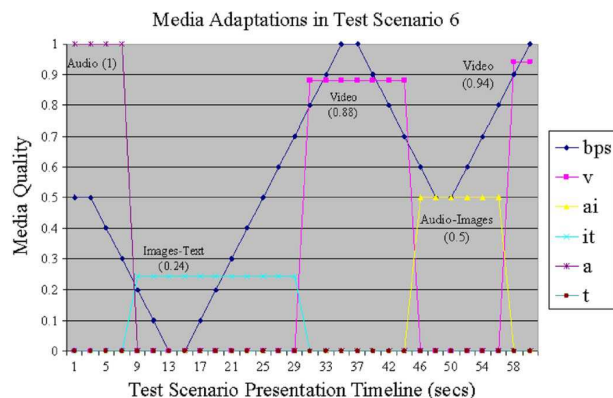


Fig. 4. Adaptation trace data from test scenario six.

data includes current bandwidth, the percentage packet loss, the value for 'adaptFreq' (adaptation frequency), current, lowest, and highest qualities available for video, audio-images, images-text, audio, and text, as well as five fuzzy output values representing the new qualities for video, audio-images, images-text, audio, and text following each adaptation. Also, the percentage packet loss prevented through the current adaptation is detailed as well as the time at which each adaptation occurred during the presentation.

The objective evaluation of TeleMorph was completed using manually generated sample bandwidth data which intentionally features frequent erratic fluctuations in available network bandwidth to enable comprehensive stress testing of TeleMorph. This enabled the evaluation of TeleMorph's performance in worst case scenario network conditions, thus proving the system's ability to perform acceptably even in the worst possible network bandwidth circumstances. Trace data recorded for all the adaptations that occur during each of the six test scenarios was recorded and analyzed. The trace data for test scenario six, graphically represented in Fig. 4, is indicative of the cross-modality adaptive behavior embodied in TeleMorph and so was utilized to evaluate TeleMorph objectively.

In test scenario six, TeleMorph prevented a packet loss of 60% by initially adapting across modalities from audio to a presentation composed of images and text, compared to

transcoding audio to its lowest available quality. In this way, TeleMorph addresses the first evaluation metric. The average percentage packet loss prevented for all six test scenarios is 44.35%. The first adaptation in test scenario six from unimodal audio to a presentation composed of images and text is the most effective adaptation given that there is no alternative multimodal presentation available which could be transmitted at the available bandwidth of 0.5 (172 Kbps). This complies with the second evaluation metric, as a multimodal output presentation was maintained even under very low bandwidth conditions.

This behavior is preferable based on the principles outlined within the cognitive load and dual coding theories. Similarly, the second and fourth cross-modality adaptations in the test scenario comply with the second evaluation metric as they maintain multimodal output presentations during increase in the presentation's quality despite the existence of a unimodal audio presentation with a higher output quality. TeleMorph's frequency of adaptations in all test scenarios complies with the third evaluation metric. An average of 19 s elapsed before TeleMorph executed cross-modality adaptations involving increases in the output presentation's quality, and an ideal average of 11 s passed in all six test scenarios before adaptations occurred which involved decreases in the output presentation's quality. TeleMorph's 'adaptFreq' fuzzy input was designed to impose greater restrictions on adaptations involving increases in the presentation's quality than decreases in its quality. This is because decreases are crucial in maintaining the presentation's usability, whereas presentation quality increases are treated as being complimentary for the presentation but not essential for its successful delivery, presentation and usability. Thus, TeleMorph fulfils the third evaluation metric successfully.

B. Subjective Evaluation

TeleMorph's bandwidth determined cross-modality adaptations were evaluated subjectively by 30 subjects. The questionnaire-based experiment was conducted using a wirelessly connected tablet computer running TeleTuras. Subjects viewed eight test scenarios, six of which were unique, and completed a printed questionnaire based on each test scenario. The majority of evaluation participants fell within the age bracket 26–30 (mean age of 28), with 24 out of the 30 subjects being male. Regards levels of computer literacy, two, three, eight, and 12 of the subjects graded themselves as having a novice, literate, experienced and expert knowledge of computers, respectively.

