Bandwidth determined transmoding through fuzzy logic in mobile intelligent multimedia presentation systems

Anthony J. Solon · Kevin Curran · Paul Mc Kevitt

Abstract   Mobile intelligent multimedia presentation systems are subject to various resource constraints including mobile network characteristics, mobile device capabilities and user preferences. Those presentation systems which incorporate remote multimedia content accessed across HTTP (Hyper Text Transfer Protocol) or RTP (Real Time Transfer Protocol) protocols are particularly reliant on the capabilities of the connecting mobile network (i.e. minimum, average and maximum bandwidth) and in particular on the real-time constraints (i.e. currently available bandwidth, packet loss, bit error rate, latency) which prevail during actual content transmission. One approach to address this is to scale content, thus reducing its datarate requirement, although this technique is inherently limited by the lowest acceptable quality of that media element. Alternatively, content can be converted from one modality to another with a lower resource requirement. TeleMorph, a cross-modality adaptation control platform is detailed here. Initially a brief introduction to Intelligent Multimedia and to Mobile Intelligent Multimedia is given, and key systems discussed. The main premise of TeleMorph is that cross-modality adaptations in mobile presentation systems must be controlled in a manner which gives primary consideration to bandwidth fluctuations as well as the constraints listed above. The current prototype of TeleMorph, which uses a fuzzy inference system to control cross-modality adaptations between video and audio, is described. Particular focus is given to the fuzzy inputs, fuzzy control rules and fuzzy outputs which have been utilised in decision making. TeleTuras, a tourist information application which has been implemented as a testbed for TeleMorph, gives promising evaluation results based on multimedia and bandwidth specific test scenarios. TeleMorph is related to other approaches in the area of Mobile Intelligent Multimedia Presentation Systems. TeleMorph differs from

A. J. Solon (✉) · K. Curran · P. Mc Kevitt
Intelligent Multimedia Research Group, School of Computing & Intelligent Systems, Faculty
of Computing & Engineering, University of Ulster, Magee, Derry BT48 7JL, Northern Ireland, UK
e-mails: AJ.Solon@ulster.ac.uk; tonysolon@gmail.com

K. Curran
e-mail: KJ.Curran@ulster.ac.uk

P. Mc Kevitt
e-mail: P.McKevitt@ulster.ac.uk

other approaches in that it focuses specifically on the challenges posed by controlling bandwidth determined cross-modality adaptations in a mobile network environment. Future work on TeleMorph's output presentation composition will incorporate images and text also, thus allowing for extended adaptation between video, audio, images and text, as well as multimodal combinations of these media elements.

1 Introduction

Mobile networks impose various constraints on connected mobile multimedia presentation systems including: fluctuating bandwidth, transmission latency, packet loss, bit error rate and disconnections. These constraints vary on mobile networks as they do with all networks, but as a result of the intrinsic ability of 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 media content using RTP (Real-time Transport Protocol) or download media content using HTTP (Hyper Text Transfer Protocol) 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. Universal multimedia access (UMA) (Smith 2001) has become the driving concept behind enabling ubiquitous access to rich multimedia content on any type of terminal connected across heterogeneous networks, whilst also considering user preferences and maintaining an acceptable quality of service for the user (QoS) (Chalmers and Sloman 1999a, b).

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 (Curran and Annesley 2005; Fox et al. 1996, 1998; Fox and Brewer 1996) 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. For example, this technique would allow transmission of a video file with an original bandwidth requirement of 1.5 Mbps (Mega bits per second) on a third generation mobile network, following distillation to a quality corresponding to the available network bandwidth, e.g. 384 Kbps (Kilo bits per second). 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 datarate corresponding to that which is available on the host mobile network, its quality may be below this lowest usable quality, which would result in unacceptable presentation quality, and thus insufficient usability.

An alternative to multimedia distillation or transcoding strictly within modalities is to transmode, or adapt multimedia content across modalities (Boll et al. 1999) based on content-based 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). 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 goal of the platform prototype described here, TeleMorph, is not to create a multimedia document model for semantically representing content semantics, but to control cross-modality adaptations given these semantics, currently only manually created media alternatives in various modalities are utilised. 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 original source 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: images and audio, images and text, text and audio, text and video. This facilitates the fulfilment of TeleMorph's primary goal—to control cross-modality adaptation based on fluctuating bandwidth, and enables controlled evaluations of these cross-modality adaptations. As the multimedia alternatives are created manually this also enables TeleMorph to make adaptation decisions on the assumption that as the bandwidth for a media element increases, so too does the semantic quality/value of the media. However the semantic worth of multimedia could also easily be taken into account in adaptation decisions by adding an additional input fuzzy set to TeleMorph's fuzzy inference system.

Having metadata or a semantic representation associated with multimedia enables presentation systems to carry out cross modality adaptations on presentations in order to maintain the optimum output usability irrespective of composition modalities. With multimedia semantics available a presentation system can present information in varying modalities based on the requirements of these modalities relative to mobile network constraints. A major problem 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 repeating cross modality adaptations within one presentation is that the end-user's attention may be distracted from the actual information being portrayed, to the physical 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 the primary questions in controlling cross-modality adaptations are:

- When bandwidth decreases, should a cross-modality adaptation occur as opposed to further content scaling?
- To what extent must bandwidth fall below the lowest acceptable quality for a modality before a cross-modality adaptation is required (bearing in mind that if bandwidth decreases only a small amount, such as 5 Kbps, below the lowest acceptable quality it is just as likely to increase by that amount again within a short amount of time, e.g. 5 s)?
- What is the most effective and acceptable destination modality given the current presentation composition and content quality, as well as the bandwidth which is currently available?
- At what fidelity or quality should each new modality be incorporated into the presentation?
- What is the maximum acceptable frequency of cross-modality adaptations, and how does this vary based on the current and destination modalities involved in an adaptation?

● What direct effect does type and frequency of cross-modality adaptations have on presentation effectiveness and user acceptance?

The effects of cross-modality adaptation between both discrete and temporal multimedia elements must be considered before morphing presentations across modalities, as should the effectiveness and user-acceptance of the output presentation. As Elting et al. (2002) discovered, 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 & 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 & Speech). These points which are based on the Cognitive Load Theory (CLT) (Elting et al. 2002) should be incorporated into multimodal presentation systems to achieve ultimate usability.

Automated bandwidth-specific cross-modality adaptation is the challenge addressed in TeleMorph. The system detailed here is a prototype, which focuses on adaptations between video (audio–visual) and audio with future versions allowing for adaptation between video, audio, images and text, as well as multimodal combinations of these. In Sect. 2 an overview of transmoding in mobile intelligent multimedia presentation systems is given. Section 3 describes TeleMorph and how it has been implemented as a fuzzy inference system, with subsections detailing inputs, rules, outputs. TeleTuras a tourist information application, which has been developed as a testbed for testing TeleMorph using specific test scenarios, is detailed in Sect. 4. Following this, Sect. 5 discusses evaluations of TeleMorph using TeleTuras and a disparate selection of video and audio media files with varying bandwidth requirements. TeleMorph is then related to other work in Sect. 6 and finally Sect. 7 concludes and discusses future work.

2 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 (2000, 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. DEEP MAP (Malaka 2000, 2001; Malaka and Zipf 2000; Malaka et al. 2000; EML 2006) is a prototype of a digital personal mobile tourist guide which integrates research from various areas of computer science: geo-information systems, databases, natural language processing, intelligent user interfaces, knowledge representation, and more. CRUMPET (Creation of User-friendly Mobile services Personalised for Tourism) (EML 2006; Crumpet 2006; Zipf and Malaka 2001) implements, validates, and tests tourism-related value-added services for nomadic users across mobile and fixed networks.

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. Boll et al. (1999) 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 their application, Gallery of Cardiac Surgery (Cardio-OP) (Klas et al. 1999), 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 (Boll and Klas 2001), Tiempo (Wirag 1999) and REVEAL THIS (Pastra 2006) are other examples of multimedia document models which provide some element of flexibility and adaptation. Thang et al. (2003, 2005a, b) 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.

MPEG-21 Digital Item Adaptation (DIA) (Pereira et al. 2005; ISO/IEC 21000-7 2004; Sun et al. 2003; Vetro et al. 2005; Vetro and Timmerer 2005) is a standard which specifies tools to assist with the adaptation of multimedia. Specifically, DIA details metadata for assisting with the adaptation of digital items according to constraints on storage, transmission and consumption, thereby enabling various types of quality of service (QoS) management. MPEG-21 compliments MPEG-7, although whilst MPEG-7 deals with the semantic description of content, it does not incorporate information detailing compatible and format usage environments.

3 TeleMorph

TeleMorph (Solon et al. 2004a, b, c, d) makes content adaptation decisions for mobile intelligent multimedia presentation systems. Specifically it is a cross-modality adaptation control platform which bases adaptation decisions on relevant constraints such as bandwidth fluctuations. TeleMorph utilises fuzzy logic to control these cross-modality adaptation decisions, which occur between video and audio in the current prototype of the platform. TeleTuras, a tourist information application for the city of Derry in Northern Ireland, has also been implemented in order to provide one testbed for TeleMorph. Using TeleTuras, evaluations of TeleMorph's performance during dynamic cross-modality adaptations have been carried out. This was enabled using manually distilled multimedia and bandwidth-specific 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 presentation system to exploit its decision-making abilities. Thus it is not dependant on any particular application domain or implementation, but instead, can provide multimedia adaptation decisions to any presentation system given the required inputs. Figure 1 depicts the role TeleMorph performs in a mobile intelligent multimedia presentation system.

Zadeh (1965) 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 significance and consequence of decision outputs. When performing dynamic multimedia adaptation in a live presentation system, priority lies in the system's ability to maintain continuity of output presentation and seamlessness during adaptations whilst observing restrictions imposed by the constraints acting upon the system, as well as determining output fidelities for multimedia content. However, 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 any precise output modality fidelity.
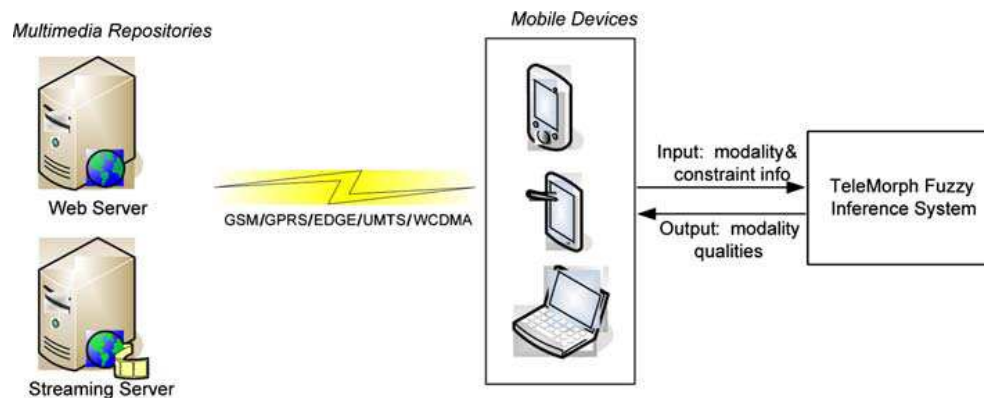
Fig. 1 Role of TeleMorph in a mobile intelligent multimedia presentation system

TeleMorph's FIS was implemented using the matlab Fuzzy Logic Toolbox (Babuska 1993). 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. 2. The prototype of TeleMorph described here focuses solely on multimedia adaptations between video and audio only, thus the inputs and outputs, as shown in Fig. 2, reflect this. In order to make cross modality adaptation decisions between audio and video the following inputs are used: current bandwidth, current video quality, lowest available video quality, highest available video quality, current audio quality, lowest available audio quality, highest available audio quality and adaptation frequency. The outputs include video (v) and audio (a). The membership functions which describe the linguistic values in the input and output fuzzy sets use a combination of polynomial based Z-curves (asymmetrical polynomial curve open to the left) and S-curves (asymmetrical polynomial curve open to the right), as well as triangularly shaped 'trimf' Membership Functions (MFs) to fuzzify inputs.

