CANADIAN NETWORK *for* INCLUSIVE CULTURAL EXCHANGE

General Guidelines for Inclusive Online Cultural Content

ATRC Adaptive Technology Resource Centre
UNIVERSITY OF TORONTO

Patrimoine canadien Canadian Heritage

Acknowledgements

Canadian Network for Inclusive Cultural Exchange (CNICE)

Creating Accessible Online Cultural Content Discussion Document Series

CNICE Partners

Adaptive Technology Resource Centre
Banff New Media Institute
Canadian Abilities Foundation
Canadian Centre on Disability Studies
Canadian Cultural Society of the Deaf
Canadian Hearing Society
Canadian Learning Television
Centre for Learning Technology
Centre for Research in Disability Studies
Digital Frog International
marblemedia
National Broadcast Reading Service

Editors and Contributors:

Michael Boyce, Sara Diamond, Deborah Fels, Elle Gadsby, Rob Harvie, Wendy Porch, Lake Porter, Jan Richards, Vera Roberts, Marc Rosen, Nancy Sicchia, Charles Silverman, Jutta Treviranus

This project was made possible through the funding by the New Media Research Networks Fund of the Canadian Heritage Canadian Culture Online Programme.

Canadian Heritage Patrimoine canadien

Executive Summary:

These discussion documents have been prepared as part of the Canadian Network for Inclusive Cultural Exchange (CNICE) project.

The package includes a single "core" document followed by four "companion documents," each devoted to closer examination of particular areas:

1. Online Enhanced Captioning
2. Online Video Description
3. Remote Real-time ASL Interpretation
4. Representations of Visual Geo-spatial Information.

The core document, *General Guidelines for Inclusive New Media Cultural Content*, is organized into six main sections:

Section 1.1 is the introduction to the core document and outlines its. This section also includes definition of terms used in the document as well as a brief discussion of disability culture.

Section 1.2 is an overview of existing accessibility principles, of which, developers of cultural content should be aware.

Section 1.3 lays out a number of challenges faced by developers of accessible online cultural content. These include a discussion about the right to artistic expression and accessibility, as well discussions of the aesthetics, entertainment value, perspective and interpretation of cultural works as related to accessibility.

Section 1.4 introduces the concept of modality translations. Aspects of the translation of art are discussed such as emotion, perspectives, presentation, cultural differences and workflow.

Sections 1.5 and 1.6details the possible modality transformations and begin the discussion of how translation of cultural content from one modality to another might be carried out.

Section 1.7 is the conclusion of the core document.

The four companion documents follow in the next four chapters. The first companion document, chapter 2, looks at online captioning and details captioning in flash as well as introduces the *CapScribe* tool for caption and video description that was developed as part of the CNICE project.

Chapter 3 details Online video description. This discussion extends a discussion that is started in chapter 2 and provides guidance and principles for this modality translation.

Chapter 4 studies in great detail the challenge of creating remote real-time ASL interpretations for cultural events. Along with comprehensive guidelines, the chapter provides first-hand viewpoints and an overview of relevant technologies.

The final companion document, chapter 5, discusses the special case of representing visual geospatial information in alternative modalities. The authors discuss solutions that use SVG as well as VRML. The authors close with guidelines for creating inclusive navigation and way-finding content for the web.

1 General Guidelines for Inclusive New Media Cultural Content 9
1.1 INTRODUCTION 9
1.1.1 Scope 10
1.1.2 Terms Defined 10
1.1.3 Discussion of Disability Culture 12
1.2 EXISTING PRINCIPLES OF ONLINE ACCESSIBILITY 12
1.2.1.1 General Web Content Accessibility Guidelines 13
1.2.1.2 Authoring Tools and User Agents Guidelines 14
1.2.1.3 Language or Format Specific Accessibility Guidelines 15
1.2.1.4 General Software Accessibility Guidelines 16
1.2.1.5 Operating System Specific Accessibility Guidelines 17
1.2.1.6 Education Focussed Accessibility Guidelines 17
1.2.2 XML and Interoperable Information 19
1.2.3 Accessibility Focussed Metadata and Information Architecture 19
1.2.4 Inclusive Usability Evaluation Methods 20
1.3 KEY BENEFITS OF ACCESSIBLE DESIGN FOR ONLINE CULTURAL CONTENT 21
1.3.1 Accessible Equivalents for Deliberately Challenging Interfaces 23
1.3.2 The Role of Aesthetics 24
1.3.3 Entertainment and Engagement Values 25
1.3.4 Perspective: A Cultural and Technical Consideration 26
1.3.5 Interpretation 27
1.4 MOVING ARTWORK ACROSS MODALITIES 28
1.4.1 Translating Emotion Across Modalities 28
1.4.2 Provide Multiple Perspectives 29
1.4.3 Consider the Presentation Context 31
1.4.4 Considering Cultural Differences 31
1.4.5 Integrate with Workflow 32
1.5 INTRODUCTION TO ONLINE CONTENT MODALITIES 32
1.5.1 Visual Modality 33
1.5.2 Audio Modality 33
1.5.3 Haptic Modality 34
1.5.3.1 Haptic Devices 35
1.5.4 Language Modality 36
1.6 MODALITY TRANSLATIONS 36
1.6.1 Twelve Possible Modality Translations 37
1.6.2 Alternative Modality Equivalents of Visual Content 39
1.6.2.1 Visuals to Language 39
General Techniques for using Language to Describe Visuals 40
1.6.2.2 Visuals to Audio 41
Video Descriptions 42
Music 43
Sound Effects 44
Automated auditory displays 48
1.6.2.3 Visuals to Haptics 49
1.6.3 Alternative Modality Equivalents of Audio Content 51