The questionnaire responses for test scenarios 1 to 6 were collated and statistics derived from the user responses and their feedback. The mean evaluation score for all subjects was 5.6, which implies an above average level of user-acceptance to TeleMorph-adapted multimedia presentations. 5.6 is considered to be a very positive for this subjective evaluation given that irregular bandwidth sample data was utilized during the eight test scenario presentations, as outlined above. This mean score of 5.6 does not correspond to the maximum frequency of occurrence which lies from 6-7, as depicted in the histogram in Fig. 5(b), because of the outliers within TeleMorph's subjective evaluation results. As a result of these outliers a standard deviation of 2.2 exists in the subjective evaluation scores. The outliers, with values that deviate most from the

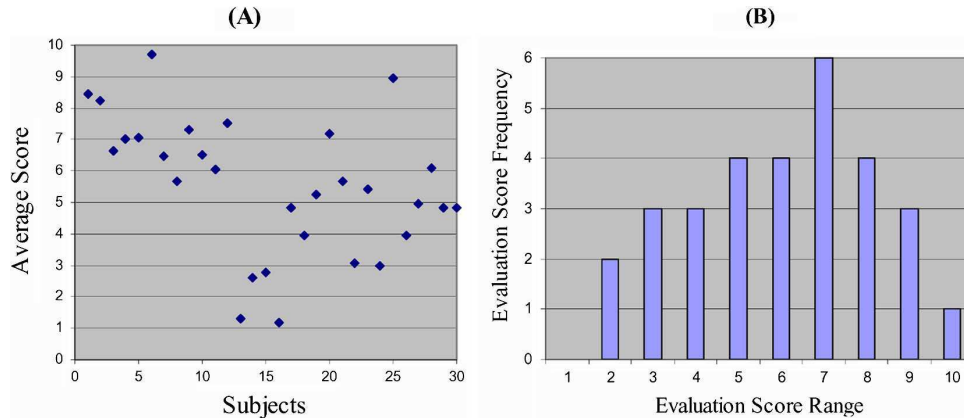


Fig. 5. Subjective evaluation mean scores and histogram. (a) Average score for 30 subjects and (b) histogram of subjective results.

mean score, represent the acceptance ratings given by specific subject-groups, such as those with novice and literate levels of computer knowledge where mean scores were 2.0 and 4.1, respectively.

Each of the eight test scenarios were evaluated using the following four questions. A) Given the bandwidth fluctuations and resulting adaptations which occurred during this test scenario, would you consider the multimedia presentation acceptable? B) Would you consider the number of adaptations and the frequency at which they occurred during the presentation acceptable? C) How would you rate the media adaptations which occurred during the presentation? D) Did you find the adaptations which occurred during this presentation distracting? Each of the aforementioned evaluation metrics are assessed based on subject responses to these four questions.

The mean evaluation scores for all 30 participants are depicted in Fig. 5(a). The questions within the evaluation questionnaire, labeled A, C and D above, facilitated the assessment of TeleMorph’s performance in relation to the first two evaluation metrics. The average scores for these questions were 5.6, 5.4, and 5.6, respectively, implying an above average acceptance level for these aspects of TeleMorph’s adaptations and compliance with evaluation metrics one and two. The second question in the questionnaire, labeled B above, addresses the remaining evaluation metric related to the frequency of adaptations. The average score for this question was 5.6, thus indicating an above average high level of acceptance for this vital element of TeleMorph’s adaptation behavior.

Test scenarios one and three were repeated during the subjective evaluation to assess the consistency of user’s evaluation scores. Test scenario 7, which was a concealed repeat of the presentation in scenario 3, had an average score of 5.8 compared to 5.4 for scenario 3. Test scenario 8, which was an acknowledged repeat of scenario 1, had an average score of 6.1 compared to 5.8 for scenario 1. Both of these statistics imply a slightly improved level of acceptance towards cross-modality adaptations following the completion of the previous six test scenarios. However, the similarity between the scores of the two repeated scenarios, 1 and 3, implies a general consistency of user’s scores throughout the subjective evaluation. In addition, when the av-

erage user score for each question is calculated for all test scenarios, we also see a very high level of consistency of scoring.

Further to these results, obvious correlations were discovered between users’ ages, levels of computer literacy and acceptance of cross-modality adaptation as performed by TeleMorph. In the age group 18–25 TeleMorph achieves a very high rating of 8.1, with those aged 26–30 giving an average score of 6. In addition, a direct correlation exists between an increasing level of users’ computer literacy and their average acceptance scores for TeleMorph-adapted multimedia presentations.

VI. CONCLUSION

TeleMorph addresses a number of challenges which have been used as metrics during evaluations associated with controlling cross-modality adaptations including 1) determining whether to transcode media further or transcode across modalities, 2) determining the most effective and acceptable destination media types following adaptation and the quality/fidelity of this new content, and 3) cautiously controlling the timing and frequency of cross-modality adaptations.

Results obtained during objective evaluations of TeleMorph have been presented. The average percentage packet loss prevented for all test scenarios involved in the objective evaluation process was 44%. As detailed previously TeleMorph’s cross-modality adaptations throughout the test scenarios were shown to have complied with all the evaluation metrics outlined above. Also, TeleMorph’s frequency of adaptations in all test scenarios fulfilled the third evaluation metric, as an average of 19 s elapsed before TeleMorph executed cross-modality adaptations involving increases in the output presentation’s quality, and an ideal average of 11 s elapsed on average during the six test scenarios before adaptations occurred which involved decreases in the output presentation’s quality. The empirical figure of 11 s is based on initial experiments carried out with prototypes of TeleMorph. During subjective testing, the average evaluation score of the subjects was 5.6, which implies an above average level of user-acceptance to TeleMorph-adapted multimedia presentations. Average evaluation score 5.6 is considered to be very positive for this subjective evaluation given that ir-

regular bandwidth sample data was utilized during the eight test scenario presentations, as outlined above.

