Server Transcoding of Multimedia Data for Cross-Disability Access

Vidhya Balasubramanian and Nalini Venkatasubramanian

Information and Computer Science Department
University of California, Irvine, Irvine California-92697-3425, USA
{vbalasub,nalini}@ics.uci.edu

ABSTRACT

Multimedia transmission over wide-area networks currently only considers the server and network resource constraints and client device capabilities. It is also essential that the accessibility of the multimedia content for different users with diverse capabilities and disabilities be considered. In this paper we develop a transcoding technique to present the multimedia content to suit diverse disabled user groups by using an ability based classification approach. Using ability-based prioritization, the appropriate alternate modalities and quality levels are chosen to replace the inaccessible modalities. The transcoding process allows for refinements to cater to specific types and degrees of impairments. Our performance results illustrate the benefits of the ability-based transcoding approach.

Keywords: Ability-Driven, transcoding, adaptation, cross-disability, modalities

1.INTRODUCTION

The recent years have seen a proliferation of multimedia information and applications. Some of these applications like distance learning are of high potential to users with disabilities. The integration of different data types into applications offers more flexibility and interactivity. While this represents unprecedented opportunities it is also essential to ensure that the resulting benefits also enrich the lives of the disabled, and the elderly, by making information more accessible. To date, application developers and service providers have focused on extensive functionality and optimal performance (possibly, driven by commercial needs) to ensure that customers can be delivered the best possible Quality of Service (QoS) with the least possible resource consumption overhead, often ignoring the goal of universal accessibility. For example, increased network bandwidth and real-time video capture technologies may prompt transmission of a live event using streaming video but such a media is of little use to the blind. The goal of the SUGA project [SUGA02] is to enhance existing infrastructures to account for user abilities (and disabilities) so that personalized content can be delivered in a cost-effective manner.

A major step in the realization of the "universal access" goal is that of a transcoding framework that supports personalized (cross-disability) access to information. Transcoding refers to the adaptation of multimedia information to enhance usability and manage the heterogeneity. Content adaptation refers to the modification of the parameters of a specific media type (for instance, an image can be encoded using various resolutions, color levels and intensities). There are multiple facets to developing such a framework foremost of which is a generalized representation of content to allow access to individuals with varying disabilities. Another critical issue is that of cost-effective mechanisms to achieve the transcoding. Specialized mechanisms built into the specialized adaptive user-interfaces and access devices for each and every service may not represent the optimal usage of resources. In the envisioned approach, discussed in the paper, user abilities are considered, to provide the best alternate modality of information to improve the accessibility of the information. For instance closed captioning could be provided in place of audio, to a person with a severe hearing impairment. By taking advantage of the information about the end-user ability to optimize resource utilization without sacrificing on the utility of the delivered content, the number of users admitted into the system can be increased. For example, video objects in many distance-learning sites without proper captioning are of little use to a user who is blind. Instead, video/image content can be transmitted as a text summary that can be accessed by screen readers. The bandwidth saved can be used to send some other modalities that may be relevant to other users, thereby admitting a larger number of users into the system.

Our focus is on cross-disability access that caters to a wide range of users with different disabilities, including the blind and visually impaired, deaf and hearing impaired, cognitively impaired, motor impaired and the deaf-blind. In order to provide the relevant media types, and modalities, we use information about the ability that an end user possesses to determine which content to deliver to specific clients. A rule-base is computed using the different user abilities based on which the transcoding is performed. Based on the user profile and system resource constraints, appropriate modalities are selected so as to maximize the utility of the information communicated to the user with disabilities. The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes the ability based transcoding framework. It addresses how the priority values for the multiple modalities are calibrated, describes the architectural layout of the transcoding system and introduces the priority based transcoding algorithm. We also present a refinement of the transcoding process for vision impairments using existing studies about perception of vision. Section 4 evaluates the performance of the proposed transcoding mechanisms. We conclude in Section 5 with future research directions.

2.RELATED WORK

Problems of accessibility apply to both hardware and software systems. An individual who cannot use her hands or her eyes cannot use a mouse, type on a keyboard, see a screen, or navigate easily through a user guide. Software components such as operating systems and application programs include many roadblocks to users with disabilities. There have been many different solutions proposed to tackle the problem of accessibility. Most universal accessibility solutions have been studied in the interface level, by improving the software interfaces to suit the users [K98]. Other works address the use of assistive technology such as access equipment and input devices, and show how such tools can be made more popular by alleviating some of the accessibility issues [JV01]. Universal Design techniques have been applied to information technologies to make all information perceivable, including keys and controls, facilitate navigation, and ensure compatibility with commonly used assistive technology [JV01, SS01]. Adding the adaptation functionality in the interface and specialized devices is very important for improving accessibility. However with more and more information being accessed over networks, it is more efficient and flexible to put the adaptation functionality in an intermediate layer in the network, and not burden the client.

Blindness is a widely studied disability issue; tools such as screen readers, audio HTML interfaces and auditory navigation are proposed techniques to alleviate some problems faced by users with impaired vision [TA00]. Image exploration and accessing graphical user interfaces are the focus of research in developing accessible interfaces for the blind [RM96]. Systems catering to users with hearing-impairments deal with recognition of sign languages, online generation of closed captions, and voice to text recognition capabilities. [HH99]. Some of the proposed solutions to alleviate problems faced by users with motor impairments are use of gestures, force-feedback technology, eye tracking and vocal programming [KR98]. For the benefit of cognitive and language impaired users certain solutions proposed are context-aware assistance, like hints, descriptions etc and natural language processing [CD98,CP02]. Most of these techniques described are tailored to suit individual disabilities and do not cater to the general population of users with varied disabilities. In our proposed work we cater to cross-disability access, by providing a generic solution to a variety of disability groups, which is also tailored to suit individual preferences. While this provides flexibility, it is also cost effective as a wider range of users is catered to.

Accessibility to multimedia on the web is increasingly an important area of research [ZV01] since most content is embellished with multimodal data. One way to make multimodal information accessible to users with disabilities is by transcoding the information appropriately [HS00, TA00]. Transcoding techniques have been applied to adapt to client device heterogeneity [MS99], to adapt to bandwidth heterogeneity [HS01], to ensure that some level of QoS is maintained despite network fluctuations [SEV00] etc. Object based transcoding of multimedia information [VW01] presents another approach to content adaptation where objects that constitute a multimedia scene are identified and prioritized. While many of these transcoding techniques help to adapt multimedia information for varying conditions, they do not consider the capability of the user to access the information. Our framework aims to precisely do that, by adapting these transcoding techniques to suit the problem at hand of adapting multimedia information for accessibility over a cross disability platform.