A sample test scenario using TeleMorph is shown in Fig. 3 where the currently available bandwidth represented by 'bpsCurr' has a value of 0.1, current video quality represented by 'vCurr' is 0.8, minimum video quality represented by 'vLowest-bps' is 0.8, maximum video quality represented by 'vHighest-bps' is 1, current audio quality represented by 'aCurr' is 0, minimum audio quality represented by 'aLowest-bps' is 0, maximum audio quality represented by 'aHighest-bps' has a value of 0.3, and adaptation frequency denoted by 'adaptFreq' has a value of 0.9. Given these input values the outputs are 0 for video 'v' and 0.0892 for audio 'a'. This implies an adaptation from video at a quality of 0.8 to audio at a quality of approximately 0.1 thus meeting the bandwidth constraint. As can be seen from Fig. 3 the antecedents (inputs) for rules numbered 5, 6 and 11 fire and thus their consequents (outputs) are successful and are aggregated together to provide a value for centroid defuzzification, which results in an output of 0.0892 for audio. It is important to note that the fifteen rules in Fig. 3 are purely indicative of TeleMorph's logic. In total there are over 200 rules which determine the output decisions of TeleMorph.
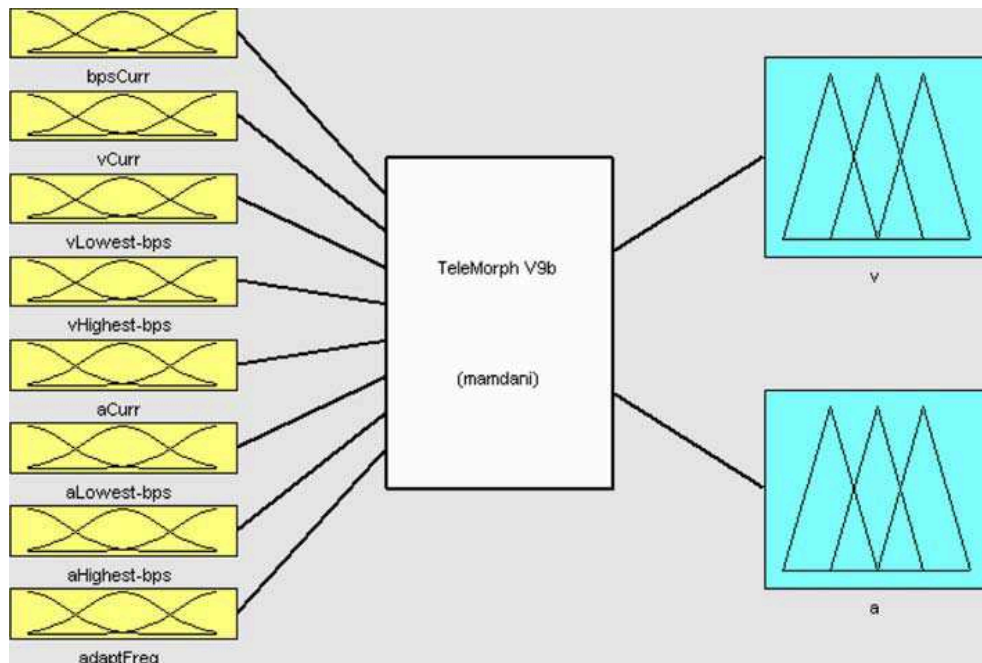
Fig. 2 Architecture of TeleMorph's FIS for video and audio (adaptations)
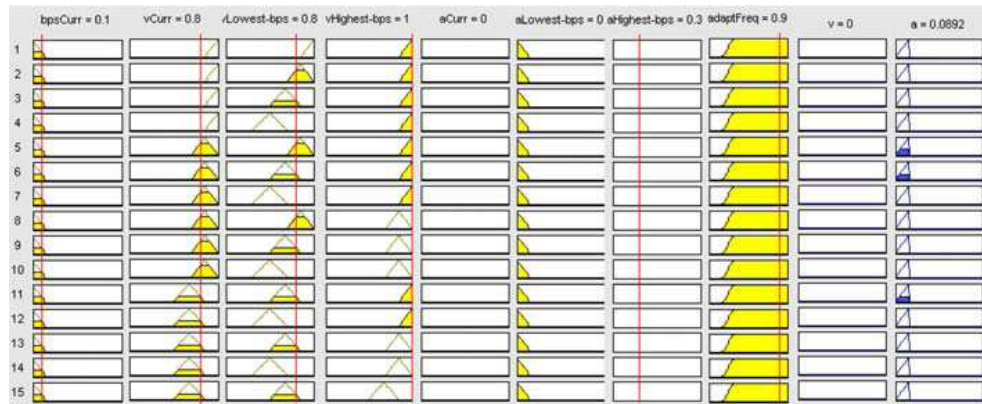
Fig. 3 Fuzzy inference in TeleMorph

3.1 TeleMorph FIS inputs

Given inputs including current bandwidth ('bpsCurr'), current modality qualities ('vCurr', 'aCurr'), potential modality qualities ('vLowest-bps', 'vHighest-bps', 'aLowest-bps', 'aHighest-bps') and a value representing the frequency of adaptation ('adaptFreq'), TeleMorph provides outputs ('v', 'a') detailing the quality at which each modality should be present in the output presentation. Figure 2 shows the structure of a fuzzy logic system which controls adaptations between two modalities—video and audio. This system has eight normalised inputs including: current bandwidth, current video quality (relative to the minimum and maximum video qualities available), minimum video quality (video requiring least bandwidth), maximum video quality (video requiring most bandwidth), current audio quality (relative

Fig. 4 TeleMorph's FIS input membership functions

to the minimum and maximum audio qualities available), minimum audio quality (audio requiring least bandwidth), maximum audio quality (audio requiring most bandwidth) and adaptation frequency. Using these eight inputs the inference system's rules determine whether to instruct an adaptation from audio to video (or vice versa) and at what fidelity this targ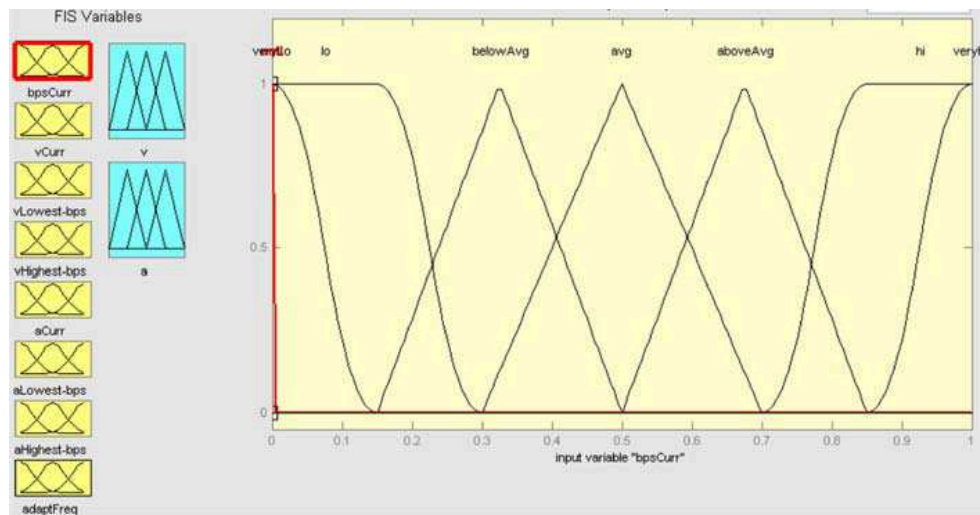et modality should be used in the output presentation. The system can also determine that no compositional changes should be made to a presentation given certain parameters.

Figure 4 shows an example of an input fuzzy set in TeleMorph and the shape of the Membership Functions (MFs) in this set. The input fuzzy sets in TeleMorph use a combination of polynomial based Z-curves (asymmetrical polynomial curve open to the left) and S-curves (asymmetrical polynomial curve open to the right), as well as triangularly shaped 'trimf' MFs to fuzzify inputs. The MFs in the input fuzzy sets include null, 'veryLo', 'lo', 'belowAvg', 'avg', 'aboveAvg', 'hi', and 'veryHi'. The 'veryLo' and 'lo' MFs were designed using Z-curves as this allows them to incorporate a range of decreasing values to the left of their negatively sloped curves, whilst 'hi', and 'veryHi' utilise S-curves as this allows them to cover a range of increasing values to the right of their positively sloped curves. The triangularly shaped 'trimf' function is used to describe the 'belowAvg', 'avg', 'aboveAvg' MFs as this allows for full membership in these MFs at a singleton point whilst also maintaining a symmetric shape. Combined, this range of MFs enables TeleMorph to comprehensively and effectively fuzzify all inputs to the system.

These MFs are relative to the normalised Universe of Discourse (UoD) which is limited by the highest bandwidth datarate 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 requiring 64 Kbps) to 0. This means that the membership functions always represent the maximum possible area within the UoD and thus the system always has optimal granularity in its decision making. This can best be described by comparing this technique to an approach where the UoD range is static in representing a range from 0 bps to the maximum bandwidth feasible with that mobile network (e.g. 1 Kbps).

If the audio inputs spanned a region of this domain from 0 to 100 Kbps and video spanned from 250 to 400 Kbps the decision-making effectiveness of the MFs covering the region of

the UoD from 0 to 400 Kbps is degraded. In contrast, if the UoD (always using a normalised range from 0 to 1) is adapted to only incorporate the pertinent region of the input space from 0 to 400 Kbps, then much more granular, and thus more effective, decision making is feasible. In Fig. 5 the gauge from TeleMorph's GUI shows the current bandwidth available as 189,110 bps out of a maximum of 309,168 bps, where 309,168 is the bandwidth required to stream the highest quality multimedia file (video in this example) and not the maximum bandwidth feasible with the mobile network. All input fuzzy sets in TeleMorph incorporate a 'null' membership function which lies in the region of the UoD from 0 to 0.000001. This exists due to the fact that there is no method of representing a null value of 0 in matlab's fuzzy logic toolbox. Creating a membership function with such a minuscule range effectively provides an appropriate workaround for this issue.

3.1.1 Current bandwidth

'bpsCurr' Is the input fuzzy set which represents the normalised value for currently available bandwidth in the mobile network which is being used to transmit the output multimedia presentation. The value is normalised against the maximum datarate required by available output multimedia files. The membership functions in the 'bpsCurr' fuzzy set are the same as those shown in Fig. 4. This input can be set using a variety of techniques including manually, automatically using sample data loaded from a file, or by passively and actively detecting live network throughput. During testing bandwidth can be set manually using the bandwidth gauge in TeleMorph's GUI (Graphical User Interface) shown in Fig. 5. The small grey triangle in this figure indicates current bandwidth. This gauge in the GUI allows the user to test the effectiveness of TeleMorph's adaptation decisions based on specific bandwidth fluctuations which they control. This method of controlling the 'bpsCurr' input has proven invaluable in testing TeleMorph.