TeleMorph differs from other approaches which have been presented, in that it focuses specifically on the challenges posed by controlling bandwidth determined cross-modality adaptations in a mobile network environment. The essence of mobile networks implies that the bandwidth of the network which is connecting the mobile device to its multimedia sources will vary continuously as the location of this device does. The necessity to control on-the-fly dynamic adaptations explicitly is therefore of key importance in a mobile intelligent multimedia presentation system. Thus, TeleMorph's control logic, which is represented by its inputs, fuzzy rules and outputs, has been implemented with an inherent awareness of these specific issues related to the mobile domain. In managing these aspects of mobile cross-modality adaptations is where TeleMorph differs most from other systems which carry out trans-media presentation modifications.

Future work on TeleMorph will incorporate additional fuzzy inputs including: bandwidth prediction algorithms, user preferences, and user modelling based on the demographic and computer literacy trends identified within the subjective results. These inputs will enhance TeleMorph's ability to provide more intelligent adaptation decisions to applications such as TeleTuras. The semantic worth of multimedia relative to a user query could also potentially be taken into account in TeleMorph's logic by implementing an additional semantics-related input fuzzy set. This would facilitate bespoke responses to disparate user queries based on semantic comparisons and similarities between linguistic queries and the semantics of available multimedia. Alternative testbed applications to TeleTuras will also be explored in the future, focusing in particular on the medical and educational domains.