Designing a transcoding system includes many design challenges [VG94, SGV95]. Techniques have been proposed to address placement of media on disk to ensure real-time retrieval [AO92]; similarly admission control procedures to maximize server throughput and buffer management policies to minimize memory requirements have been developed. Replication and striping strategies or optimizing storage across disk arrays are described in [TP93]. Data placement techniques help to efficiently transcode the multimedia data without imposing undue load on the system [BS00, MW97]. However data placement in the context of our problem should consider the pattern of user accesses to efficiently support adaptation for cross-disability access. The trade-offs between storage space and transfer bandwidth are highlighted in [DS95]. A number of broadcast and multicast techniques have been studied for effective utilization of network bandwidth [GK98] for video-on-demand systems. The above-mentioned design considerations cater to the network parameters, and quality of service in mind. In order to effectively build the adaptation framework these design techniques should be appropriately modified so as to also cater to the user accessibility.

3.AN ABILITY DRIVEN APPROACH TO TRANSCODING

Several approaches can be used for the purpose of quantifying the accessibility of the information provided to the user. We analyze the different approaches that exist to model user abilities, and describe their relative advantages and disadvantages. This will be followed by a detailed description of an ability driven approach to quantify the accessibility of the provided media type to the particular user.

The cognitive sciences community has conducted abundant studies for characterizing the abilities of humans. One approach is the Ability Requirements approach adopted from Fleishman and Quaintance [FQ84]. Several abilities, required for performance of tasks have been identified and classified into cognitive, perceptual, physical and psychomotor abilities as shown in Table 2. In this approach tasks are described, contrasted and compared in terms of the abilities that a given task requires of an individual performer. Ability rating scales have been developed and validated after a lot of experiments and systematic studies. (see Ability Requirements Approach, Chapter 12 in FQ84] . Human tasks are classified based on a seven-point scale using definitions of high and low ability requirements as scale anchors. It has been proved that this method of rating tasks provides a statistically reliable tool for assessing amount of ability requirement. Table 1 gives one such example of tasks representing the ability 'Near Vision ', adopted from FQ84-Appendix C.

Another approach for quantifying the ability to assimilate information would be to use the individual mathematical models of human vision systems, auditory systems etc [SW99]. The vision models help calibrate visual acuity, based on illumination, distance and size of object, etc. These models help quantify the distortion of image as viewed by the user, giving better view of the perceptual distortion than the regular Mean Square Error Metric. In our application, quantification of the ability to assimilate the information helps in selecting the most appropriate form of information to the user.

Specific models of the human vision and auditory system are more perceptual in nature, they tend to ignore the motor and cognitive aspects involved in information assimilation, and hence are more useful for applications requiring mainly perceptual ability. The ability requirements strategy is a more wholesome approach, since it applies not only to information assimilation, but also to interactive computing, which requires more motor skills. This approach is also helpful when considering people with multiple impairments, a factor missing in other models. Furthermore, the ability-requirements approach also accounts for the abilities that are primarily affected as well as other abilities that are affected as a result. For instance, even a mild hearing loss of (25 – 40dB) might result in mild language retardation [JV01].

After considering the relative advantages of these different approaches, we decided to use the ability classification approach to derive initial accessibility information. We use the specific mathematical models of human faculties (vision, auditory etc.) to further fine-tune the transcoding process and improve accessibilities (See Section 3.3). The following paragraphs describe how we use this ability-requirements approach to derive the accessibility of the modality for a specific user class.

Task Items for the Selected ability– Near Vision	Mean	Standard Deviation
Read the fine print of legal journals	5.89	0.90
Cut and mount color film transparencies	4.67	0.97
Plug in a TV set	1.44	0.62

Table 1: Ability levels required for different tasks (Near Vision)

***Prioritization of Modalities*:** The first step is to find the ability required to perform the task of assimilation of each of these modalities, by a normal user. The following modalities are considered, audio, video, text and image. The alternate modalities presented in case these are not accessible are, closed-captioned audio, captioned video, audio description of video, tactile text (we do not send the text in tactile form, but assume that the user has the appropriate device, like Braille reader to view the text in tactile form),text summary, summary of audio and video, text description of image and audio description of image.

Cognitive Abilities	Perceptual Abilities	Physical Abilities	Psychomotor Abilities
1.Oral Expression	1.Speed of Closure	1.Static Strength	1.Control Precision
2.Written Expression	2.Flexibility of Closure	2.Dynamic Strength	2.Multilimb Coordination
3.Oral Comprehension	3.Spatial Orientation	3.Explosive Strength	3.Response Orientation
4.Written Comprehension	4.Visualization	4.Trunk Strength	4.Rate control
5.Fluency of Ideas	5.Perceptual Speed	5.Extent Flexibility	5.Reaction Time
6.Originality	6.Near Vision	6.Dynamic Flexibility	6.Arm-Hand Steadiness
7.Memorization	7.Far Vision	7.Gross Body Coordination	7.Manual Dexterity
8.Problem Sensitivity	8.Visual Color Discrimination	8.Gross Body Equilibrium	8.Finger Dexterity
9.Mathematical Reasoning	9.Night Vision	9.Stamina	9.Wrist-Finger Speed
10.Number Facility	10.Peripheral Vision		10.Speed of Limb Movement
11.Deductive Reasoning	11.Depth Perception		11.Selective Attention
12.Inductive Reasoning	12.Glare Sensitivity		12.Time Sharing
13.Information Ordering	13.General Hearing		
14.Category Flexibility	14.Auditory Attention		
15.Speech Hearing	15.Sound Localization		
16.Speech Clarity			

Table 2: Ability Classification from Fleishman and Quaintance

For each of these modalities, we select the abilities that are applicable, to the information retrieval. For assimilation of video information, we do not consider abilities like trunk strength, dynamic strength, etc. For the abilities selected, we decide the closest matching task given in the task classification from Fleishman and Quaintance[FQ84], to the application at hand. Let us consider the ability 'Near Vision' in table 1. It is defined as "the capacity to see close environmental surroundings" [Appendix B in FQ84]. Considering the small size of fonts of text data, we correlate the assimilation of text data to that of reading fine print of legal journals. This gives us the required ability of near vision to read text as 5.89. Similarly the values for other abilities for normally accessing that modality are calibrated.