Using a software package called Network Simulator 2 (NS/2) (Kurkowski et al. 2005; Qiu et al. 2004) 36032 normalised sample datarate values were generated. This allows for simulations and test runs of TeleMorph by reading this sample bandwidth data from the generated file. Using NS/2 this normalised data varies based on proven wireless network emulation models. With this data it was possible to automate tests of TeleMorph and to run extensive simulations over long durations of time, whilst recording adaptation information for subsequent subjective analysis. In Table 1 below the first 20 samples of the bandwidth data generated by NS/2 is shown. In order to custom design tests of TeleMorph, sample data was also created manually to incorporate all possible bandwidth fluctuations. This allowed for specific scenarios to be evaluated comprehensively as detailed in Sect. 5.
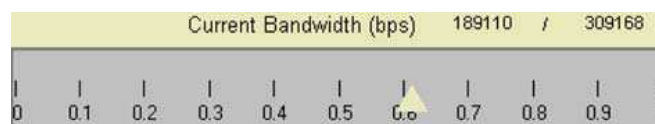

Fig. 5 Bandwidth gauge in TeleMorph's GUI

Table 1 Sample of bandwidth data generated by NS/2	0.5,0.45848,0.52384,0.08,0.37768,0.5,0.5,0.55072,0.42,0.5, 0.6,0.59648,0.44872,0.49,0.5,0.65744,0.1952,0.4,0.51904,0.08

In order to deploy TeleMorph in a live intelligent multimedia presentation system, real-time bandwidth measurements must be used to test its effectiveness and ability to react promptly to fluctuating bandwidth in a live situation. In order to obtain values for 'bpsCurr' in a deployed scenario real-time bandwidth measurement techniques have been utilised. For the purpose of testing however, these techniques use basic active download timing measurements. The active measurements use one specific multimedia file and compare the download start and end times to ascertain throughput. Active measurements have the disadvantage that they impose additional traffic on the network which can affect the throughput in the attempt to measure it. Time delays are also an inherent disadvantage of active measurements as a specific file must be transmitted. Passive measurements offer a better alternative as they occur incidental to presentation content transmissions which obviously occur regardless of whether the bandwidth is being measured.

3.1.2 'vCurr', 'vLowest-bps' & 'vHighest-bps'

The 'vCurr' input fuzzy set represents the datarate of the video file currently part of the output presentation. 'vCurr' is normalised against the maximum bandwidth required by any potential presentation output in the system (e.g. 300 Kbps video or 450 Kbps combination of audio and images). Thus it is mapped to a normalised value relative to the highest bandwidth datarate required by one of the available multimedia files in the system. If no video is included in the current output presentation then 'vCurr' is set to 0. 'vLowest-bps' represents the lowest datarate of the available video, whilst 'vHighest-bps' is the highest bandwidth required by the available video files.

3.1.3 'aCurr', 'aLowest-bps' & 'aHighest-bps'

These input fuzzy sets represent similar variables to those described for video. Audio generally requires a lower datarate than video due to its unimodal nature, although one cannot always assume this. TeleMorph allows for this in it ability to degrade/increase presentation quality from audio to video as effectively and seamlessly as it can degrade/increase video to audio.

3.1.4 'adaptFreq'

One of the principal limitations of cross modality adaptation is the effect these adaptations have on the flow and continuity of the output presentation. 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 effecting the frequency of cross modality adaptation, namely the time elapsed since the previous adaptation and the number of adaptations carried out so far during the multimedia presentation. 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, as shown in Fig. 6(a) and (b) respectively. The normalised value for time elapsed is calculated in relation to a threshold constant of 11 s which is an empirical value for the ideal amount of time which should have elapsed before a subsequent adaptation occurs.
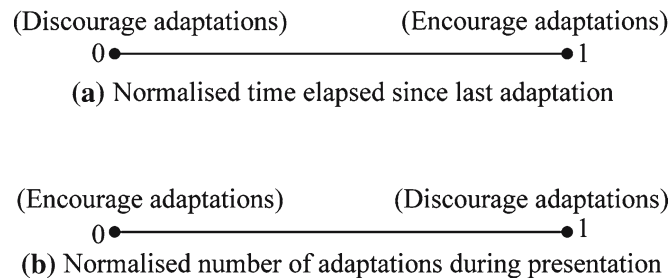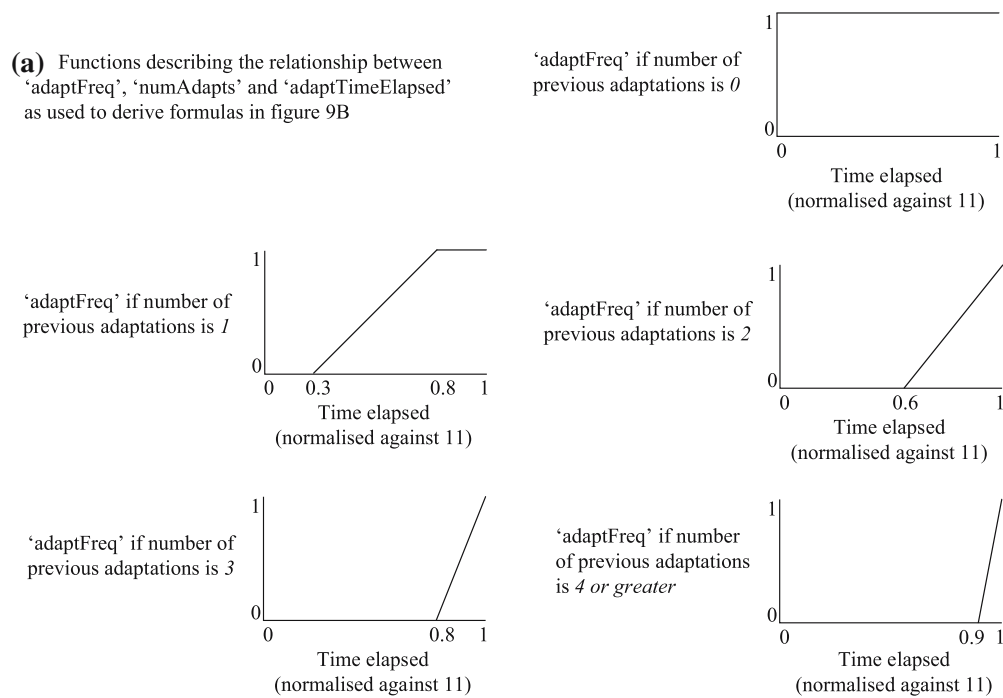
(Discourage adaptations)   (Encourage adaptations)

0 ●————————————————————● 1

(a) Normalised time elapsed since last adaptation

(Encourage adaptations)   (Discourage adaptations)

0 ●————————————————————● 1

(b) Normalised number of adaptations during presentation

Fig. 6 Factors combining to generate 'adaptFreq' input values

In order to generate input values for 'adaptFreq' it was necessary to create functions to describe the relationship between the domain depicted in Fig. 6(a) and the number of previous adaptations as it is represented in Fig. 6(b). To this effect functions were created, based on expert knowledge, to describe 'adaptFreq' in terms of the 'time elapsed' and 'number of previous adaptations' factors. The values describing 'adaptFreq' when the number of adaptations is 0, 1, 2, 3 and 4 or greater are shown in Fig. 7(b) and include formulas which are unique for each value of 'numAdapts'. Also the functions from which these formulas are derived are depicted in Fig. 7(a). Combining these two complimenting factors in this way reduces the number of rules required in the system thus increasing the design, development and execution efficiency of the FIS.

Following calculations to compute a combined value for 'adaptFreq', the value is passed as a parameter into in TeleMorph's FIS where it is represented as a fuzzy logic set. The shapes of the MFs in this fuzzy set are shown in Fig. 8. All of the eight MFs are described using S-curves which are asymmetrical polynomial curves open to the right.

In TeleMorph, as described previously, rules are designed to use 'adaptFreq' to control the frequency of cross-modality adaptations. This ensures a consistent degree of usability throughout a presentation. Different types of cross-modality adaptations (e.g. linguistic text to non-linguistic images) have a varying reliance on this temporal limitation as the transition across modalities is more noticeable and has a greater impact on the seamlessness of the transition and thus the user's attention. This is based on research related to the dual coding theory (Bernsen 1994; Clark and Paivio 1991; Uden and Campion 2000). Having MFs which represent the temporal requirement to varying degrees allows adaptations to use varying prerequisite values for 'adaptFreq'. An adaptation from audio to video which increases output quality requires a 'hi' value for 'adaptFreq', whereas increasing from video to audio requires an 'average' value, or higher. Having S-shaped MFs for 'adaptFreq' enables the inclusion of all values in the domain which are greater than the initial input value. For example, in the scenario where the presentation is increasing quality from audio to video, at least an 'average' value for 'adaptFreq' is required (as shown in Fig. 9(a)), so even if the value is higher and falls within the MFs for 'aboveAvg', 'hi' or 'veryHi', rules controlling quality increases from audio to video will fire. Alternatively if an adaptation involves an increase in output quality from video to audio, the value for 'adaptFreq' must be 'hi' (as shown in Fig. 9(b)). 'adaptFreq' has the ability to restrict adaptations in this way if rules utilise the available MFs effectively. All of the values of 'adaptFreq' which are required for audio and video adaptations in the current prototype of TeleMorph are shown in Fig. 9, including those required for reductions in output quality (shown in Figs. 9(c) and (d)). Effectively the lower the value for 'adaptFreq' is, the easier it is for adaptations to occur. Thus, only rules controlling seamless adaptations which have a negligible impact on usability (e.g. audio and images

(a) Functions describing the relationship between 'adaptFreq', 'numAdapts' and 'adaptTimeElapsed' as used to derive formulas in figure 9B

'adaptFreq' if number of previous adaptations is *0*

Time elapsed (normalised against 11)

'adaptFreq' if number of previous adaptations is *1*

Time elapsed (normalised against 11)

'adaptFreq' if number of previous adaptations is *2*

Time elapsed (normalised against 11)

'adaptFreq' if number of previous adaptations is *3*

Time elapsed (normalised against 11)

'adaptFreq' if number of previous adaptations is *4 or greater*

Time elapsed (normalised against 11)

(b) Values (derived from functions shown in figure 9A) for 'adaptFreq':
- for all whole values of numAdapts (adaptations occurred)
- and given normalised values for adaptTimeElapsed (time elapsed since last adaptation)

```
if (numAdapts == 0) {
        adaptFreq = 1
} elseif (numAdapts == 1) {
        adaptFreq = 2(adaptTimeElapsed) – 0.6
} elseif (numAdapts == 2) {
        adaptFreq = 2.5(adaptTimeElapsed) – 1.5
} elseif (numAdapts == 3) {
        adaptFreq = 5(adaptTimeElapsed) – 4
} elseif (numAdapts >= 4) {
        adaptFreq = 10(adaptTimeElapsed) – 9
}
```

Fig. 7 Functions and subsequent formulas describing 'adaptFreq' given values for 'numAdapts' and 'adaptTimeElapsed'

adapting to audio; or images and text adapting to images i.e. maintaining linguistic and/or non-linguistic content across adaptations) have been designed to have a 'veryLo' prerequisite value for 'adaptFreq', whereas the higher the antecedent value for 'adaptFreq' is, the more effect the adaptation in question has on users' attentions and the system's usability.