REFERENCES

- [1] R. Babuska, "Fuzzy toolbox for MATLAB," in *Proc. 2nd IMACS Int. Symp. Mathematical and Intelligent Models in System Simulation*, Brussels, Belgium, 1993.
- [2] A. D. Baddeley and R. H. Logie, "Working memory: The multiple-component model," in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake and P. Shah, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1999, pp. 28–61.
- [3] S. Boll and W. Klas, "ZYX-A multimedia document model for reuse and adaptation of multimedia content," *IEEE Trans. Knowledge Data Eng.*, vol. 13, no. 3, pp. 361–382, 2001.
- [4] S. Boll, W. Klas, and J. Wandel, "A cross-media adaptation strategy for multimedia presentations," in *Proc. ACM Multimedia*, 1999, pp. 7–46.
- [5] D. Chalmers and M. Sloman, "QoS and context awareness for mobile computing," in *Proc. Int. Symp. Handheld and Ubiquitous Computing (HUC99)*, Karlsruhe, Germany, 1999, pp. 380–382.
- [6] D. Chalmers and M. Sloman, "A survey of Quality of Service in mobile computing environments," *IEEE Commun. Surv.*, vol. 2, no. 2, pp. 2–10, 1999b.
- [7] Crumpet (2006) [Online]. Available: <http://www.ist-crumpet.org> Nov. 2007
- [8] K. Curran and S. Annesley, "Transcoding media for bandwidth constrained mobile devices," *Int. J. Network Manage.*, vol. 15, no. 2, pp. 75–88, 2005.
- [9] C. Elting, J. Zwickel, and R. Malaka, "Device-dependant modality selection for user-interfaces-an empirical study," in *Proc. 6th Int. Conf. Intelligent User Interfaces*, San Francisco, CA, 2002, pp. 55–62.
- [10] A. Fox and E. A. Brewer, "Reducing WWW latency and bandwidth requirements by real-time distillation," in *Proc. 5th Int. World Wide Web Conf.*, Paris, France, 1996, pp. 1445–1456.
- [11] A. Fox, S. D. Gribble, E. A. Brewer, and E. Amir, "Adapting to network and client variability via on-demand dynamic distillation," in *Proc. 7th Int. Conf. Architectural Support for Programming Languages and Operating Systems*, Cambridge, MA, 1996, pp. 160–170.
- [12] A. Fox, S. D. Gribble, Y. Chawathe, and E. A. Brewer, "Adapting to network and client variation using infrastructural proxies: lessons and perspectives," *IEEE Trans. Personal Commun.*, vol. 5, no. 4, pp. 10–19, 1998.
- [13] A. Hildebrand, "EMBASSI: Electronic Multimedia and Service Assistance," in *Proc. IMC'2000*, Nov. 2000, pp. 50–59.
- [14] W. Klas, G. Christian, and F. Reinhard, "Cardio-OP: Gallery of Cardiac Surgery," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Jun. 1999, vol. 2, pp. 07–11, 1992.
- [15] R. Malaka, "Artificial intelligence goes mobile - artificial intelligence in mobile systems 2000," in *Proc. Workshop in Conjunction with ECAI 2000*, Berlin, Germany, Aug. 22, 2000, pp. 5–6.
- [16] R. Malaka, "Multi-modal interaction in private environments," in *Proc. Int. Seminar on Coordination and Fusion in MultiModal Interaction, Schloss Dagstuhl Int. Conf. Research Center for Computer Science*, Wadern, Saarland, Germany, 2001 [Online]. Available: http://www.dfki.de/~wahlster/Dagstuhl_Multi_Modality/Dagstuhl-2001.pdf
- [17] A. Paivio, *Imagery and Verbal Processes*. New York: Holt, Rinehart F& Winston, 1971.
- [18] A. Paivio, *Mental Representations..* New York: Oxford Univ. Press, 1986.
- [19] K. Pastra, "Beyond multimedia integration: Corpora and annotations for cross-media decision mechanisms," in *Proc. 5th Language Resources and Evaluation Conf. (LREC)*, 2006.
- [20] J. R. Smith, "Universal multimedia access," in *Proc. Multimedia Systems and Applications*, Boston, MA, 2001, pp. 21–32.
- [21] A. J. Solon, K. J. Curran, and P. M. Keivitt, "Bandwidth determined transmoding through fuzzy logic in mobile intelligent multimedia presentation systems," in *Artificial Intelligence Review*. Dordrecht, The Netherlands: Springer-Verlag, 2006, vol. 26.
- [22] J. Sweller, J. V. Merrienboer, and F. Paas, "Cognitive architecture and instructional design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, 1998.
- [23] T. C. Thang, Y. J. Jung, and Y. M. Ro, "Modality conversion in content adaptation for universal multimedia access," in *Proc. Int. Conf. Imaging Science, Systems and Technology (CISST)*, Las Vegas, NV, 2003, pp. 434–440.
- [24] T. C. Thang, Y. J. Jung, J. Lee, and Y. M. Ro, "Effective adaptation of multimedia documents with modality conversion," *Signal Process.: Image Commun.*, vol. 20, no. 5, pp. 413–434, 2005.
- [25] T. C. Thang, Y. J. Jung, and Y. M. Ro, "Modality conversion for QoS management in universal multimedia access," *Proc. Inst. Elect. Eng.*, vol. 152, no. 3, pp. 374–384, 2005.
- [26] W. Wahlster, "SmartKom: Foundations of Multimodal Dialogue Systems," in *Cognitive Technologies Series*. Heidelberg, Germany: Springer, 2006.
- [27] S. Wirag, "Scheduling of adaptive multimedia documents," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems (ICMCS)*, Florence, Italy, 1999, pp. 307–311.
- [28] L. A. Zadeh, *Inf. Contr.*, vol. 8, pp. 338–353, 1965.
- [29] A. Zipf and R. Malaka, "Developing location based services (LBS) for tourism—The service providers view. ENTER 2001," in *Proc. 8th. Int. Congr. Tourism and Communications Technologies in Tourism*, Apr. 2001, pp. 24–27.



Anthony J. Solon received the B.Sc. (Hons.) degree in computing science and the Ph.D. degree in mobile intelligent multimedia from the University of Ulster, U.K., in 2002.

His main research interests include mobile intelligent multimedia, location-based services, and mobile multimodal interfaces. He is a Research Associate with the Wireless Technology Research Centre. This position has seen him research and develop various novel location based systems and mobile/wireless applications. He has published over 15 research papers in conferences and journals and has also served as a reviewer for several conferences and international journals.



Paul McKeivitt received the B.Sc. (Hons.), M.S., and Ph.D. degrees.

He is Chair in Intelligent MultiMedia at the University of Ulster, U.K.. Previously, he was Associate Professor with the School of Computer Science at The Queen's University of Belfast, Northern Ireland. He has been Visiting Professor of Intelligent Multi-Media Computing in the Institute of Electronic Systems at Aalborg University, Denmark and a British EPSRC Advanced Fellow in the Department of Computer Science, University of Sheffield, U.K. His primary research interests are in NLP, including the processing of pragmatics, beliefs and intentions in dialogue. He has published over 50 research papers in international conferences and journals and is Editor of the *Artificial Intelligence Review Journal*.



Kevin Curran received the B.Sc. (Hons.) and D.Phil. degrees.

He is a Lecturer in computer science at the University of Ulster, U.K. His achievements include winning and managing European Framework projects, U.K. Government Council funded projects, and many Technology Transfer Schemes. He has published over 250 research papers to date in the field of distributed computing. He is a member of the editorial committee of the *Journal of Universal Computer Science*.

Dr. Curran is a member of numerous conference organizing committees.