1.6.3.1 Audio to Language 51
Transcription of dialogue (Captions) 51
Description of Music 52
Description of Sound Effects 55
1.6.3.2 Audio to Visuals 56
ASL Translation of Speech 56
Graphical Component of Enhanced Captions 56
Visual Displays of Music or Sound Effects 56
1.6.3.3 Audio to Haptics 57
1.6.4 Accessible Collaborative Tools 57
1.6.4.1 A-Chat: Accessible Online Chat Tool 57
Key Features 57
Screenshots: 58
Suggested Use of A-Chat: 59
1.6.4.2 A-Comm: Client-Side Chat/Whiteboard Tool 60
Key Features: 60
Screenshot: 60
Suggested Description Strategies: 61
Using Pre-Authored Drawing Objects: 62
1.6.5 Ensuring Accessible Control 62
1.6.5.1 Keyboard Accessibility 62
1.6.5.2 Mouse Adaptation 63
1.6.5.3 Voice Recognition 63
1.6.5.4 Gestural Interfaces 64
1.7 CONCLUSION 65
2 Online Enhanced Captioning 67
2.1 THE CURRENT STATE OF CAPTIONING AND DESCRIPTIVE VIDEO ON THE WEB 67
2.2 THERE'S NO ONE STANDARD FOR CAPTIONING AND DESCRIPTION ON THE WEB 68
2.3 STYLE GUIDES FOR CAPTIONING AND DESCRIBING ON THE WEB 68
2.4 TOOLS FOR CAPTIONING AND DESCRIBING ON THE WEB 68
2.5 POSSIBILITIES FOR CAPTIONS AND DESCRIBING ON THE WEB 69
2.5.1 Captioning and the Web 69
2.5.2 Description and the Web 73
2.6 TAKING TIME TO ADD ACCESS TO THE WEB 73
2.7 THE CAPSCRIBE TOOL 74
2.8 ONLINE CAPTIONING OF FLASH 75
2.8.1 The Flash Captioning Tool 75
2.8.2 Using the marblemedia Flash Captioning tool 76
2.8.3 Techniques, Usability Issues and Further Considerations 76
2.8.4 Summary 77
3 Online Video Description 79
3.1 INTRODUCTION TO VIDEO DESCRIPTION 79
3.1.1 Real-Time Online Video Description 80
3.1.2 System Description 82
3.1.3 Procedure for describing a live stream 84

3.1.4 Preferences 85
3.1.5 Technical Difficulties 87
3.2 DISCUSSION AND RECOMMENDATIONS 88
4 Remote Real-Time ASL Interpretation 90
4.1 INTRODUCTION 90
4.2 TECHNOLOGY OVERVIEW 91
4.2.1 Connecting IP technologies with ISDN 94
4.2.2 Bandwidth recommendations 94
4.2.3 Applications 94
4.2.4 Physical space/Room technologies 95
4.2.5 Environmental Considerations 95
4.2.6 Future Considerations 95
4.3 TECHNOLOGY ISSUES RELATED TO ACCESSIBILITY 95
4.3.1 Video conferencing and use with sign language interpreters for people who are deaf 96
4.3.1.1 Sign Language Interpretation 96
4.3.1.2 Video remote interpreting 98
4.3.1.3 Challenges of Video Remote Interpreting 98
4.3.1.4 Considerations for video remote interpreting 99
Eye contact/gaze 99
Seating 101
Environmental and technical issues 102
Detailed Examples of Visual Noise 103
Other physical factors to consider 103
Turn-taking 105
Confidentiality 106
Special Considerations for managing multipoint or multi-application conferencing 106
Other factors 107
4.3.1.5 Considerations for Remote Sign Language Interpreters 108
Preparation 109
During the video conference 109
Processing in public 109
Lag Time 110
Interpreters, deaf people and hearing people at several sites 110
Interrupting the speaker (either deaf or hearing) 110
Reduced Signing Space 111
Auditory Referencing 111
Team Interpreting 111
Deaf Interpreter 111
4.3.1.6 Skills Needed for Remote Interpreters 112
Other factors to consider 113
4.4 ILLUSTRATIVE CASES AND PERSONAL ACCOUNTS HIGHLIGHTING ISSUES 113
4.4.1 Personal Account: My First Impressions of Videoconferencing 114
4.4.2 Personal Account: My First Impressions of Videoconferencing 114
4.4.3 Personal Account: Web Applications Specialist 115