3.2 TeleMorph FIS rules and related adaptations

There are 221 rules in use in the current version of TeleMorph. These rules are divided into four groups: rules controlling increases in output presentation quality from video to audio, increases from audio to video, decreases in quality from audio to video and decreases from video to audio. The MFs from 8 input fuzzy sets which form the basis of the rules in
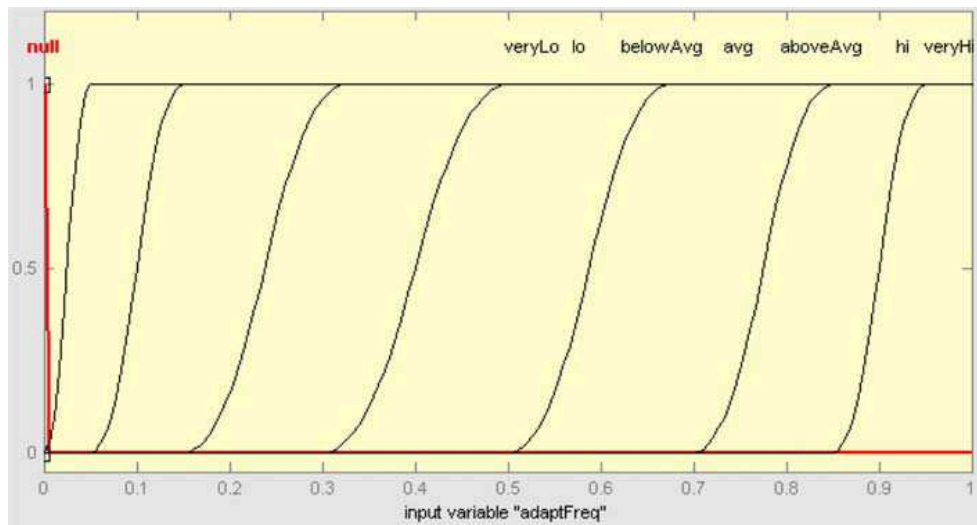
Fig. 8 'adaptFreq' Fuzzy set

Adaptations increasing presentation quality

(a) Audio to Video   **(b)** Video to Audio

Adaptations decreasing presentation quality

(c) Audio to Video   **(d)** Video to Audio
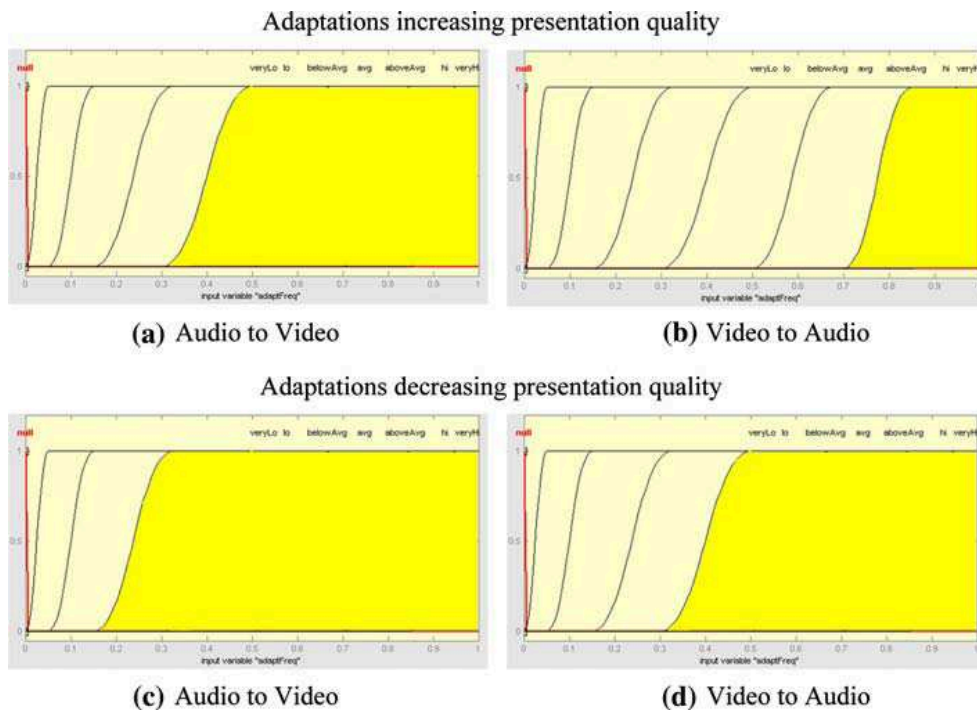
Fig. 9 Values of 'adaptFreq' which are required for audio ↔ video adaptations

TeleMorph's FIS are shown in Fig. 10 for convenience when considering the rule explanations which follow in Sects. 3.2.1–3.2.4.

TeleMorph's rules have been designed to only allow adaptations when the quality transition across the Universe of Discourse is substantial e.g. from 'avg' quality to 'veryLo' or 'veryHi' to 'avg'. By designing rules in this manner adaptation frequency is inherently
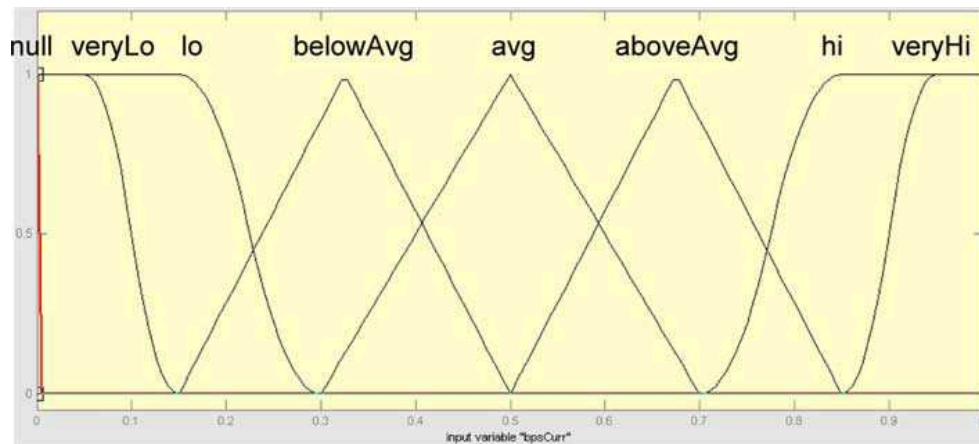
Fig. 10 Input set MFs used as antecedents in FIS rules

restricted in order to reduce the impact adaptations have on the flow and continuity of the presentation.

3.2.1 Quality decrease from video to audio

The first group of adaptation rules, which control quality decreases from video to audio, consists of 56 rules in total. This set of rules is further broken into four subsets of rules, each having an increasing antecedent value for 'bpsCurr'. Sample rules from each subset can be seen in Table 2 (three rules, 1, 17 and 35, are shown from the subset where 'bpsCurr' is 'veryLo'). 'aHighest-bps' is none for all rules in this group as it is has no influence on this group of rules.

The first subset (rules 1, 17, 35 in Table 2), with a 'veryLo' value for 'bpsCurr', has 35 rules which detail adaptations from video to 'veryLo' audio with 'vCurr' memberships ranging from 'veryHi' to 'avg'. Adaptations from video to audio given these conditions are not enabled if the value for video is lower than 'avg', as this would not constitute a substantial transition in quality. The resulting qualities for audio in this subset vary from 'veryLo' to 'aboveAvg' depending on the range of video qualities. For example, when video is 'veryHi' the system can adapt to a value for audio ranging from 'aboveAvg' to 'veryLo'. But when the value for video is 'avg' the system can only adapt to a 'veryLo' value for audio. This ensures compliance with the inherent necessity apt for rules to only allow substantial quality adaptations. The second subset in this group, which handles 'lo' values for 'bpsCurr', has 15 rules in total. The rule numbered 43 in Table 2 is a sample rule from this subset. The third subset in this group incorporates 'belowAvg' values for 'bpsCurr'. There are 5 rules in this subset, an example of which is the one numbered 54 in Table 2. The final subset within the rules which control quality decreases from video to audio consists of just one rule. Numbered 56 in Table 2, this rule manages adaptations when 'bpsCurr' is 'avg'.

3.2.2 Quality decrease from audio to video

This group of rules controls adaptations where the presentation quality decreases from audio to video. The 56 rules in the group are similar in format to those described in Sect. 3.2.1,

Table 2 Sample rules for video to audio quality reducing adaptations

Rules	Antecedents If {bpsCurr...}								Consequents Then {v...}	
	bpsCurr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq	v	A
1	veryLo	veryHi	veryHi	veryHi	null	veryLo	none	belowAvg	null	veryLo
17	veryLo	avg	Avg	veryHi	null	veryLo	none	belowAvg	null	veryLo
35	veryLo	veryHi	veryHi	veryHi	null	Avg	none	belowAvg	null	avg
43	lo	aboveAvg	aboveAvg	veryHi	null	Lo	none	belowAvg	null	Lo
54	belowAvg	hi	Hi	hi	null	belowAvg	none	belowAvg	null	belowAvg
56	avg	veryHi	veryHi	veryHi	null	Avg	none	belowAvg	null	avg

and similarly, are broken up into four subsets. Samples from these four subsets are shown in Table 3.

3.2.3 Quality increase from audio to video

Rules in TeleMorph were designed to impose greater restrictions on adaptations involving increases in presentation quality than decreases. 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 usability.

The third grouping of rules in TeleMorph comprises cross-modality adaptations from audio to video which result in an increased presentation quality. These rules can be seen in Table 4. There are 52 rules in total in this group, which can be divided into four subsets, each with a lower value for 'bpsCurr', and each dictating under what parameters an audio presentation should adapt to video. Rules 113, 132 and 144 below are deal with adaptation scenarios when 'bpsCurr' is 'veryHi', whereas rules 145, 160 and 165 manage adaptations when bpsCurr is 'hi', 'aboveAvg' and 'avg' respectively.

3.2.4 Quality increase from video to audio

The last group of adaptation rules includes those which control cross modality adaptations from video to audio where the presentation's quality is increased, due to fluctuating bandwidth. Table 5 depicts six samples from this group of rules, with those numbered 166, 185 and 200 having a 'veryHi' value for the 'bpsCurr' antecedent and those numbered 201, 216 and 221 having 'hi', 'aboveAvg' and 'avg' values for 'bpsCurr' respectively.

3.3 TeleMorph FIS outputs

TeleMorph's output fuzzy sets represent the modalities which are available to the presentation system. However, as this prototype of TeleMorph deals solely with adaptations between video and audio, only two output fuzzy sets 'v' and 'a' are used. Video and audio output sets are named 'v' and 'a'. Following the implication process using the 221 rules in the FIS, the 221 resulting output sets are aggregated to one fuzzy set. This is followed by defuzzification, using the centroid method, of the two aggregated output sets to give two crisp values for 'v' and 'a'. Only one of TeleMorph's outputs is a whole number whilst the remaining output is always 0, e.g. v = 0.812, a = 0. These output values represent the quality at which the chosen modality (video or audio) 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 datarate (required bandwidth) closest to 0.812 of the maximum bandwidth will be used within the presentation. Both outputs can also be zero meaning that no cross modality adaptation should occur. The membership functions within TeleMorph's two output fuzzy sets are shown in Fig. 11.