Let $A_1, A_2, .. A_n$, be the abilities selected in that order.

Let $W_1, W_2, .. W_n$ be the weight associated, that is the value required for each of these selected abilities for a specific task. Then the total ability required by a normal user to access this modality of information is given by

$$AB_{n,m} = \sum_{i=1}^{n} W_i A_i$$

The next step is to calculate the total ability the impaired users possess to access the above information. The different user groups considered are blind, low-vision, deaf, hard of hearing, motor impaired, cognitively impaired and deaf-blind. Based on different statistics on these impairments, and how these different levels of impairments affect the information processing [LH02,WHO2,SS01], we try to correlate with the tasks selected as required for the normal user, the maximum possible task the disabled user can perform based on this disability. Consider the same ability as before, i.e., near vision. The required task level was to read fine print. For a blind person (no light perception), even the minimum task possible cannot be performed, relying on this ability. Hence this ability is absent and is assigned a value zero. A vision-impaired user might be able to perform the task of plugging a TV set(considering the worst case) , but with difficulty. We calculate the value of that ability the user possesses using the mean and standard deviation values from [FQ84] (see sample values in Table 1). This yields a value of 1.3 as the amount of near-vision a severely vision impaired user possesses to access fine print. For specific vision impairments we can refine these values, based on the vision models and visual acuity values of different levels of vision impairments.

For a given disability category the ability possessed to access this information is calculated as follows. For the selected abilities A_1, A_2, … A_n, for the modality, let U_1, U_2, $U_3 \ldots U_n$, be the corresponding weights denoting the corresponding amount of that ability the user possesses (the upper limit being that required by that of a normal user). The total ability the user possesses to access the particular modality m is given by

$$AB_{d,m} = \sum_{i=1}^{n} U_i A_i$$

The accessibility of the modality to the user group is defined by the accessibility ratio value that is calculated as follows.

$$AR = \frac{\sum \textit{Values of abilities the user possesses}}{\sum \textit{Values of abilities required to access the modality}}$$

$$= \frac{AB_{d,m}}{AB_{n,m}}$$

Once the accessibility ratio values are calibrated, an accessibility matrix is developed where the cells indicate the accessibility of the corresponding modality to the person with the disability in the corresponding row. Table 3 below gives a sample accessibility matrix. An extended version of this accessibility matrix is also created for the alternate modalities like closed captioned audio, tactile text and auto-summarization.

Disability	Audio	Video	Text	Image
Blind	0.990	0.063	0.172	0.145
Low Vision	0.990	0.608	0.554	0.669
Deaf	0.090	1.000	0.996	1.000
Hardof Hearing	0.575	1.000	0.996	1.000
Cognitive	0.685	0.830	0.776	0.888
Motor	0.903	0.833	0.751	0.870
Deaf blind	0.090	0.063	0.172	0.145

Table 3: Accessibility Matrix showing the accessibility of the particular modality to different user groups

3.1.System Architecture

The architecture of proposed transcoding system is shown in Figure 1. The essential components of the proposed transcoding system are as follows: a) Content Source b) Transcoding Meta-data c) Fine-Grained Transcoder d) Coarse-Grained Transcoder

Content Source: The content source is the place where the information content is stored. All the other modules of the transcoding system access the information from the content source. As and when the content adaptation is done on-line or offline by the content adapter the information in the content source is updated. The information in the content source can be accessed via the transcoding meta-data where different media types at different qualities are represented.

Transcoding Meta-data: The transcoding meta-data represents the following information:

1. Accessibility information: An accessibility matrix containing the accessibility of each of these modalities to each of the user groups is created based on the application and the abilities required to access the information. This accessibility matrix encodes the utility of each modality *(and quality level)* to each user class for that specific application.

2. Media related information: This part of the transcoding meta-data represents the information about the media at different quality levels. Specifically, it stores information about the different quality parameters for each media-type (e.g. color, intensity, resolution for images, sampling rate for audio, frame rate for video etc.) and encodes the importance of each parameter to that media type. The selection of appropriate information content for the particular user is based on both the media information and the accessibility value – the accessibility value defined earlier is refined based on the quality level.

3. Resource requirements of different media types at different quality levels: This component represents the resource requirements for the different media types at the different quality levels. Resources required include storage, network bandwidth, memory and CPU resources. To represent resource requirements of different modalities at different quality levels, we will leverage work on efficient multidimensional data structures [ZM99] to accurately represent resource information and derive the resource needs for a particular user session. A simple example is a storage matrix, which represents the storage required for the media type, say an image at different resolutions and different color encoding schemes. We can also use other parameters to efficiently utilize system resources. Detailed resource models have been studied [LL99, RL97]. Though it is beyond the scope of this paper to be developing extensive resource models, we will use simplistic resource considerations. Detailed models will be used for future work.

Coarse-Grained Transcoder: This module performs coarse grained-transcoding, which is primarily modality selection. Here the appropriate modalities are selected from the content source based on the meta-data information and accessibility matrix. The modalities are selected so that they maximize utility to the user while minimizing system load and resource requirements.

Fine-Grained Transcoder: This is the module where the finer granularity of transcoding is performed. The transcoding could result in the creation of new modalities such as close-captioned audio, video summarization (if they do not already exist) or new quality levels e.g. adapted images or video with changes in resolution, frame rate and deletion of irrelevant components. Adaptation of quality levels for specific media types is largely dependent on the nature and extent of the disability. A specialized treatment of transcoding for the vision-impaired is discussed in Section 3.3. We assume that content for fine-grained transcoding is created offline.

The above architecture follows the following operational dynamics. When a request for the information comes in the user profile is checked with the existing user profiles stored. Any new profiles are added. The client profile is checked to calculate the client capabilities and available client resources are recorded for that client. The coarse grained transcoder picks the most appropriate content from the content source for that data requested, based on the accessibility values in the meta-data calculated, and the resources in the client and server. More fine-grained adaptation is provided by the fine-grained transcoder. This content is created based on the accessibility values and other parameters in the meta-data. The fine- grained transcoding is done offline and stored in the content source for future use. The coarse grained transcoder finally transmits the adapted content to the client.