4.5 USER ENVIRONMENT SCENARIOS: ... 116
4.5.1 Health care Environments ... 116
4.5.1.1 Tasks 116
4.5.1.2 Physical scenarios 117
4.5.1.3 Issues unique to health care settings 117
Physical Positioning ... 117
Sign language production ... 117
Socio/Political ... 117
4.5.1.4 Guidelines and Protocols specific to Health Scenarios 118
4.5.2 Education Environments ... 119
4.5.2.1 Typical tasks 119
4.5.2.2 Physical scenarios 120
4.5.2.3 Issues unique to education settings 121
Turn-taking ... 121
Shared applications ... 121
Other important factors ... 121
4.5.2.4 Guidelines and Recommendations specific to education 121
4.5.3 Business Environments ... 122
4.5.3.1 Tasks 122
4.5.3.2 Scenarios 122
One to One Meetings ... 122
Small Group Meeting ... 123
Large meetings ... 123
4.5.3.3 Issues unique to meeting settings 123
Turn-taking ... 123
Visual materials ... 123
Socio/political ... 124
Technology alternatives ... 124
4.5.3.4 Guidelines and Recommendations specific to meetings 124
4.6 LIVE CAPTIONING: AN OVERVIEW ... 124
4.7 RECOMMENDATIONS AND GUIDELINES – SUMMARY ... 125
4.8 GLOSSARY OF TERMS ... 129
5 Representations of Visual Geo-Spatial Information ... 130
5.1 INTRODUCTION ... 130
5.2 2D CONTENT ... 131
5.2.1 Translations: Visual to Sound ... 131
5.2.1.1 Spatialized Audio 134
5.2.1.2 The Aural Legend 136
5.2.1.3 Information Scalability 136
5.2.1.4 Translation Algorithms 137
5.2.2 Translations: Visual to Touch ... 141
5.2.2.1 Full Force and Tactile Feedback Distinctions 141
5.2.2.2 Haptic Effects Provided by Immersion Corporation 141
5.2.2.3 Tactile Effects Applied to Geospatial Content 142
5.2.3 Format: Scalable Vector Graphics (SVG) ... 144
5.2.3.1 Resources 145

5.2.3.2 Methodology 146
5.2.3.3 Code Sample 1: An SVG that draws a fill red circle. 146
5.2.3.4 Code Sample 2: Using Sample 1 code, group objects together, add sound and assign JavaScript functions as needed (shown in bold). 147
5.2.3.5 Code Sample 3: The minimum HTML code required to implement haptics with SVG. 147
5.2.3.6 GIS Applications and SVG for Government On-Line (GOL) Content 148
5.3 3D CONTENT 149
5.3.1 Transformations: Visual to Sound 149
5.3.1.1 Visual to Touch 150
5.3.1.2 Visual to multimodal 151
5.3.2 Format : Web3D - Virtual Reality Modeling Language (VRML)/X3D .. 152
5.3.2.1 Methodologies 152
5.3.2.2 Code Sample 1: Initial HTML with JavaScript to enable immersion effects 153
5.3.2.3 Code Sample 2: VRML generating a red sphere that will detect collisions 154
5.3.2.4 Code Sample 3: Java applet used with VRML from Step 2 to detect collisions. 154
5.3.2.5 Code Sample 4: Embedding VRML and applet into the HTML from Step 1 155
5.3.2.6 VRML/X3D and GIS 157
GeoVRML 157
Resources 157
5.3.3 Navigation and Way-finding Guidelines 158
Learning about an Environment 158
Landmark Types and Functions 158
Landmark Composition 159
Landmarks in Natural Environments 160
Combining Paths and Landmarks: Landmark Placement 160
Using a Grid 161
5.4 MODALITY COMPLEMENTARITY 161
5.5 APPENDIX: CODE SAMPLE FOR ADDING BOTH IMMERSION WEB PLUG-IN AND USER ENVIRONMENT DETECTION TO HTML. 163
References 167