4 TeleTuras

The following two sections detail TeleTuras, an example application developed as a testbed for TeleMorph's 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. Specifically TeleTuras integrates 40 multimedia and bandwidth specific test scenarios which were created manually to facilitate

Table 3 Sample rules for audio to video quality reducing adaptations

Rules	Antecedents If{bpsCurr...}								Consquents Then{v...}	
	bps Curr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq	V	a
57	veryLo	null	veryLo	none	veryHi	veryHi	veryHi	avg	veryLo	null
76	veryLo	null	VeryLo	none	avg	Avg	avg	avg	VeryLo	null
91	veryLo	null	Avg	none	veryHi	veryHi	veryHi	avg	avg	null
92	lo	null	Lo	none	veryHi	veryHi	veryHi	avg	lo	null
107	belowAvg	null	belowAvg	none	veryHi	veryHi	veryHi	avg	belowAvg	null
112	avg	null	Avg	none	veryHi	veryHi	veryHi	avg	avg	null

Table 4 Sample rules for audio to video quality increasing adaptations

Rules	Antecedents If {bpsCurr …}								Consequents Then {v…}	
	bpsCurr	vCurr	vLowest -bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq	V	a
113	veryHi	null	none	veryHi	veryLo	veryLo	veryLo	avg	veryHi	null
132	veryHi	null	none	veryHi	avg	avg	avg	avg	veryHi	null
144	veryHi	null	none	avg	veryLo	veryLo	veryLo	avg	avg	null
145	hi	null	none	hi	veryLo	veryLo	veryLo	avg	hi	null
160	aboveAvg	null	none	aboveAvg	veryLo	veryLo	veryLo	avg	aboveAvg	null
165	avg	null	none	avg	veryLo	veryLo	veryLo	avg	avg	null

Table 5 Sample rules for video to audio quality increasing adaptations

| Rules	Antecedents If {bpsCurr...}									Consquents Then {v...}	
	bpsCurr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq		v	A
166	veryHi	veryLo	veryLo	veryLo	null	none	veryHi	hi		null	veryHi
185	veryHi	avg	avg	avg	null	none	veryHi	hi		null	veryHi
200	veryHi	veryLo	veryLo	veryLo	null	none	avg	hi		null	avg
201	hi	veryLo	veryLo	veryLo	null	none	hi	hi		null	hi
216	aboveAvg	veryLo	veryLo	veryLo	null	none	aboveAvg	hi		null	aboveAvg
221	avg	veryLo	veryLo	veryLo	null	none	avg	hi		null	avg
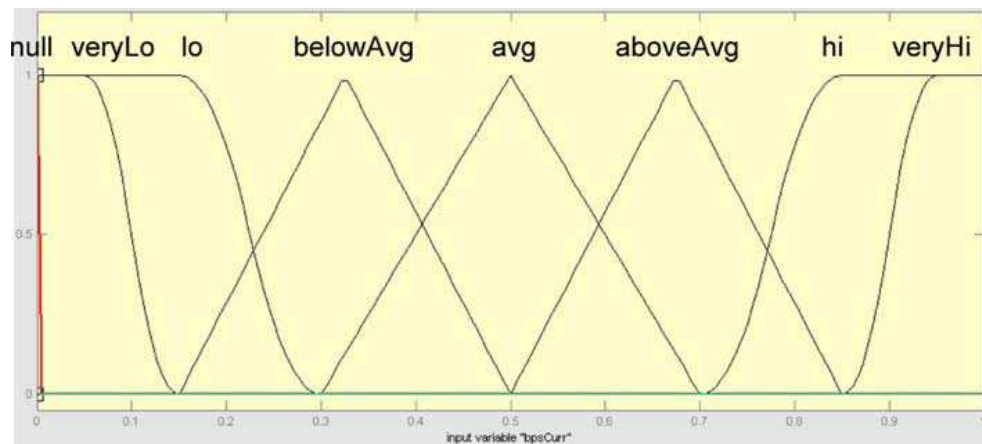
Fig. 11 Output fuzzy set MFs as used for audio and video

comprehensive testing of TeleMorph. TeleTuras enables input queries such as: "Take me to the Millennium forum?" "What is the GuildHall". Hence, TeleTuras provides tourists with instantaneous access to relevant information.

4.1 TeleTuras' GUI

In order to test, and demonstrate, TeleMorph's ability to control cross-modality adaptations, an example test application, TeleTuras, which is shown in Fig. 12, was created using Matlab's Graphical User Interface Development Environment (GUIDE) (Babuska 1993; DeMoyer and Mitchell 1999). GUIDE allows a developer to not only lay out a GUI easily using the GUIDE Layout Editor and program the logic to control the components, but one can also integrate ActiveX controls easily, thus enabling the integration of enhanced functionality. In TeleTuras Apple's QuickTime ActiveX control is used as a media client to output presentations.

This QuickTime control provides a media client which is used to present video and audio content as shown in Fig. 13 (a) and (b) respectively. Apple's QuickTime ActiveX control also provides access to multimedia details such as required datarate, duration, time position and timing scale which are necessary for decision making in TeleMorph's fuzzy inference system as outlined in Sect. 3.2. ActiveX controls are also used for the gauges that control current bandwidth (Fig. 5), video quality and audio quality (Fig. 14).

The 'current bandwidth' gauge (Fig. 5) has a range from 0 to 1 allowing it to represent the normalised 'bpsCurr' values as described in Sect. 3.1.1. The current pointer value of this gauge can be changed manually allowing for specific user-controlled testing of TeleMorph, but can also be changed automatically using sample network bandwidth data (e.g. data shown in Table 1). The values for current bandwidth and maximum bandwidth are shown above the current bandwidth gauge, as can be seen in Fig. 6 where the current data rate is 189,110 out of a maximum bandwidth of 301,168 bps. The gauges representing current quality of video and audio, as shown in Fig. 14, include pointers coloured red, blue and white that correspond to minimum, maximum and current media (video/audio) quality values respectively.

Only one of the pointers used to represent current media (i.e. audio and video) qualities can have a value greater than 0 as only one modality can be presented at any one time. Sometimes one pointer conceals another if their values are the same, as can be seen in Fig. 14(a) where the minimum quality pointer for audio is hidden behind the current quality pointer.
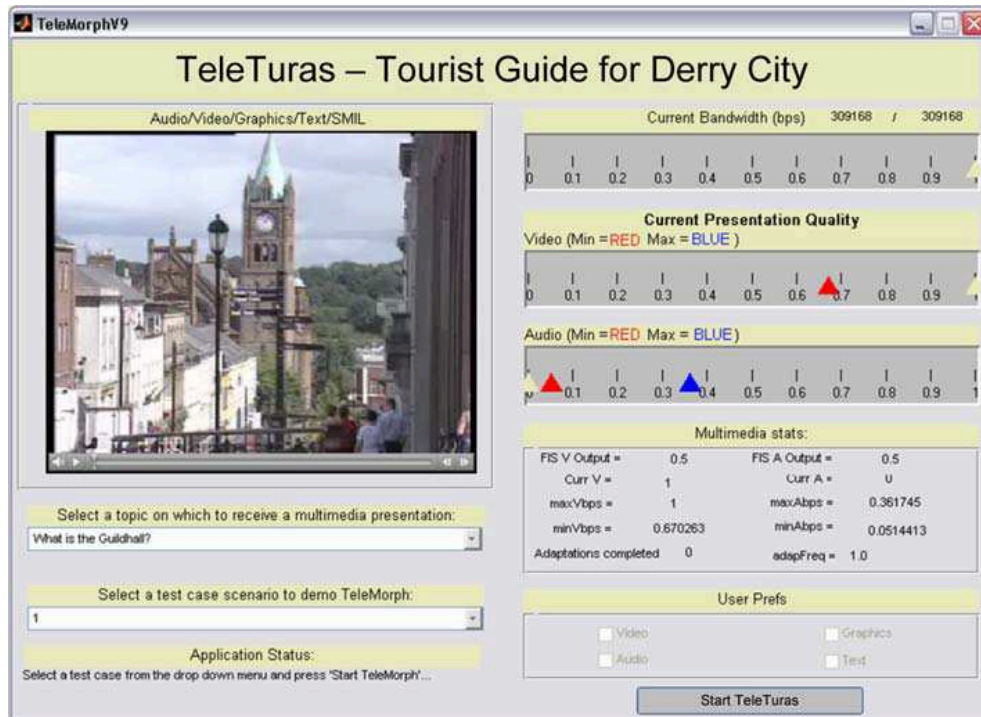
Fig. 12 TeleTuras' GUI

Fig. 13 Multimedia output in TeleTuras using Apple's QuickTime ActiveX control
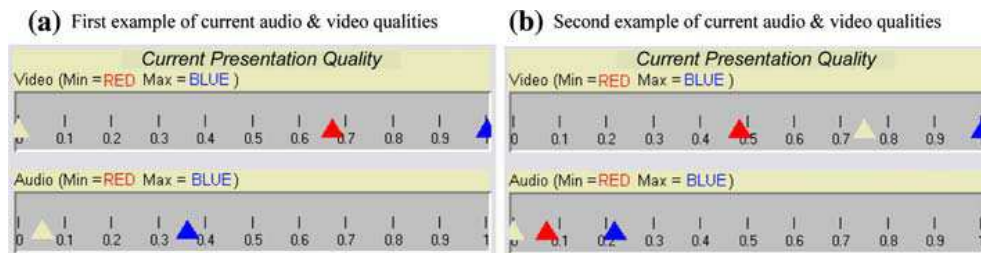

Fig. 14 ActiveX gauge control for video and audio qualities

These values for minimum and maximum media quality are derived from the minimum and maximum datarates required by the multimedia that is available to a presentation. The highest datarate for any one multimedia file (e.g. video with a datarate of 885 Kbps) is used to normalise all other media datarates against, i.e. if 885 Kbps is the datarate for a video file which is the maximum bandwidth required by one media file out of all the multimedia available to the presentation, it takes on a normalised quality value of 1, and thus an audio file with a datarate of 225 Kbps takes on a normalised value of 0.25 (i.e. 225/885).

4.2 TeleMorph's test scenarios

Forty test scenarios have been created to test TeleMorph's adaptation decisions within TeleTuras. The 'test case scenario' popup menu in TeleTuras allows for the selection of any one test case scenario during testing. Each test scenario contains a selection of audio and video files. The variations in qualities and ranges of multimedia available in each test case can be seen in Table A.1 in Appendix A. These test cases were designed to cover a comprehensive set of adaptation scenarios. They also provide test cases to detect bugs in TeleMorph's FIS whilst allowing a wide variation in presentation adaptations during future user acceptance testing. The 40 test scenarios are divided into similar logical groups as the rules within TeleMorph's FIS (Sect. 3.2). The first 20 test scenarios involve adaptations from audio to video where the presentation quality increases, as well as adaptations from video to audio where the presentation quality decreases. The remaining 20 scenarios demonstrate video to audio adaptations that result in improvements in presentation quality and audio to video adaptations that result in degraded presentation quality. Test scenarios 1–10 have the same range of audio files available to them with datarates of 8, 16, 32, 40, 48, and 56 Kbps, whilst the bandwidth range of the video files varies from 665 to 887 Kbps for the first scenario to 44–887 Kbps for the 10 scenario. All of the test scenarios are shown in detail in Table A.1 in Appendix A. Results and analysis of the adaptations carried out by TeleMorph during each test scenario are detailed in Sect. 5.

5 Evaluation of TeleMorph

This section provides a detailed objective analysis of the performance of TeleMorph's adaptation decisions in relation to the aforementioned test scenarios. This analysis includes a detailed description of what adaptations occurred during each of the 40 test scenarios. To facilitate this analysis, TeleTuras was used to create trace data for each test scenario including details of the timing and frequency of adaptations and as well as details of what input parameters brought these about. Specifically the trace data includes current bandwidth, current video, lowest and highest quality video available, current audio, lowest and highest quality audio available, the adaptation frequency variable, number of adaptations carried out during the presentation, and the two FIS output values representing the new qualities for video and audio following adaptation.

The test runs of TeleTuras were completed using manually generated sample bandwidth data. The video and audio files used had 55 s durations and various bandwidth ranges as shown in Table A.1 in Appendix A. The audio files were distilled versions of the video files (with the video track removed) which allowed for seamlessly synchronised adaptations between the two modalities. Bandwidth values were changed every 0.6 s using the sample data shown in Table 6. This data describes erratic bandwidth fluctuations which stress test all possible adaptation scenarios.