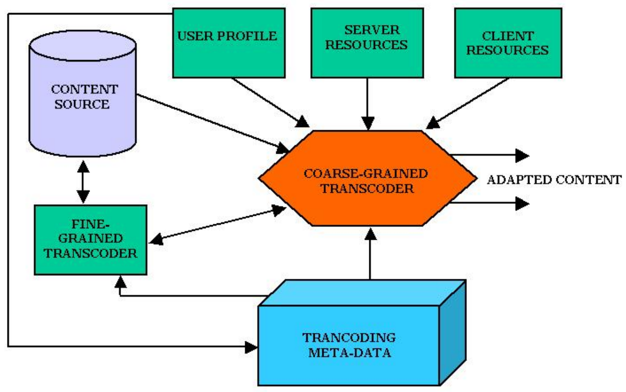

Fig 1: Overall System Architecture

3.2.An Ability Driven Transcoding Algorithm

To effectively transcode the information we use a priority based transcoding algorithm that uses accessibility information from the accessibility matrix described above along with resource constraints within the environment to prioritize content that may be delivered for each incoming request. The transcoding technique employs an iterative procedure to select the appropriate modalities. When a request comes in, the following factors are considered – (i) User profile (consisting of the user abilities); (ii) Server Resources, (iii) Client Resources and (iv) Resources required by the request. Based on the user profile the appropriate accessibility is determined from the accessibility matrix. The resource consumption for each of the modalities at various resolutions is represented in a resource consumption matrix. The value R_{jk} at each cell represents the value of the resources required for the particular modality j, at the resolution k. The resources can include server bandwidth, link bandwidth, storage, CPU, etc. The primary bottleneck resource we have considered is server bandwidth since the paper does not focus on CPU-intensive online transcoding and compression techniques. The QRAM model provides heuristics for optimal allocation of resources to applications with different QoS requirements [RL97, RL98]. The ability driven resource allocation problem at hand is similar to the QoS Based resource allocation, however with accessibility requirements instead of QoS requirements. The QRAM model works on the fact that the quality and hence utility increases with increased allocation of resources. However accessibility does not increase with increased resources allocated, (infact defining an analytical relation between accessibility and resources is not straightforward). Hence adapting such detailed resource allocation schemes in the context of maximizing accessibility is a complex problem and is beyond scope of this paper. Here we consider a simple resource allocation scheme with the primary goal of providing maximum accessibility to the user that is based on a greedy strategy. According to the accessibility requirement of the user, the resources are allocated. If the resources are constrained the modalities with the best accessibility requirements that is satisfiable is provided.

Given the following parameters:

- M_{jk} represents the index to the modality j at the resolution k
- P_{ijk}, which gives the ability based priority value for the particular user group i for accessing M_{jk}
- P_{ij}^{min} is the minimum accessibility value required for modality j as required by user group i. For now we assume the user provides the minimum requirements. Calculating the minimum accessibility requirements will require detailed user studies that will be done as part of future work.
- P_{ij}^{max} is the maximum accessibility beyond which there is no significant improvement in user experience of accessing modality j. This value also requires user studies and for now we assume the maximum possible accessibility as that for a normal user (which is 1.0 per modality).
- R_{jk}, the resource requirement for M_{jk}
- Available server resource R_s, and
- Client resource R_c.
- ε is the factor by which priority value is reduced in consecutive iterations. ε can be calculated based on the accuracy requirements. One simple assignment would be to partition the interval between P_{ij}^{min} and P_{ij}^{max} into small intervals then assign the size of the interval to ε.

$$\varepsilon = (P_{ij}^{max} - P_{ij}^{min}) / 100$$

The objective of the transcoding process can be stated as a maximization problem with the following specifications: Choose M_{jk} to maximize ΣP_{ijk} subject to the following constraints

$$\Sigma(P_{ijk}.R_{jk}) <= R_s$$
$$\Sigma(P_{ijk}.R_{jk}) <= R_c$$

In other words the objective is to maximize the total priority value so that the total storage consumption is less than available server resources R_s, and client resources R_c. A greedy generic, algorithm for the above maximization problem is given in Figure 2.

From the meta-data, initially the modalities with accessibility values above the threshold value, for the particular user group are chosen. Admission control is then performed at the server end where the request can either be admitted or rejected, based on resource availability. If the required resources exceed the available resources, then we change the resolution of the modalities. To do this we pick the same modalities with changed resolution from the meta-data, which gives the maximum accessibility value and can be satisfied by currently available resources at the server or client end. The admission control process is repeated. If the request is still not admissible, the least accessible modality is dropped and the process is repeated again.

```
While there is a user request:
        Select all modalities with P_ij > P_ij^min from priority
        matrix
          If admissible by server // first come first served ·
admission control
        If within resource availability
          Send the appropriate modalities requested
     Else
     Repeat
          Select the modalities from the info-cube so
that P_ijk is maximized
        If  (R_jk <R_s) && (R_jk<R_c)
           Send the information to the user
        Else If (resources exist)
             Reduce priority of least priority modality by ε
            Else drop the least priority modality.
        End If
     Until the request is admitted or no modalities remain
     Reject request if any P_ij <  P_ij^min
     Else Admit request with selected modalities
    End If
End While
```

Figure 2: The Ability Driven Transcoding Algorithm

3.3.Specialized Transcoding for Vision Impairments – A Case Study

In this case study we consider different vision disorders that exclude legal blindness, where transcoding techniques are applied to change the quality of the image and text information. The quality in this context encompasses resolution, color contrast, brightness, and font size. The transcoding is done to adapt to the varying perception levels of vision that a user possesses. For effective transcoding it is necessary to have a good understanding of the human visual system and its functions. According to the different studies on the human visual system, [AO02,CS02,VC02-1,WH01] effective visual perception is dependent on a number of factors. These factors are as follows:

Visual Acuity (V): It is the ability to resolve a spatial pattern separated by a visual angle (α) of 1 minute of arc. In the context of textual information, it is the ability to resolve contrast on a white background that subtends 5 minutes of arc. Using the common Snellen fraction, visual acuity is 20/20 for a normal user

$$V = 1/a \quad \text{where} \quad \alpha = h/d * 3438 \text{ minutes}$$

where h and d are the height and viewing distance of the media information.