Perspectives:

BOX 1: ACCESSIBILITY FROM AN ARTIST'S PERSPECTIVE *BY SARA DIAMOND* 29
BOX 2: AN ASL CONTENT DEVELOPER'S PERSPECTIVE *BY ELLE GADSBY* 78

Figures:

FIGURE 1: POTENTIAL MODALITY TRANSLATIONS. 37

Figure 2: Language bridges all three computer modalities and is a flexible medium for modality translation. 38
Figure 3: This screenshot shows one of the A-Chat user preference screens. 58
Figure 4:This screenshot shows the main A-Chat screen. The message area in on the top-left, the compose message area under it and the ootions and user list and history areas on the right. 59
Figure 5: An example of a teacher-led lesson using drawings to demonstrate a concept, three students ("Amanda", "David" and "Cathy") submit peer descriptions. The most recently drawn shape, an ellipse has not yet been described. 61
Figure 6: Example of captioned streamed video. 67
Figure 7: Sample of captioned PBS video. 67
Figure 8: Example of annotations with captions. 67
Figure 9: Example of traditional appearance, mixed case captioning. 69
Figure 10: Example of captions with enhanced font styles. 70
Figure 11: Example of enhanced caption font style used to convey tone and energy. 70
Figure 12: Example of multilingual captions. 71
Figure 13: Example of graphics to convey language and sounds. 71
Figure 14: Example of graphics to convey silences and pauses. 72
Figure 15: Example of animation to convey sound elements of video 72
Figure 16: CapScribe Tool 74
Figure 17: Silence indicators that move horizontally as the video progresses 82
Figure 18: Describer control and status indicators. 83
Figure 19: Describer interface. 85
Figure 20: Preferences screen. 86
Figure 21: Different furniture and placement configurations. 105

Tables:

Table 1: Usability evaluation methods Web resources. 21
Table 2: Suggested seating arrangements for all participants. 101
Table 3: Summary of technology considerations for video conferences involved hearing and deaf participants. 103
Table 4: Behavioural, communication and etiquette issues. 108
Table 5: Real world or aural metaphor sounds for an intersection map. 138
Table 6: Abstract sounds for an intersection map. 138
Table 7: Geopolitical map translation matrix 141
Table 8: Translation algorithm for representing typical map elements with haptic effects. 143
Table 9: Five landmark types. 159
Table 10: Sample landmark items. 160

1 General Guidelines for Inclusive New Media Cultural Content

1.1 Introduction

The purpose of this document is to describe techniques by which online cultural material may be made more accessible to people with sensory, motor and learning disabilities.

The accessibility of online content to people with disabilities is an active area of research that has produced a variety of guidelines and best practice documents. However, at the time of writing, these documents have tended to be either general-purpose guidelines (e.g. the W3C-WAI Web Content Accessibility Guidelines—WCAG) or guidelines that have a relatively narrow focus on content in the educational domain (e.g. the NCAM “Math and Science Education Guidelines” and the IMS Accessibility Guidelines). In both cases, the guidelines do little to address issues, such as aesthetic experience, that are of crucial importance in the cultural domain. The intent of this discussion document, then, is not to replace the existing guidelines documents, but rather to build on their strengths to provide more comprehensive coverage of content within the cultural realm. Naturally, as this document is an early attempt to investigate this large and complex area, we fully expect this document to raise as many questions as it answers.

One of the main assumptions of this document is that all of the users of the content are able, in one way or another, to access online content. In reality, access to online content (via computers and any required assistive technology) is often subject to social, cultural and economic constraints. These issues lie outside of the current scope. However, considerations of bandwidth will be addressed.

Another concern is that cultural content is quite often set within the context of cultural practices and norms. Efforts have been made to be sensitive to these aspects within this document. This effort is a balancing act: on the one hand, we must respect certain cultural proprieties, and on the other, explore definitions of what constitutes a good, inclusive, communicative and attractive design.

The core of this document concerns “modality translation.” This term refers to the representation of some online entity by an entirely different modality for example, an image described by text. By its nature, such transformation will always be unable to convey fully the aesthetic qualities conveyed by the original. Therefore, the interest is in defining the specific contextual parameters of such a transformation while registering, in the face of its limitations, both its value as an interpretive act and its commitment to being suitably comprehensive and duly unabridged. In the end, the techniques herein are designed to expand and enhance content, for those who wish it. There is absolutely no intention to "dilute" the richness of the experience for any users.