Table 6 Manually designed bandwidth test data

Normalised raw bandwidth data	Normalised bps range tested
0.5, 0.4, 0.3, 0.2, 0.1, 0,	$0.5 \rightarrow 0$
0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1,	$0 \rightarrow 1$
1, 0.9, 0.8, 0.7, 0.6, 0.5,	$1 \rightarrow 0.5$
0.5, 0.6, 0.7, 0.8, 0.9, 1,	$0.5 \rightarrow 1$
1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0,	$1 \rightarrow 0$
0, 0.1, 0.2, 0.3, 0.4, 0.5	$0 \rightarrow 0.5$

5.1 Test scenarios 1–10

The trace data recorded for test scenarios 1–10 is shown in Table 10. During test scenario one, three adaptations occurred. The first adaptation, which occurred after 5 s, was due to a bandwidth decrease from 0.3 to 0.2 and resulted in a presentation change from video at a quality of 1.0, to audio at 0.13197. This adaptation was facilitated by the fact that the adaptFreq prerequisite for audio to video presentation modifications with an increase in quality ('belowAvg') was fulfilled with a value of 0.29127 (see Sect. 3.1.4 for further details on the adaptFreq FIS input). TeleMorph's FIS rules which controlled this adaptation are shown in Table 7. It is possible to determine the rules associated with each adaptation using matlab's Fuzzy Logic toolbox Rule Viewer, which primarily serves to enable tests of what outputs are generated by specific inputs. The aggregated and defuzzified outputs from these three rules resulted in the audio output of 0.13197. From the second adaptation sample shown in Table 10 (in test scenario 1), one can see that the actual quality of audio which was incorporated into the output presentation following the first adaptation in test scenario one was 0.06343. This is due to the fact that outputs from TeleMorph's FIS are only as precise as the number of membership functions describing values within the output fuzzy sets. The multimedia presentation system which is utilising TeleMorph must determine the audio file which has the closest datarate (e.g. 0.06343) to the FIS output value (e.g. 0.13197).

The second adaptation during test scenario one happened 16 s into the presentation following a bandwidth increase from 0.7 to 0.8. This bandwidth fluctuation resulted in a cross modality adaptation from audio at a quality of 0.63, to video at 0.87941. The rules which controlled this adaptation are shown in Table 8.

The adaptations in test scenarios 2–4 vary slightly from those in test scenario one in that the video to audio changes resulted in lower qualities of output audio (0.07). This is because the bandwidth decrease which caused these adaptations in test scenarios 2–4 is lower (0.1) than the datarates involved in the video to audio modifications during test scenario one (0.2). The rules which caused the video to audio degradation adaptations which occurred during test scenarios 1–4 are shown in Table 9.

No adaptations were performed during test scenarios 5–10 as the values for 'vLowest-bps' in these scenarios were too close to the values for 'aHighest-bps'. This is because TeleMorph's rules have been designed to only allow adaptations when the quality transition across the Universe of Discourse is substantial, as previously stated in Sect. 3.2, e.g. from 'avg' quality to 'veryLo' or 'veryHi' to 'avg'.

By designing rules in this manner adaptation frequency is inherently restricted. This restriction controls the number of adaptations which can occur, hence reducing the impact adaptations have on the flow and continuity of the presentation. In Table 10 one can see that the values for 'vLowest-bps' decrease gradually from 0.74863 in scenario 1 to 0.048357 in

Table 7 Rules enabling the first adaptation in test scenario 1

Rules	Antecedents If {bpsCurr...}									Consequents Then {v...}	
	bpsCurr	vCurr	vLowest -bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq		v	A
38	lo	veryHi	aboveAvg	veryHi	Null	lo	none	belowAvg		null	lo
40	lo	hi	aboveAvg	veryHi	Null	lo	none	belowAvg		null	lo
42	Lo	hi	aboveAvg	Hi	Null	Lo	none	belowAvg		null	lo

Table 8 Rules enabling the second adaptation in test scenario 1

Rules	Antecedents If {bpsCurr...}								Consequents Then {v...}	
	bpsCurr	vCurr	vLowest -bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq	v	A
145	Hi	null	none	Hi	veryLo	veryLo	veryLo	Avg	hi	null
146	Hi	null	none	hi	veryLo	veryLo	lo	Avg	hi	null
148	Hi	null	none	hi	lo	veryLo	lo	Avg	hi	null
150	Hi	null	none	hi	lo	lo	lo	Avg	hi	null

Table 9 Rules controlling video to audio degradation adaptations in test scenarios 2–4

Rules	Antecedents If {bpsCurr...}									Consequents Then {v...}	
	bpsCurr	vCurr	vLowest -bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq		v	A
4	veryLo	veryHi	avg	veryHi	Null	veryLo	none	belowAvg		null	veryLo
7	veryLo	hi	avg	veryHi	Null	veryLo	none	belowAvg		null	veryLo
10	veryLo	hi	avg	hi	Null	veryLo	none	belowAvg		null	veryLo

scenario 10. In test scenario 8, 9 and 10 the lowest quality of video is very similar to the highest quality for audio, thus inter-modality video adaptation or transcoding would be preferential here, as apposed to a cross-modality adaptation from video to audio. Specifically, the current, lowest and highest values for audio in Table 10 all fall in the 'veryLo' membership functions (max value 0.15) of their respective fuzzy sets, thus a value of 'avg' (0.3) or greater for 'vLowest-bps' is required (as is the case in test scenarios 1–4) before an adaptation can occur. When the values for 'vLowest-bps' go below 0.3 (e.g. 0.23492 in test scenario 5) adaptations no longer occur as is the case in test scenarios 5–10 (depicted in Table 10).

5.2 Test scenarios 11–20

Over test scenarios 11–20 the value for 'aHighest-bps' increases gradually across the values 0.018, 0.045, 0.063, 0.090, 0.126, 0.181, 0.217, 0.253, 0.29, 0.362, whilst all other input values remain static. The resulting adaptations for test scenarios 11–20 are the same as those described in Sect. 5.1 for test scenario 1, as are the rules which brought about these adaptations (as shown in Figs. 2 and 3). The fact that adaptations which resulted across test scenarios 1–20 showed similarities acts as evidence of the consistency with which TeleMorph makes cross modality decisions. The adaptation traces recorded during test scenarios 11–20 are shown in Table 11.

5.3 Test scenarios 21–30

The adaptations which occurred during the presentation of test scenarios 21–28 further go to demonstrate the consistency of TeleMorph's decision making, as increases in presentation quality involve the same output value (0.87941) as the quality increasing adaptations in test scenarios 1–20, although the roles of audio and video are reversed with audio representing the higher quality modality in test scenarios 21–28 (as in all scenarios from 21 to 40). Thus, the rules which control the video to audio quality increasing adaptations in test scenarios 21–28 are the same as those shown in Table 8.

The first and third adaptations in test scenarios 21–28, as shown in rows 1 and 3 in Table 12, which resulted in degradations from audio to video at qualities of 0.14437 and 0.12569 respectively, are associated with the rules shown in Table 13. The audio to video adaptations for test scenarios 29 and 30, as shown in Table 11, are also controlled by the rules depicted in Table 13. The video to audio quality increasing adaptation which occurs during test scenario 29 (where Audio = 0.93111) is as a result of the rule detailed in Table 13.

5.4 Test scenarios 31–40

The modifications which were made to the output multimedia presentation during test scenarios 31–37, as detailed in Table 14, reflect those described previously in Sects. 5.1–5.3. As stated in Sect. 5.1 TeleMorph's rules have been designed to only allow adaptations which represent a substantial and worthwhile cross-modality transition as regards changes to the output presentation's quality. This accounts for the lack of adaptations in test scenarios 38–40.

The 40 test scenario evaluations have shown a great degree of consistency in how TeleMorph controls cross-modality adaptations. This is essential so that decisions involving increases and decreases in presentation quality display similar patterns for both audio to video adaptations and vice versa. Also one can see from the traces in Tables 5–7 and 10 that the FIS was successful in carrying out cross-modality adaptations in a cautious, prudent, and controlled manner.

Table 10 Cross-modality adaptation traces from test scenarios 1–10

Test scenario	bpsCurr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	Adapt. freq.	Number adapts	Video	Audio
1	0.2	1	0.74863	1	0	0.009015	0.06343	0.29127	0	0	0.13197
	0.8	0	0.74863	1	0.06343	0.009015	0.06343	1.2207	1	0.87941	0
	0.2	1	0.74863	1	0	0.009015	0.06343	3.6868	2	0	0.13197
2	0.1	1	0.48301	1	0	0.009069	0.063494	0.396	0	0	0.071212
	0.8	0	0.48301	1	0.063494	0.009069	0.063494	1.0678	1	0.87941	0
	0.1	1	0.48301	1	0	0.009069	0.063494	3.9625	2	0	0.071212
3	0.1	1	0.35049	1	0.063494	0.009069	0.063494	0.38873	0	0	0.074021
	0.8	0	0.35049	1	0	0.009069	0.063494	1.0478	1	0.87941	0
	0.1	1	0.35049	1	0	0.009069	0.063494	3.9534	2	0	0.074021
4	0.1	1	0.3025	1	0	0.009069	0.063494	0.37691	0	0	0.075
	0.8	0	0.3025	1	0.063494	0.009069	0.063494	1.0515	1	0.87941	0
	0.1	1	0.3025	1	0	0.009069	0.063494	3.9214	2	0	0.075
5–10	0.3	1	0.23492	1	0	0.009069	0.063494	5.0664	0	0	0

Table 11 Cross-modality adaptation traces from test scenarios 11–20

Test scenario	bpsCurr	vCurr	vLowest -bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	Adapt. freq	Number adapts	Video	Audio
11–20	0.2	1	0.74863	1	0	0.009015	0.01803	0.29318	0	0	0.13197
	0.8	0	0.74863	1	0.01803	0.009015	0.01803	1.2135	1	0.87941	0
	0.2	1	0.74863	1	0	0.009015	0.01803	3.678	2	0	0.13197

Table 12 Cross-modality adaptation traces from test scenarios 21–30

Test scenario	bpsCurr	vCurr	vLowest -bps	vHighest -bps	aCurr	aLowest -bps	aHighest -bps	Adapt. Freq.	Number Adapts	Video	Audio
21–28	0.1	0	0.13334	0.15167	1	0.80024	1	0.36591	0	0.14437	0
	0.8	0.14425	0.13334	0.15167	0	0.80024	1	1.0571	1	0	0.87941
	0.2	0	0.13334	0.15167	0.80024	0.80024	1	3.332	2	0.12569	0
	0.8	0.13334	0.13334	0.15167	0	0.80024	1	5.6959	3	0	0.87941
29	0.1	0	0.13334	0.64779	1	0.80024	1	0.36509	0	0.14459	0
	0.9	0.14425	0.13334	0.64779	0	0.80024	1	1.2118	1	0	0.93111
	0.2	0	0.13334	0.64779	1	0.80024	1	3.1091	2	0.12569	0
30	0.1	0	0.13334	0.76992	1	0.80024	1	0.36682	0	0.14413	0

Table 13 Audio to video degradation adaptation rules as used in test scenarios 21–30

Rules	Antecedents If {bpsCurr...}									Consequents Then {v...}	
	bpsCurr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq		v	A
78	veryLo	null	lo	none	veryHi	hi	veryHi	avg		lo	null
79	veryLo	null	lo	none	veryHi	aboveAvg	veryHi	avg		lo	null
80	veryLo	null	lo	none	hi	hi	veryHi	avg		lo	null
81	veryLo	null	lo	none	veryHi	aboveAvg	veryHi	avg		lo	null
82	veryLo	null	lo	none	veryHi	hi	hi	avg		lo	null
83	veryLo	null	lo	none	veryHi	aboveAvg	hi	avg		lo	null
93	lo	null	lo	none	veryHi	hi	veryHi	avg		lo	null
94	lo	null	lo	none	veryHi	aboveAvg	veryHi	avg		lo	null
95	Lo	null	lo	none	Hi	hi	veryHi	avg		lo	null
96	lo	null	lo	none	veryHi	aboveAvg	veryHi	avg		lo	null
97	lo	null	lo	none	veryHi	hi	hi	avg		Lo	null
98	lo	null	lo	none	veryHi	aboveAvg	hi	avg		Lo	null