Contrast Sensitivity (S): It is a measure of the limit of visibility of low contrast patterns before they become indistinguishable from a uniform field. It is the function of the coarseness/fineness of image features or spatial frequency. It is defined as the reciprocal of contrast.

$$S = 1/\text{contrast, where contrast} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} corresponds to the maximum intensity level and I_{min} the minimum intensity level of the media.

Brightness (B): Brightness is the subjective evaluation or interpretation of the amount of light reaching the visual system. For simplicity it can be related to the luminance of the image, which is the product of light, reflected and light reaching the surface. It varies logarithmically with luminance. For a RGB image where, R, G, B are the pixel intensity values for red, green and blue pixels brightness is calculated by averaging the luminance for all the pixels.

$$\text{Luminance} = 0.3R + 0.59G + 0.11B$$

Color Perception (C): For an arbitrary image the three responses are X, Y, Z (also called tristimulus values) for red, green and blue, can be calculate from the RGB color values as follows

$$X = 0.6067 R + 0.1736 G + 0.2001 B,$$
$$Y = 0.2988 R + 0.5868 G + 0.1143 B,$$
$$Z = 0.0000 R + 0.0661 G + 1.1149 B$$

Visual Field (θ): the visual field is the total area where objects can be seen in the peripheral vision while the eye is focused on a central point. Each eye has a horizontal field of view of 150 deg (60 deg towards the nose, 90 deg to the side) and a vertical field of view of 120 deg. A visual field of 20 deg indicates blindness. Visual field affects the visual angle of the user, which influences the amount of information received by the eye.

Hence the visual accessibility VA can be defined as a function of the above factors.

$$VA = f(V,S,B,C,\theta) = \text{norm}(V) + \text{norm}(S) + \text{norm}(B) + \text{norm}(C) + \text{norm}(\theta)$$

Here norm(i) is a normalized value for each of these factors, and 0<=norm(i)<=1. For simplicity, we first normalize each of these factors to fit into a uniform scale and sum the normalized values to obtain a measure of visual accessibility. Our goal is to maximize the visual accessibility for users with varying degrees and types of vision impairments. The normalization process is specific to the parameter;

Impairment	Affected Part of eye and effect	Factor affected	Transcoding required
Refractive errors (near vision, far vision, presbiopia)	Lens which results in poor focus and hence blurred image	Visual acuity (v) Norm(v) = v Other factors unaffected by transcoding	For given acuity v and size of image or text y Y = size /v New norm(v) = y*v/size
Cataract and Corneal Pathology	Affects lens resulting in hazy, distorted vision	Visual acuity (v), contrast sensitivity Norm(v) =v; Norm(s)= 1-1/s Other factors unaffected	For given acuity v and size of image or text y Y = size /v For given contrast sensitivity s, original contrast c1, the required new contrast is 1/s. Choose appropriate contrast level c2 closest to 1/s. New norm(v)=y*v/size; New norm(c)= c2;
Diabetic Retinopathy and Macular Degeneration	Macula, retina and fovea resulting in loss of detail, distortion and reduced color perception	Visual acuity (v), contrast sensitivity (s), color(c) Normalization for v and s same as above Norm(c) = (X+Y+Z)/3	Transcoding for factors v and s same as above For given responses X,Y,Z, RGB color values calculated and closest color level chosen. New norm(c) = Total sensitivity of Selected colors/no of colors
Glaucoma, Retinis Pigmentosa	Retina, optic nerve, resulting in poor peripheral vision and brightness perception	Field of View (θ), brightness (b) Norm(θ)= θ / 60; Norm(b) = 1-1/b	For field θ and size of image in x dimension, new size = $\frac{\theta * x}{60}$; Norm(θ) = ((x*θ)/ new size)/60 For brightness b the selected intensity level is e^b; For selected intensity I new norm(b) = 1-(1/logI)
Color Blindness	Cones, can affect red, green or blue cones	Color Perception (c) Norm(c)= (X+Y+Z)/3	For given responses X,Y,Z, RGB color values calculated and closest color level chosen. New norm(c) = Total Sensitivity of selected colors/no of colors

Table 4: Transcoding for vision impairments using detailed vision studies.

we reduce each factor to a numerical value between 0 and 1 where 0 represents a high degree of impairment and 1 represents the value of that factor for a normal unimpaired user. While visual acuity (low in users with refractive errors like myopia) can be directly represented without further calibration, the normalized value of contrast sensitivity is calculated as 1-1/S, where S is the measured contrast sensitivity for a user. For color sensitivity the normalized values is the average of the individual sensitivities to red, blue and green colors.

There are also intricate dependencies between the different factors. We model the dependency of brightness on contrast - i.e. as contrast increases the perceived brightness increases. Though other factors like luminance also affect brightness we do not model these dependencies due to complexities involved in changing luminance. To simplify the transcoding process, we only model those dependencies that have a significant impact on visual perception. Furthermore, the transcoding decisions may conflict in the case when multiple visual factors are impacted. For instance, a low visual acuity requires enlargement of image/text while a low field of vision will trigger a size reduction. The transcoding system must contain rules to handle such conflicts; for instance, in the above case, a bounded enlargement factor to suit the feasible field of view must be calculated.

The effect of different impairments with respect to the above factors is given in [SAP, RI02]. Based on the information we decide the level of transcoding to be performed for text and image data. Table 4 gives the general level of transcoding required for different impairments. For simplicity we consider only a limited number of resolution levels and a limited number of color-coding representations. For e.g. we consider the maximum size of an image to be 1024x768, beyond which a magnification is not feasible. Similarly the minimum size of the image or text is also defined. Similarly the color levels considered are Black and White images, 16 colors, 256 colors and 24 bit colors. From the available representations the appropriate quality level for the specific impairment is chosen based on the rules defined in the table 4.

4. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the ability driven transcoding scheme for a variety of disabilities and under varying resource constraints.

***System Model*:** The basic server configuration consists of a single data source, with storage of 50 GB. We assume that the server has enough storage to hold all the required modalities for the users considered. The server has a transfer bandwidth of 100Mbits/second in the basic configuration. For simplicity, CPU and memory resources of the data sources are not considered as bottlenecks. The data source consists of a single video with resolution – 640x480, and a frame rate of 30Frames per second. The audio has the following characteristics, 44KHz sampling rate, and 32 bits per sample with two channels. The data source additionally has text associated with each clip of video. The average duration of each session is 30 to 60 minutes. The alternate modalities associated are, video captioning and summarization (text and tactile), audio captioning (text, and tactile), summary of text, tactile representation of text and summary. We assume that these representations are generated and already stored along with the original piece of information.