1.1.1 Scope

This document is intended to be applicable to a wide, and ever-expanding, variety of online cultural content. This scope includes content that is exclusively online content (e.g. a computer graphic created with freehand drawing tools) as well as online content that depicts physical material (e.g. a digital picture of a "real" painting in a "real" gallery).

Examples of online cultural content can include the following:

> **Visual Art:** Computer art, interactive multimedia, digital images of paintings, sculpture, installation art, etc., scanned photographs, as well as online tools for creating art.
>
> **Architecture:** Images of buildings, architectural fly-throughs as well as online architectural design tools.
>
> **Dance:** Images or video of dancers, as well as online interactive avatar choreography tools.
>
> **Music:** Audio or video of music performance, singing, opera, as well as music making tools.
>
> **Museum-type exhibits:** Images or video of museum artefacts as well as artefact organization or exploration tools.
>
> **Virtual Galleries:** Art displays including visual, audio, multi-media, games, etc. also interactive/participatory elements.

In recent years, cultural production has increasingly expanded into electronic forms. The methods and strategies discussed throughout this document will ensure that people with disabilities also have access to this growing body of cultural content.

1.1.2 Terms Defined

The underlying theme of this document is that a goal of accessible online culture content is possible, inspires creativity and provides a broader conception of the audience. In order to convey these ideas, some novel terms may be used to explain the creation of accessible content. These terms are defined here:

Alternatives-refers to cultural content that may be accessed through another perceptual modality. For example, a spoken description of the visual events in a video is an auditory alternative for visual content.

Accessible equivalent-refers to an alternative form of the work that does enables the inclusion of an individual who would otherwise be excluded from the original due to a sensory impairment.

Audience-refers to the people who observe or participate in the experience of the cultural object/work of art. Participation generally depends on the medium of the art and, for example, may refer to viewing, listening or reading.

Cultural Content/Object-refers to an expression that is generally artistic in nature or to the more broadly conceived of expression of group identity that is formed in part from artistic expression and in part from shared experience.

Equivalents-refers to content that has the same (or as close to the same as possible) sense of the aesthetic, meaning and intention of content in another modality. This term is an acknowledgement of the esoteric nature of cultural objects that are created in a specific medium and is intended to encourage inclusion of these aspects in any translation to another modality.

Haptic Device or Interface-'Haptic' comes from the Greek word haptesthai, meaning 'to touch' (Wall, 2000) and entails both controlling human movement and getting feedback through our sense of touch. A haptic interface transmits forces to a person's hand or fingers in a way that mimics the sensation of touching real objects. Virtual haptic touch can be particularly useful for people with visual impairments. It makes it possible for a blind person to touch virtual objects, corresponding to the way a sighted person can see objects on a computer screen. (Sjöström, 2002)

Interpretation-refers to the culturally sensitive rendering of a message from one medium of dialogueue to another. Mediums of dialogueue may include language as well as artistic forms such as visual art.

Map Symbols-The graphic elements shown on a map designed to represent geographic features or communicate a message, at a given scale. There are three basic forms of map symbology: point, line, and area. Map symbols may represent qualitative data and be shown as a distribution, or they can express quantitative data by showing proportionality or ratio. Map symbols are usually explained in a legend.

Modality-refers to the sensory realm of the content. For example music is in the auditory mode while a video of a musical performance will have both auditory and visual modes. Modalities are further defined in section 1.5, .

Modality Complimentarity-Sensory perception of information about the world around us is normally a cross modal experience, involving a combination of more than one engaged sensory channel (sight, hearing, touch, taste, smell) at a time. Congruency of these multiple signals lead to a more optimal or synergistic comprehension, including easier, faster recognition, with less confusion.

Transformation-has the same meaning as translation and may be used interchangeably with it.

Translation-refers to the process of interpreting content typically from one modality to another. For example, captions are text translations of the spoken part of a video. The creation of captions is a translation from the auditory mode to the visual mode.

1.1.3 Discussion of Disability Culture

This document is concerned with artistic representations of culture, it is, therefore, important to also acknowledge the existence of disability culture. For many people with disabilities, there is a sense of shared experience that exists as a result of having been marginalized and excluded from mainstream culture. This sense of shared identity and experience, along with the human rights activism undertaken in the mid 1980's, has contributed to many people with disabilities identifying with and partaking in the creation of disability culture. As a result, for some artists with disabilities, their identification with disability culture remains central to an authentic representation of their work. For these artists, it may be important when describing their work to acknowledge that it was produced by a person with a disability. Other artists with disabilities, however, may feel that the issue of their own disability has little relevance to their work or their identity as an artist. Should you be responsible for integrating the kinds of accessibility practices recommended in this document and be unsure as to whether or not an artist with a disability wants to be identified as such, please ask him or her their preference on the issue. For a more extensive view of disability culture please see Brown (2002) and Peters (2000).