Table 14 Video to audio adaptation rules as utilised in test scenario 29

Rules	Antecedents If {bpsCurr…}								Consequents Then {v…}	
	bpsCurr	vCurr	vLowest - bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	adaptFreq	V	A
175	veryHi	lo	lo	avg	null	none	veryHi	hi	null	veryHi

Table 15 Cross-modality adaptation traces from test scenarios 31–40

Test Scenario	bpsCurr	vCurr	vLowest-bps	vHighest-bps	aCurr	aLowest-bps	aHighest-bps	Adapt. freq.	Number adapts	Video	Audio
31	0.8	0.15167	0.13334	0.16783	0	0.80024	1	1.2316	1	0	0.87941
	0.2	0	0.13334	0.16783	0.80024	0.80024	1	3.3707	2	0.12569	0
	0.8	0.133334	0.13334	0.16783	0	0.80024	1	5.7141	3	0	0.87941
32	0.1	0	0.13334	0.16783	1	0.7002	1	0.36691	0	0.14411	0
	0.8	0.144425	0.13334	0.16783	0	0.7002	1	1.0605	1	0	0.87941
	0.2	0	0.13334	0.16783	0.80024	0.7002	1	3.3432	2	0.12569	0
	0.8	0.13334	0.13334	0.16783	0	0.7002	1	5.65	3	0	0.87941
33	0.1	0	0.13334	0.16783	1	0.60017	1	0.35509	0	0.14667	0
	0.8	0.14425	0.13334	0.16783	0	0.60017	1	1.0515	1	0	0.87941
	0.2	0	0.13334	0.16783	0.80024	0.60017	1	3.3432	2	0.13219	0
	0.8	0.13334	0.13334	0.16783	0	0.60017	1	5.7232	3	0	0.87941
34–37	0.1	0	0.13334	0.16783	1	0.49991	1	0.37236	0	0.075	0
	0.8	0.13334	0.13334	0.16783	0	0.49991	1	1.0696	1	0	0.87941
	0.1	0	0.13334	0.16783	0.80024	0.49991	1	3.5368	2	0.075	0
	0.8	0.13334	0.13334	0.16783	0	0.49991	1	5.295	3	0	0.87941
38–40	0.2	0	0.13334	0.16783	1	0.17508	1	4.5912	0	0.5	0.5

Also it is noted that the time which passed between adaptations averages at 13, 13, 9, and 16 s for the four test scenario groups (1–10, 11–20, 21–30, 31–40) respectively, more than surpassing the 11 s minimum stated early in Sect. 3.1.4. This figure will be analysed further in future subjective tests of applications utilising TeleMorph for cross modality adaptation decision making, such as TeleTuras.

Further objective tests of TeleMorph will occur following integration of other inputs including—predicted bandwidth, user preferences and packet loss. In order to evaluate TeleMorph's outputs subjectively future tests using TeleTuras will be carried out involving end users. This will comprise user's feedback and opinions on the frequency of cross modality adaptations resulting from TeleMorph's decision making, as well as presentation effectiveness evaluations, compared with similar metrics for presentation systems which do not incorporate cross-modality adaptations.

6 Relation to other work

Substantial research has been carried out to address the specific problems which exist in the area of mobile intelligent multimedia presentation systems. A variety of mobile intelligent multimedia systems were presented in Sect. 2.2 such as SmartKom (Wahlster 2003), DEEP MAP (Malaka 2001), CRUMPET (Zipf and Malaka 2001), VoiceLog (Bers et al. 1998), MUST (Almeida et al. 2002), EMBASSI (Hildebrand 2000) and QuickSet (Oviatt et al. 2000). Though all of these systems focus on some aspect of mobile intelligent multimedia presentation, none have specifically addressed the problem of dynamically transmoding based on network constraints, and in particular available network bandwidth, as TeleMorph has. Pedersen and Larsen (2003) describe a test system which analyses the effect of user acceptance when output modalities are changed automatically or are changed manually by the users, however this research was based on a user base of only six test subjects and no final system was ever developed as part of the project.

Other research which tackles the problem of cross modality adaptation have primarily focused on multimedia document models with inherent substitution or alternative semantics. Boll et al. (1999) proposed Cardio-OP (Gallery of Cardiac Surgery) which utilised a cross-media adaptation and substitution model which focused on maintaining the semantics of a presentation's information flow using semantically equivalent media alternatives. Boll et al. also briefly mention the importance of on-the-fly adaptations resulting from unpredictable changes in the user-context, and the importance of maintaining the presentation's information flow during any such content alterations. Despite this brief mention of on-the-fly adaptation no attention is given to the effect that an uncontrolled frequency of cross-modality adaptations has on the usability of the output presentation during continuous dynamic adaptations, e.g. continuously fluctuating bandwidth. ZyX (Boll and Klas 2001), Tiempo (Wirag 1999) and REVEAL THIS (Pastra 2006) were also detailed in Sect. 2.2 as document models providing some element of flexibility and adaptation, although the control and execution of these adaptations were not dealt with in any of these projects.

Thang et al. (2005a) propose a content value model which 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. If, using the technique proposed by Thang et al. (2005a), the resource(s) controlling content adaptation fluctuate around a "conversion boundary" the result would

be continuous cross-modality conversions between the adjacent (multi)modality qualities. Practical adaptation evaluations have not been presented using this adaptation model. As a result of these shortcomings this method is inadequate to deal with the indeterminable bandwidth variations which occur on a network connecting a mobile multimedia presentation system. TeleMorph's adaptation thresholds, which are integrated within fuzzy sets as membership functions, and resulting adaptation decision are dynamically modified based on the frequency of cross-modality adaptations thus enabling the system to deal with the indeterminable bandwidth variations.

7 Conclusion and future work

For the purpose of efficient and effective cross-modality adaptation we have presented a platform, TeleMorph, which was implemented using fuzzy logic. Also we have detailed a testbed application, TeleTuras, which provides tourist information for the city of Derry in Northern Ireland. TeleMorph addresses a number of questions associated with controlling cross-modality adaptations including:

(1)   When bandwidth decreases, should a cross-modality adaptation occur as opposed to further content scaling?
(2)   To what extent must bandwidth fall below the lowest acceptable quality for a modality before a cross-modality adaptation is required (bearing in mind that if bandwidth decreases only a small amount, below the lowest acceptable quality it is just as likely to increase by that amount again within a short amount of time)?
(3)   What is the most effective and acceptable destination modality given the current presentation composition and content quality, as well as the bandwidth which is currently available?
(4)   At what fidelity/quality should each new modality be incorporated into the presentation?
(5)   What is the maximum acceptable frequency of cross-modality adaptations, and how does this vary based on the current and destination modalities involved in an adaptation?
(6)   What direct effect does type and frequency of cross-modality adaptations have on presentation effectiveness and user acceptance?

TeleMorph addresses (1) above in the way in which its rules have been designed; that is, to only allow adaptations which represent a substantial and worthwhile cross-modality transition as regards changes to the output presentation's quality (represented by the bandwidth requirement of the media). If the decision engine determines that a cross-modality adaptation is not suitable then alternatively further inter-modality scaling may occur. Initial evaluations of TeleMorph's performance regarding this problem have been presented. The platform also addresses (2) above, regarding adaptation control when the bandwidth available is close to a potential adaptation threshold. Fuzzy logic handles this very effectively in that no static value represents this threshold value, but a combination of thresholds, which are implicit in the 221 fuzzy rules which have been defined. The most effective destination modality as described in (3) is determined by using a combination of output from work on the cognitive load theory by Elting et al. (2002), the dual coding theory (Uden and Campion 2000), and rules which use the 'adaptFreq' fuzzy input to determine the inclination to adapt from linguistic (text, speech audio) to non-linguistic content (images, video, non-speech audio), and vice versa. The problem of maintaining the semantic flow of information during cross-modality adaptations is currently resolved in TeleMorph by using multimedia derived from

one high quality video file (audio–visual). With respect to (4) above, the two outputs from TeleMorph's fuzzy inference system specify the quality at which new media elements (currently audio and video) should be incorporated into the adapted presentation. TeleMorph addresses cross-modality adaptation frequency as outlined in (5) inherently by incorporating an input 'adaptFreq' into the fuzzy inference system which controls this aspect of transmoding. The test scenario evaluations which have been carried out prove the efficiency with which TeleMorph manages this problem. The challenge presented by (6) above regarding the effect that type and frequency of cross-modality adaptations have on presentation effectiveness and user acceptance is something which will be addressed during future subjective evaluations of the platform.

TeleMorph differs from other approaches 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's output presentation composition will incorporate images and text also, thus allowing for extended adaptation between video, audio, images and text, as well as multimodal combinations of these media elements (e.g. text & images). To enable this, additional inputs and outputs will be incorporated into TeleMorph's FIS. Also, rules to process these inputs to the output space will be created. Images will be presented at various fidelities in a similar fashion to audio and video, and will be included in presentations along with audio or text. Text will be used in combination with audio, as well as images or video to deliver multimodal presentations. Alternatively text will be utilised in unimodal presentations as a base line representation of the semantic content when network throughput is very low and higher quality modalities are unfeasible. TeleMorph exploits "auxiliary inputs", such as 'adaptFreq', which compliment the basic media-specific inputs. These additional inputs enable more intelligence adaptation decision making in TeleMorph, based on comprehensive contextual information. "Auxiliary inputs" that will be incorporated into future versions of TeleMorph include: bandwidth prediction, user preferences, adaptation follow-up flexibility of the media, packet loss, and system response time/latency. These inputs will enhance TeleMorph's ability to provide more intelligent and fully informed decisions to applications such as TeleTuras and will be subsequently evaluated using statistical user evaluation and further objective analysis. Testbed applications other than TeleTuras will be explored, focusing in particular on the medical and educational domains.

Appendix A: Variations in test scenario multimedia datarates

Forty test scenarios have been created to test TeleMorph's adaptation decisions within TeleTuras. Each test scenario contains a selection of audio and video files. The variations in qualities and ranges of multimedia available in each scenario can be seen in Table A.1.