Request Model: We assume the requests for the data follows a Zipf-like distribution and model the requests using the same. Accordingly, the popularity for the video follows Zipf's law, with the request arrivals per day for the data v given by: Probability (v_i is requested) $=K_V / i$ where,

$$K_V = \left(\sum_{1}^{V} \frac{1}{i} \right)^{-1}$$

The requests arrivals per hour are assumed to follow a Zipf-like distribution[BP95]. From the request arrivals-per day for the data and the probability distribution of requests each hour, the number of requests that arrive in each hour for the information is computed.

***User Model*:** We also model the user arrival, based on the demographics of the different types of user categories. Classification of the different disabilities, and the different demographic statistics has been based on the reports from the World Health Organization [WHO1]. The visual acuity values and hearing values , that help in deciding the user groups are derived from the different studies on these impairments. [WHO1, RL01]. Some studies on motor and cognitive impairments have been researched, but more towards these areas have to be studied. [CP02, SCI]. (Note, motor impairments here more specifically concerns spinal cord injuries, as it deals with limited use of upper limbs, other motor impairments are being studied to perform efficient transcoding for motor impairments). Once all the above-mentioned information is collected, the user model is developed. The user profile includes, visual acuity, auditory perception, motor ability and cognitive ability, based on which the user is grouped into one of the existing classifications. A random mix of users is generated based on the demographic information. We assume that all the users fall into any of the categories defined.

4.1.Performance of the Ability Driven Transcoding Scheme

We study the performance of the proposed algorithm under varying bandwidth restrictions. We compare the performance of the ability-based transcoding algorithm with the following cases- one when the information is sent without any transcoding, and the other is the case where the best accessible information is sent (e.g., for a blind person we send audio, captioned video, and tactile text). The relative performances of these techniques are given by the graph in Figure 3. Note that this result is based on the general user population based on the demographics as previously discussed.

The results illustrate that by considering accessibility information alone (given by the Best Modalities policy) we can improve the overall bandwidth utilization; hence, the best-modalities policy performs marginally better than the case with no transcoding with respect to the acceptance rate. However this result is not indicative of the improvements in the accessibility to the users. On the other hand, the ability based transcoding algorithm significantly outperforms the other strategies, especially under low bandwidth conditions. The irregularity in the curve is due to the random arrival of user groups, and the fact that the accessibility of information is not forsaken while considering the resource restrictions. Note that, if the accessibility drops below ε, the request is rejected.

Effects of Load on Accessibility: In this section we study the effect of the number of user requests at a time, on the accessibility of the information. The set of graphs below show the accessibility values, plotted as a function of user requests. We can see from the graphs that the accessibility is very low when information is not transcoded, and drops to zero soon, due to the resource intensive nature of information provided. On the other hand, due to transcoding the accessibility is maximized, and it is gradually reduced to admit the user, under resource restrictions. When this value falls below a threshold, which varies for different user groups, the request is rejected. Hence we can see that the proposed algorithm tries to maintain accessibility even under heavily loaded conditions, while also reducing the number of rejects. We consider each disability category separately to analyze how the accessibility is affected due to load. The maximum accessibility attainable from given information is 3.0 in our experimental case.

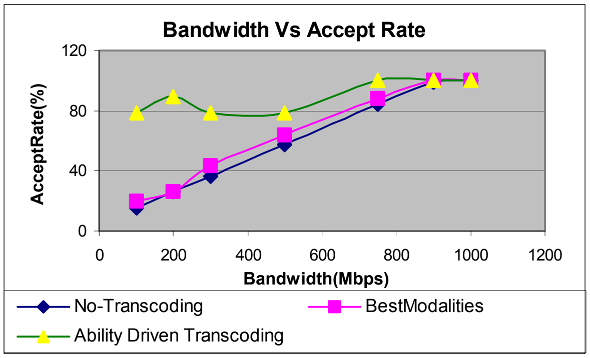

Figure 3: Graph showing acceptance rate for different bandwidths

Accessibility for Blind Users: From the graph in Figure 4 we can see that by sending the original data without transcoding, the maximum total accessibility achieved is just a little over 2.0. On the other hand by transcoding we can reach a maximum accessibility of nearly 3. The lower resource requirement of the transcoded information (containing audio, and tactile captioning and summary) also plays a pivotal role in maintaining this accessibility, even if the load is high. After a point the accessibility suddenly drops to zero (indicating rejects), attributed to the limited number of alternate modalities available.

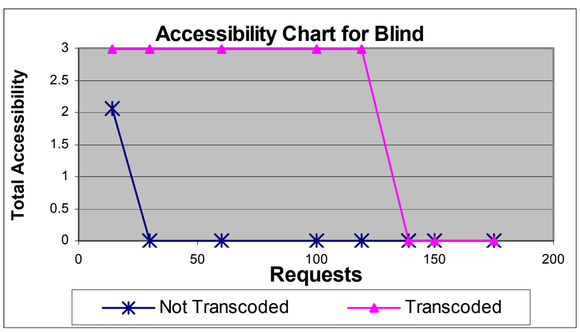

Fig 4: Effect of Load on Accessibility for users with blindness

Accessibility for Motor Impaired: We can see from the graph in Figure 5, that the originally sent modalities are almost as accessible as the transcoded information. This is because we have considered information access, which involves mainly perceptual abilities. Additionally modalities being highly textual in nature are less accessible compared to audio or video, which can be accessed without much motor involvement (assuming mot everyone have specialized devices to access information). These factors contribute to the drop in accessibility even at lower load levels.

Accessibility for deaf-blind: The only information accessible is tactile. As we send only tactile text, maximum accessibility is maintained throughout constantly at any load as in Figure 6. If we consider tactile images, we expect to see slight change in the trend followed, but this too depends on the available bandwidth and resources available. Generally the accessibility should not change much with moderate loads, as very little resources are required.