1.2 Existing Principles of Online Accessibility

The task of creating accessible online cultural content is in many basic respects, the same as that of creating any online content. In both cases, the author is assembling larger pieces of content from an array of smaller constituent parts, including text fields, multimedia, markup and program code. The special nature of cultural online content is evident, not in the types of constituents employed or how they are arranged at a the structural or architectural level, but rather by the information contained within the constituents and the way this information combines in the content as a whole.

At the structural level of online content constituent parts, therefore, the design principles that must be followed are largely the same regardless of the ultimate purpose of the content. For example, support for keyboard navigation is a requirement that is implemented at the level of the constituent parts (links, form controls, etc.), regardless of whether the final content is cultural in nature. Since a body of widely accepted standards, guidelines and techniques describing these structural level design principles already exist, no purpose is served by "re-inventing the wheel" in this document. In fact, "guideline fragmentation," the practice of creating numerous competing guidelines, is a potentially serious problem for efforts to establish a unified definition of accessibility. Therefore, we recommend that authors of cultural online content develop content that meet these the relevant existing principles of online accessibility described in section 1.2.1.1 General Web Content

Accessibility Guidelines. Furthermore, for a discussion of those aspects of accessible design that are unique to cultural online content, see section 1.3 Key Benefits of Accessible Design for Online Cultural Content, and for a discussion of modality translations see 1.4 Moving Artwork Across Modalities and 1.5 Introduction to Online Content Modalities.

The following sections list guidelines documents that may be relevant to materials displayed or distributed online. These guidelines should be followed closely, except where they conflict with the solutions discussed in 1.6 Modality Translations.

1.2.1.1 General Web Content Accessibility Guidelines

The Web Content Accessibility Guidelines (WCAG) documents produced by the Web Access Initiative of the World Wide Web consortium (W3C-WAI) are the most widely accepted authority on what constitutes accessible Web-based materials. Other guidelines do exist, for example the Section 508 regulations in the United States. However, because these other guidelines tend to be limited subsets of WCAG, web content that satisfies WCAG will also satisfy most other guidelines. Therefore, we recommend following WCAG as a general rule.

To aid in the implementation of WCAG, links are included to freely available multiple language techniques documents. We recommend following these techniques documents, where possible.

1. Web Content Accessibility Guidelines (WCAG), v. 1.0 (Recommendation)
Produced by the World Wide Web Consortium (W3C)
(Chisolm, Vanderheiden, & Jacobs, 1999)

WCAG 1.0 (http://www.w3.org/TR/WCAG10/) was published in 1999 in an effort to unify a variety of informal guideline attempts that existed at the time. The result is the following fourteen guidelines:

1. Provide equivalent alternatives to auditory and visual content
2. Do not rely on colour alone
3. Use markup and style sheets and do so properly
4. Clarify natural language usage
5. Create tables that transform gracefully
6. Ensure that pages featuring new technologies transform gracefully
7. Ensure user control of time-sensitive content changes
8. Ensure direct accessibility of embedded user interfaces
9. Design for device-independence
10. Use interim solutions
11. Use W3C technologies and guidelines
12. Provide context and orientation information
13. Provide clear navigation mechanisms
14. Ensure that documents are clear and simple

Techniques: http://www.w3.org/TR/2000/NOTE-WCAG10-TECHS-20000920/

2. Web Content Accessibility Guidelines (WCAG), v. 2.0 (Working Draft)
Produced by the World Wide Web Consortium (W3C)
(Caldwell, Chisolm, Vanderheiden, & White, 2003)

As of April 1, 2004, WCAG 2.0 (http://www.w3.org/TR/WCAG20/) is still a working draft, and therefore may change in the future. At the moment WCAG 2.0 has a more simple structure than the first version, with an organization that emphasises four guiding principles, each encompassing several lower-level guidelines. The higher level principles are:

1. Ensure content can be perceivable by any user.
2. Ensure content can be operable by any user.
3. Ensure content can be understandable to as many users as possible.
4. Use Web technologies that maximize the ability of the content to work with current and future accessibility technologies and user agents.