Table A.1 Video and audio qualities used in TeleMorph's 40 test scenarios

Test case scenarios	Available audio bandwidth range		Available video bandwidth range	
	Kbps	Normalised	Kbps	Normalised
Test A → V Quality increases &V → A quality decreases				
1 (Audio static)	8 → 56	0.009 → 0.063	665 → 887	0.748 → 1
2	8 → 56	0.009 → 0.063	431 → 887	0.483 → 1
3	8 → 56	0.009 → 0.063	314 → 887	0.350 → 1
4	8 → 56	0.009 → 0.063	271 → 887	0.302 → 1
5	8 → 56	0.009 → 0.063	212 → 887	0.234 → 1
6	8 → 56	0.009 → 0.063	162 → 887	0.178 → 1
7	8 → 56	0.009 → 0.063	109 → 887	0.120 → 1
8	8 → 56	0.009 → 0.063	79 → 887	0.088 → 1
9	8 → 56	0.009 → 0.063	57 → 887	0.062 → 1
10	8 → 56	0.009 → 0.063	44 → 887	0.048 → 1
11 (Video static)	8 → 16	0.009 → 0.018	665 → 887	0.748 → 1
12	8 → 40	0.009 → 0.045	665 → 887	0.748 → 1
13	8 → 56	0.009 → 0.063	665 → 887	0.748 → 1
14	8 → 80	0.009 → 0.090	665 → 887	0.748 → 1
15	8 → 112	0.009 → 0.126	665 → 887	0.748 → 1
16	8 → 160	0.009 → 0.181	665 → 887	0.748 → 1
17	8 → 192	0.009 → 0.217	665 → 887	0.748 → 1
18	8 → 224	0.009 → 0.253	665 → 887	0.748 → 1
19	8 → 256	0.009 → 0.29	665 → 887	0.748 → 1
20	8 → 320	0.009 → 0.362	665 → 887	0.748 → 1
Test V → A Quality increases & A → V quality decreases				
21 (Audio static)	256 → 320	0.79 → 1	44 → 50	0.133 → 0.15
22	256 → *320*	0.79 → 1	44 → 66	0.133 → 0.199
23	256 → *320*	0.79 → 1	44 → 79	0.133 → 0.24
24	256 → *320*	0.79 → 1	44 → 94	0.133 → 0.28
25	256 → *320*	0.79 → 1	44 → 109	0.133 → 0.33
26	256 → *320*	0.79 → 1	44 → 122	0.133 → 0.37
27	256 → *320*	0.79 → 1	44 → 133	0.133 → 0.4
28	256 → *320*	0.79 → 1	44 → 162	0.133 → 0.49
29	256 → *320*	0.79 → 1	44 → 212	0.133 → 0.64
30	256 → *320*	0.79 → 1	44 → 251	0.133 → 0.77
31 (Video static)	256 → *320*	0.79 → 1	44 → 55	0.133 → 0.16
32	244 → *320*	0.69 → 1	44 → 55	0.133 → 0.16
33	192 → *320*	0.59 → 1	44 → 55	0.133 → 0.16
34	160 → *320*	0.49 → 1	44 → 55	0.133 → 0.16
35	128 → *320*	0.39 → 1	44 → 55	0.133 → 0.16
36	112 → *320*	0.34 → 1	44 → 55	0.133 → 0.16
37	96 → *320*	0.3 → 1	44 → 55	0.133 → 0.16
38	80 → *320*	0.25 → 1	44 → 55	0.133 → 0.16
39	64 → *320*	0.2 → 1	44 → 55	0.133 → 0.16
40	56 → *320*	0.175 → 1	44 → 55	0.133 → 0.16

References

Almeida L, Amdal I, Beires N, Boualem M, Boves L, den Os E, Filoche P, Gomes R, Knudsen JE, Kvale K, Rugelbak J, Tallec C, Warakagoda N (2002) Implementing and evaluating a multimodal and multilingual tourist guide. International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems Copenhagen, Denmark, 28–29 June

Babuska R (1993) Fuzzy toolbox for MATLAB. In: Proceedings of the 2nd IMACS International Symposium on Mathematical and Intelligent Models in System Simulation. University Libre de Bruxelles, Brussels

Bernsen NO (1994) Modality theory in support of multimodal interface design. In: Proceedings of the spring symposium on intelligent multi-media multi-modal systems. Stanford University, USA, pp 37–44

Bers J, Miller S, Makhoul J (1998) Designing conversational interfaces with multimodal interaction. In: Proceedings of the DARPA broadcast news transcription and understanding workshop. Lansdowne Conference Resort, Lansdowne, Virginia, pp 88–92

Boll S, Klas W (2001) ZYX-A multimedia document model for reuse and adaptation of multimedia content. IEEE Transa Knowledge Data Eng, 13(3):361–382

Boll S, Klas W, Wandel J (1999) A cross-media adaptation strategy for multimedia presentations. In: Proceedings of ACM Multimedia. ACM, Orlando, pp 37–46

Chalmers D, Sloman M (1999a) QoS and context awareness for mobile computing. In: Proceedings of international symposium on handheld and ubiquitous computing (HUC99). Springer-verlag, Karlsruhe, pp 380–382

Chalmers D, Sloman M (1999b) A survey of quality of service in mobile computing environments. IEEE Commun Surv 2(2):2–10

Clark JM, Paivio A (1991) Dual coding theory and education. Educ Psychol Rev, 3(3):149–210

Crumpet (2006) http://www.ist-crumpet.org Site visited 28 July, 2006

Curran K, Annesley S (2005) Transcoding media for bandwidth constrained mobile devices. Int J Netw Manage 15(2):75–88

DeMoyer R, Mitchell EE (1999) Use of the MATLAB graphical user interface development environment for some control system applications. In: Proceedings of IEEE Computer Society Conference on Frontiers in Education. Stripes Publishing L.L.C, San Juan, pp 7–11

Elting C, Zwickel J, Malaka R (2002) Device-dependant modality selection for user-interfaces-an empirical study. In: Proceedings of Sixth International Conference on Intelligent User Interfaces. ACM, San Francisco, pp 55–62

EML (2006) http://www.eml.org/english/Research/Memory Site visited 28 July, 2006

Fox A, Brewer EA (1996) Reducing WWW latency and bandwidth requirements by real-time distillation. In: Proceedings of the Fifth International World Wide Web Conference. Elsevier, Paris, France, pp 1445–1456

Fox A, Gribble SD, Brewer EA, Amir E (1996) Adapting to network and client variability via on-demand dynamic distillation. In: Proceedings 7th International Conference on Architectural Support for Programming Languages and Operating Systems. ACM, Cambridge, pp 160–170

Fox A, Gribble SD, Chawathe Y, Brewer EA (1998) Adapting to network and client variation using infrastructural proxies: lessons and perspectives. IEEE Pers Communi 5(4):10–19

Hildebrand A (2000) EMBASSI: electronic multimedia and service assistance. In: Proceedings IMC'2000. Rostock-Warnemunde, Germany, November, pp 50–59

ISO/IEC 21000–7 2004, Information Technology—Multimedia Framework—Part 7: Digital Item Adaptation, ISO/IEC

Klas W, Christian G, Reinhard F (1999) Cardio-OP: gallery of cardiac surgery. In: Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, Volume 2, June 07–11, pp 1092–1095

Kurkowski S, Camp T, Mushell N, Colagrosso M (2005) A visualization and analysis tool for NS-2 wireless simulations: iNSpect. In: Proceedings of the 13th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. IEEE Computing Society, Atlanta, pp 503–506

Malaka R (2000) Artificial intelligence goes mobile — artificial intelligence in mobile systems 2000, Workshop in conjunction with ECAI 2000. Berlin, Germany, August 22, Workshop Notes, pp 5–6

Malaka R (2001) Multi-modal interaction in private environments. International Seminar on Coordination and Fusion in MultiModal Interaction, Schloss Dagstuhl International Conference and Research Center for Computer Science. Wadern, Saarland, Germany 29 October–2 November. (http://www.dfki.de/~wahlster/Dagstuhl_Multi_Modality/Dagstuhl-2001.pdf Site visited 28 July, 2006)

Malaka R, Zipf A (2000) DEEP MAP—challenging IT research in the framework of a tourist information system. In: Fesenmaier D, Klein S, Buhalis D (eds) Information and communication technologies in tourism 2000, Proceedings of ENTER 2000, 7th International Congress on Tourism and Communications Technologies in Tourism, Barcelona, Spain. Springer Computer Science, Wien, pp 15–27

Malaka R, Porzel R, Zipf A (2000) Integration of smart components for building your personal mobile guide. In: Malaka R (ed) Artificial intelligence in mobile systems—AIMS2000. Workshop in conjunction with ECAI 2000, Berlin, Germany, August 22, Workshop Notes, pp 22–26

Oviatt SL, Cohen PR, Wu L, Vergo J, Duncan E, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J, Ferro D (2000) Designing the user interface for multimodal speech and gesture applications: state-of-

the-art systems and research directions. Hum comput Interact 15:263–322. (to be reprinted in Carroll J (ed) Human–computer interaction in the new millennium. Addison-Wesley Press, Boston, 2001)

Pastra K (2006) Beyond multimedia integration: corpora and annotations for cross-media decision mechanisms. In: Proceedings of the 5th Language Resources and Evaluation Conference (LREC)

Pedersen JS, Larsen SR (2003) A pilot study in modality shifting due to changing network conditions. M.Sc. Thesis, Center for Person Communication. Aalborg University, Denmark

Pereira F, Smith JR, Vetro A (2005) Introduction to the special section on MPEG-21. IEEE Trans Multimedia 7(3):397–399

Qiu Q, Zhang D, Ma J (2004) GPRS network simulation model in NS-2. 2004 Joint Conference of the 10th Asia-Pacific Conference on Communications and the 5th International Symposium on Multi-Dimensional Mobile Communications Proceedings (APCC/MDMC). Institute of Electrical and Electronics Engineers Inc., New York, NY 10016–5997, United States, Beijing, China, pp 700–704

Smith JR (2001) Universal multimedia access Multimedia Systems and Applications. International Society for Optical Engineering, Boston 21–32

Solon AJ, Mc Kevitt P, Curran KJ (2004a) A bandwidth determined mobile multimodal system. 1st International Conference on E-Business and Telecommunication Networks (ICETE). INSTICC Press, pp 24–31

Solon AJ, Mc Kevitt P, Curran KJ (2004b) Design of a tourist driven bandwidth determined MultiModal Mobile Presentation System. IEEE International Conference on Mobility Aware Technologies and Applications (MATA '04), Florianopolis, Brazil, October 20–22. Lecture Notes in Computer Science (LNCS) Vol. 3284/2004, ISBN 3–540–23423–3, Heidelberg, Germany: Springer-Verlag GmbH, February, pp. 331–339

Solon AJ, Mc Kevitt P, Curran KJ (2004c) Mobile multimodal presentation. ACM Multimedia 2004—Proceedings of the 12th ACM International Conference on Multimedia. Association for Computing Machinery, New York, pp 440–443

Solon AJ, Mc Kevitt P, Curran KJ (2004d) TeleMorph: bandwidth determined Mobile MultiModal Presentation. In: Information Technology and Tourism (IT&T), vol 7, no 1, (ISSN 1098–3058), Cognizant Publishers, USA, June, pp 33–47

Sun H, Vetro A, Asai K (2003) Resource adaptation based on MPEG-21 usage environment descriptions. International Symposium on Circuits and Systems (ISCAS), IEEE, Bangkok, Thailand, pp 536–539

Thang TC, Jung YJ, Ro YM (2003) Modality conversion in content adaptation for universal multimedia access. International Conference on Imaging Science, Systems and Technology (CISST). CSREA Press, Las Vegas, NV, USA, pp 434–440

Thang TC, Jung YJ, Lee J, Ro YM (2005a) Effective adaptation of multimedia documents with modality conversion. Signal Process Image Communi 20(5):413–434

Thang TC, Jung YJ, Ro YM (2005b) Modality conversion for QoS management in universal multimedia access. IEE Proc Vis Image Signal Process 152(3):374–384

Uden L, Campion R (2000) Integrating modality theory in educational multimedia design. Proceedings of the 17th Annual Conference of the Australasian Soceity for Computers in Learning in Tertiary Education (ASCILITE 2000), Learning to Choose and Choosing to Learn. Coff's Harbour, Australia, December 11–14

Vetro A, Timmerer C (2005) Digital item adaptation: overview of standardization and research activities. IEEE Transa Multimedia 7(3):418–426

Vetro A, Timmerer C, Devillers S (2005) Digital item adaptation. In: Burnett IS, Pereira F, Van de Walle R, Koenen R (eds) The MPEG-21 Book. Wiley

Wahlster W (2003) SmartKom: symmetric multimodality in an adaptive and reusable dialogue shell. In: Krahl R, Günther D (eds) Proceedings of the Human Computer Interaction Status Conference 2003. DLR, Berlin, June, pp 47–62

Wirag S (1999) Scheduling of adaptive multimedia documents. In: Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS). IEEE Computing Society, Florence, Italy, pp 307–311

Zadeh LA (1965) Fuzzy sets. Inf Control, 8:338–353

Zipf A, Malaka R (2001) Developing Location Based Services (LBS) for tourism—the service providers view. ENTER 2001, In: Proceedings of the 8th International Congress on Tourism and Communications Technologies in Tourism, Montreal, Canada, April, pp 24–27