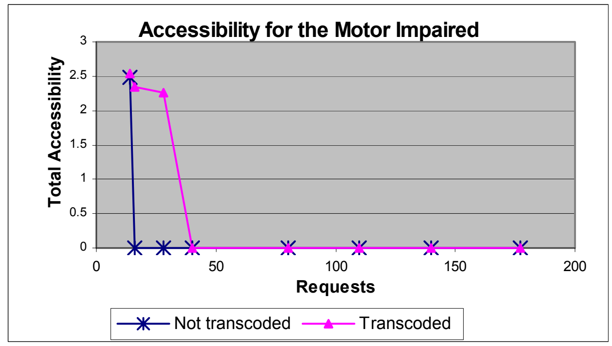

Fig 5: Effect of Load on Accessibility for the motor impaired

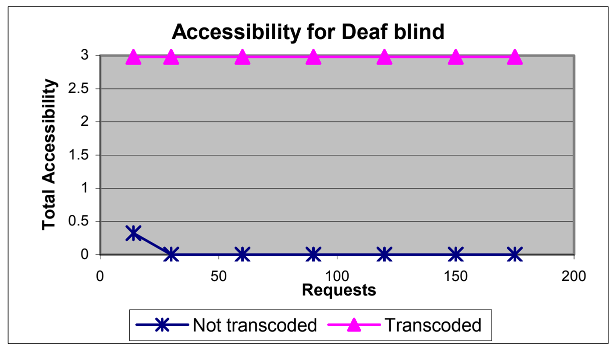

Fig 6: Effect of Load on Accessibility for the deaf-blind

4.2 Results of Specialized Transcoding for Vision Impairments

In this section we study the effect of the specialized transcoding for vision impairments. Specifically, we illustrate the effect of transcoding on refractive errors (myopia, far vision, presbiopia) and visual field errors (e.g., glaucoma). We compare the visual accessibility values for varying degrees of impairments under 3 scenarios – no transcoding, sufficient transcoding and limited specialized transcoding. The sufficient level of transcoding for a specific user and a specific media-type is calculated as the accessible transcoding level for normal viewing with the best resource utilization. The specialized transcoding process is limited by the number of quality levels available in the content source – the transcoding technique picks the nearest feasible approximation to the sufficient quality desired. A visual accessibility value higher than 1, implies that the delivered quality is better than required for useful assimilation of the information. The increased quality may improve the visual accessibility (as calculated); however the price of the enhanced quality is additional resource usage.

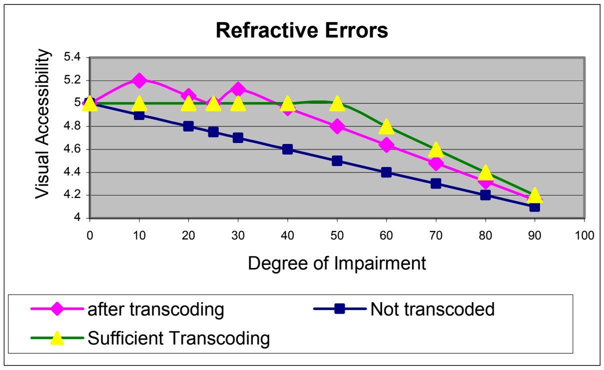

Figure 7: Effect of transcoding for refractive errors

From Figure 7, we observe that as the degree of impairment due to refractive errors increases, the visual accessibility decreases dramatically in the untranscoded case. Note that the visual accessibility, especially for a medium level impairment, improves significantly as a result of transcoding. For example, we assume the maximum magnification attainable is limited by the screen size. Hence even in the sufficient case no more magnification is possible beyond the screen height. With higher resolutions, the visual accessibility may increase beyond the sufficient value (with higher resource consumption); however, the improvement in the user's visual experience is not clear.

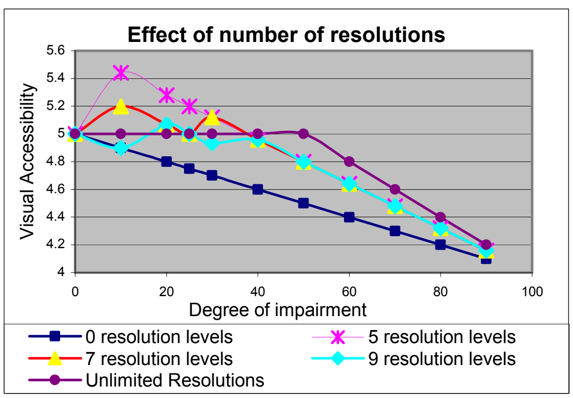

Figure 8: Effect of resolutions on visual accessibility

The second graph (Figure 8) shows the effect of increasing the available number of resolution levels in the data set. W see that as we increase the number of resolution levels the curve smoothens and gets closer to the sufficient case. When the number of levels is unlimited we can pick the required format for normal viewing without increasing resource consumption. However it is not realistic to have that many formats stored. Our experiments indicate that about 10 resolution levels are sufficient to obtain near ideal performance which gives the format that sufficiently accessible without compromising on resource usage.

The graph in Figure 9 shows the effect of visual field problems on accessibility. As the horizontal field of view decreases (as is the case in glaucoma patients), the area of the image seen is reduced. As a result, the size has to be reduced to fit the field of view. The effect of transcoding is similar to the refractive case. Here, we specify a minimum size, beyond which the image cannot be shrunk, keeping in view the effect of visual acuity.[1]

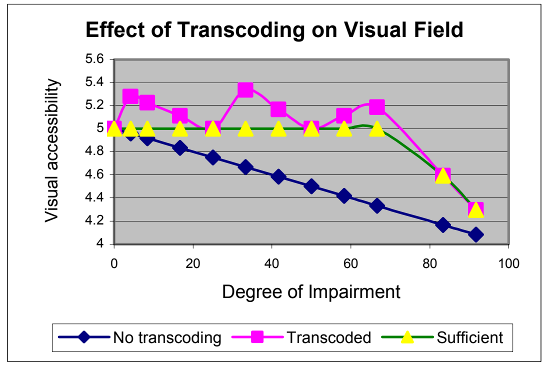

Figure 9: Graph showing the effect of transcoding on impairments affecting visual field.

5.CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper we discussed the ability based transcoding technique to improve the accessibility of the information to a disabled user. Our experiments illustrate that the proposed transcoding techniques significantly improve accessibility to multimedia information. In addition, the results show that there is an improvement in resource utilization as a result of the transcoding. We also demonstrate how specialized transcoding can help improve upon the accessibility as provided by the basic transcoding. Several challenges remain to be addressed in providing universal access for multimedia information. We are currently developing adaptive transcoding techniques to allow for dynamic reallocation of resources to ongoing streams to admit more users by allowing for simultaneous transcoding and load balancing. Other research topics currently being explored include ability-based transcoding in the presence of in the network proxy servers and the dynamic creation of transcoded content at the server and proxies. A future challenge is to address the issue of cross disability access in highly dynamic mobile environments.