Techniques: http://www.w3.org/WAI/GL/wcag20.html#techs

1.2.1.2 Authoring Tools and User Agents Guidelines

The Authoring Tool Accessibility Guidelines (ATAG) and the User Agent Accessibility Guidelines (UAAG) are companion documents to the WCAG introduced above. These two guidelines documents apply to software programs that perform authoring and content rendering, respectively.

To aid in the implementation of these documents, links are included to freely available techniques documents in several languages. We recommend following these techniques documents, where possible.

1. Authoring Tool Accessibility Guidelines (ATAG), v. 1.0 (Recommendation)
Produced by the World Wide Web Consortium (W3C)
(Treviranus, McCathieNevile, Jacobs, & Richards, 2000)

This document provides guidelines to developers of authoring tools. An "authoring tool" can be any software that is used to create Web content of any kind. Examples include HTML editors, programming tools, multimedia editors and more advanced content management-type systems. The guidelines cover the steps required to make the output of the tool and the tool interface accessible.

Techniques: http://www.w3.org/TR/ATAG10-TECHS/

3. User Agent Accessibility Guidelines (UAAG), v. 1.0 (Recommendation)
Produced by the World Wide Web Consortium (W3C)
(Jacobs, Gunderson, & Hansen, 2002)

This document provides guidelines to developers of user agents. The term "user agent" denotes any software that retrieves and renders Web content. Web browsers are user agents, as are media players, plug-ins, and some assistive technologies. The guidelines also extend to less conventional systems, such as systems for navigating the Web and reading email over the phone. The guidelines cover support for device independence, configurability and navigation.

Techniques: http://www.w3.org/TR/2002/NOTE-UAAG10-TECHS-20021217/

1.2.1.3 Language or Format Specific Accessibility Guidelines

The Web content guidelines, referenced in section 1.2.1.1, General Web Content Accessibility Guidelines, apply generally to any Web content. However, the implementation details will depend on the specific language or format used to create the online content. Along with the WCAG techniques already listed, the following documents should be consulted.

1. IBM Java Accessibility Checklist
Produced by IBM
(Schwerdtfeger, 2000)
(http://www-306.ibm.com/able/guidelines/java/accessjava.html)

This checklist applies specifically to software programs developed in Java to run as applets or stand alone applications.

2. Accessibility Features of CSS
Produced by the World Wide Web Consortium (W3C)
(Jacobs & Brewer, 1999)
(http://www.w3.org/TR/CSS-access)

This W3C note includes a technical discussion of features in Cascading Style Sheets (CSS) that support accessibility.

4. WAI Resource: HTML 4.0 Accessibility Improvements
Produced by the World Wide Web Consortium (W3C)
(Jacobs, Brewer, & Dardailler, 2000)
(http://www.w3.org/WAI/References/HTML4-access)

This W3C note includes a technical discussion of features in the Hypertext Markup Language (HTML) 4.0 support accessibility.

4. Accessibility Features of SMIL
Produced by the World Wide Web Consortium (W3C)
(Jacobs & Koivunen, 1999)
(http://www.w3.org/TR/SMIL-access/)

This W3C note includes a technical discussion of features in the Synchronized Multimedia Integration Language (SMIL) that support accessibility of multimedia.

5. Accessibility Features of SVG
Produced by the World Wide Web Consortium (W3C)
(McCathieNevile & Koivunen, 2000)
(http://www.w3.org/TR/SVG-access/)

This W3C note includes a technical discussion of features in the Scalable Vector Graphics (SVG) language that support accessibility.

6. Creating Accessible Flash [MX]
Produced by WebAIM
(http://webaim.org/techniques/flash/)

This WebAIM techniques document is a good starting place for investigating Flash Accessibility. Macromedia's own Flash MX Accessibility portal is also useful (http://www.macromedia.com/macromedia/accessibility/features/flash/).

7. Specifications for the Digital Talking Book, ANSI/NISO Z39.86-2002
(Developed by National Information Standards Organization)
(National Information Standards Organization, 2002) (http://www.niso.org/standards/resources/Z39-86-2002.html)

This digital talking book format is based on XHTML1.0 and SMIL 1.0 and allows a sequential, hierarchical information structure to be synchronized with recorded audio to create accessible "talking" books. This work is based on the DAISY Digital Talking Book (DTB) Format (The DAISY Consortium, 2001), (http://www.daisy.org/publications/specifications/daisy_202.html).