REFERENCES

[AO02] American Optometric Association, http://www.aoanet.org, 2002

[1] Very small images tend to impact the visual acuity of the user, as the visual acuity is dependent on the size.

[AO92] D. Anderson, Y. Osawa and R.Govindan, "A file system for continuous media", *ACM Transactions on Computer systems,* **10(4)**: 311-337, Nov 1992.
[BP95] Bisdikian, Patel,"Issues on movie allocation in distributed video-on-demand systems", *IEEE Conference on Communication*, 1995
[BS00] A.Brinkman, K.Salzwedel, C.Scheideler, "Efficient, Distributed Data Placement Strategies for Storage Area Networks", *ACM symposium on Parallel algorithms and architectures*, 2000
[CD98] E.Cole, P.Dehdashti, "Computer-based cognitive prosthetics: assistive technology for the treatment of cognitive disabilities"*ASSETS* ,1998
[CP02] Canadian Psychological Association, http://www.cpa.ca/factsheets/cognitive.htm ,2002
[CS02] ContrastSensitivity.net, http://www.contrastsensitivity.net, 2002
[DS95] A. Dan and D. Sitaram , "An online video placement policy based on bandwidth to space ratio (bsr)", *SIGMOD*, pp376-385, 1995
[FQ84] E. A. Fleishman and M.K. Quaintance, *Taxonomies of human Performance: The description of Human Tasks*, 1984
[GK98] L. Gao, J. Kurose, and D. Towsley , "Efficient Schemes for broadcasting popular videos", *NOSSDAV,*1998
[HH99] Harborg, Trym Holter, et al, "Subtitling of live broadcast TV-programs for the hearing impaired", *AAATE* ,1999
[HS00] A.W.Huang and N.Sundaresan, "A semantic transcoding System to Adapt Web Services for Users with Disabilities", ASSETS,2000
[HS01] Richard Han, John R. Smith, M*ultimedia Communications: Directions and Innovations*, 2001
[JV01] J.A.Jacko,H.S.Vitense , "A review and reappraisal of information technologies within a conceptual framework for individuals with disabilities", *UAIS* ,2001
[K98] S. Keates, P.J. Clarkson and P. Robinson, "Developing a methodology for the design of accessible interfaces", *ERCIM*,1998
[KR98] Simeon Keates and Peter Robinson , "The use of gestures in multimodal input", *ASSETS* ,1998
[LL99] C.Lee, J.Lehoczky, D.Siewiorik, et al, "A Scalable Solution to the Multi-Resource QoS Problem", *IEEE Real-Time Systems Symposium*, 1999
[MS99] R.Mohan, J.Smith and C.Li, "Adapting Multimedia Internet Content for universal Access", *IEEE Transactions on Multimedia*, **vol.1**, pp. 104-114, 1999
[MW97] G.K.Ma, C.S Wu, M.C Liu, P.Lin, "Efficient Real-time Data Retrieval Through Scalable Multimedia Storage", *ACM Multimedia* ,1997
[RI02] Retina International, http://www.retina-international.org/, 2002
[RL01] Robin Leonard, "Statistics on Vision Impairment", A *Resource Manual*, 2001
[RL98] R.Rajkumar, C.Lee, J.Lehoczky, D.Siewiorek, " On Quality of Service Optimization with Discrete QoS options", *IEEE Real-Time Systems Symposium*, 1997
[RM96] C.Ramstein, O.Martial, A.Dufresne et al "Touching and Hearing GUI's: Design Issues for the PC-Access System", *ASSETS* ,1996
[SAP] http://www.sapdesignguild.org/resources
[SCI] http://www.spinalcord.org
[SEV00] S.Chandra, C.S.Ellis and A.Vahdat, "Differentiated Multimedia Web Services Using Quality Aware Transcoding" *INFOCOM* ,2000
[SGV95] P.J.Shenoy, Pawan Goyal and Harrick M. Vin, "Issues in Multimedia Server Design", *ACM Computing Surveys,* **27(4)** , pp. 636-639, 1995
[SS01] C.Stephanidis and A.Savidis, "Universal Access in the Information society: Methods, tools, and Interaction Technologies", *UAIS* 2001
[SS99] S.Sahu, P.Shenoy, D.Towsley, "Design considerations for integrated proxy servers", *NOSSDAV* ,1999
[SUGA02] The SUGA Project, http://www.ics.uci.edu/~dsm/suga/, 2002
[SW99] Stefan Winkler "Issues in Vision Modeling for Perceptual Video Quality Assessment" *Signal Processing*, **78(2)**, pp. 231-252, October 1999
[TA00] H.Tagaki and Chieko Asakawa, "Transcoding Proxy for Nonvisual Web Access", *ASSETS* ,2000
[TP93] F. Tobagi, J. Pang, R. Baird, and M. Gang, "Streaming RAID: A disk storage system for video and audio files", *ACM Multimedia* ,1993
[VC02-1] Vision Channel, http://www.visionchannel.net, 2002
[VG94] H.M.Vin, A.Goyal, P.Goyal, "Algorithms for Designing Multimedia Servers", *ICMCS*, (234-243) ,1994
[VW01] A.Vetro, Y.Wang, and H.Sun, "Object-Based Transcoding for adaptable Video Content Delivery", *IEEE, transactions on Circuits and Systems for Video Content Delivery*, **11(3)**, March 2001
[WH01] J.T. Wilensky, A.Hawkins, "Comparison of Contrast Sensitivity, Visual Acuity and Humphrey Visual Field Testing in Patients with Glaucoma", *Tr. Am. Ophth. Soc*. **99**, 2001
[WHO1] http://www.who.int/whosis
[ZM99] Y.Zhuang, S. Mehrotra, and Thomas S. Huang, "A Multimedia Information Retrieval Model Based on Semantic and Visual Content",*Proceedings of the 5th International ICYCS Conference*, 1999.
[ZV01] Gottfried Zimmermann and Gregg Vanderheiden, "Modality Translation Services on demand Making the World More Accessible For All", *RESNA* ,2001