1.2.1.4 General Software Accessibility Guidelines

Sometimes the Web is used primarily as a distribution medium for software that runs primarily outside a user's browser and therefore does not usually qualify as Web content. In these cases, it is useful to apply more general software accessibility guidelines that do not assume an online environment:

1. IBM Software Accessibility Guidelines, v. 3.1
Produced by IBM
(IBM, 2002)
(http://www-3.ibm.com/able/guidelines/software/accesssoftware.html)

These guidelines are primarily intended to help developers meet the requirements of Section 508 of the U.S. Rehabilitation Act. However, with conveniently linked techniques to aid implementation, this is a good general-purpose resource, applying to the design of software programs on all platforms.

2. **ISO-16071 (Ergonomics of human-system interaction--Guidance on accessibility for human-computer interfaces)**
Produced by the International Organization for Standardization (ISO) (http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=30858&ICS1=13&ICS2=180&ICS3=)

The ISO16071 is a general-purpose standard, similar to the IBM guidelines, but is *only available for a fee*.

1.2.1.5 Operating System Specific Accessibility Guidelines

Once software that is intended to be distributed online, but run outside the browser, has been reviewed with the general software accessibility guidelines, above, it is advisable to continue by applying guidelines specific to the particular operating systems that a user might be using:

1. **Microsoft Accessibility for Application Designers**
Produced by Microsoft
(http://msdn.microsoft.com/library/default.asp?url=/nhp/default.asp?contentid=28000544)

These resources apply specifically to software programs running the Microsoft Windows platforms.

2. **MacOS Accessibility Documentation**
Produced by Apple
(http://developer.apple.com/documentation/Accessibility/Accessibility.html)

These resources apply specifically to software programs running the MacOS platforms.

3. Disability Access to GNOME
Produced by the GNOME Accessibility Project
(http://developer.gnome.org/projects/gap/)

These documents include development and testing guidance for developers of the GNOME desktop for Linux and Unix.

1.2.1.6 Education Focussed Accessibility Guidelines

The guidelines and standards referenced in the sections above are an important starting place for development of web content. However, the general-purpose nature of these documents tend to limit their usefulness in producing fully accessible online content particularly for special-purpose areas such as cultural content. Education is a special-purpose area that has received recent attention resulting in numerous guideline documents and policies. Because online cultural content often includes an educational component, we have included two references to two useful guidelines. Developers intending to implement an educational component within their online cultural content should consult these documents.

1. Making Educational Web Sites and Software Accessible: Design Guidelines Including Math and Science Solutions
Produced by the CPB/WGBH National Center for Accessible Media (NCAM) (Freed, Rothberg, & Wlodkowski, 2003)
(http://ncam.wgbh.org/cdrom/guideline/)

This document, which we will refer to as the "NCAM Math and Science Education Guidelines," builds upon WCAG v.1.0 and the various platform specific standards. Unlike both versions of WCAG, which are organized around higher-order principles, these guidelines are organized around features of online formats that are often used in science and math related online materials such as

- Images
- Multimedia
- Forms
- Tables
- Textbooks
- Interactivity
- Graphs
- Math Notation

2. IMS Accessibility Guidelines
Produced by the IMS Global Learning Consortium
(Barstow, McKell, Rothberg, & Schmidt, 2002)
(http://www.imsglobal.org/accessibility/accessiblevers/index.html)

This document, which we will refer to as the "IMS Guidelines," builds upon the NCAM Math and Science Education Guidelines to create guidelines that apply more generally to all online educational materials. Among others, these IMS guidelines include sections on:

- Using XML for Accessibility
- Developing Accessible Synchronous (Real-Time) and Asynchronous (non Real-Time) Communication and Collaboration Tools
- Guidelines for Testing and Assessment
- Legal Issues for Accessible Distance Learning

1.2.2 XML and Interoperable Information

Another important consideration for the accessibility of online cultural content is the format for data storage. While numerous proprietary formats exist, we recommend that content authors consider an XML–based (Extensible Markup Language) markup language because it is:

Interoperable:
XML documents allow interoperability between applications, including assistive technologies, because data can be automatically parsed and transformed by different applications as required.

Transformable and flexible:
XML enables and encourages the separation of informational content from the more arbitrary presentation characteristics. This separation then allows user agents to transform the presentation characteristics so that the content can be accessed in a way that meets the individual needs of user, without the author even having to consider those transformations.

Structured and validated:
XML enables and encourages use of consistent hierarchical nested structures, which make it easier for users to navigate complex content. XML documents can also be validated automatically.

Text based:
XML documents are well-suited to handling textual linguistic content. When this content is in place, transformation between modalities is facilitated, enhancing accessibility.

Note that to avoid creating accessibility problems while working with XML, developers should follow any guidelines accompanying any previously published XML language (see SVG, SMIL, above). While it is recommended that developers use these previously published XML languages where possible, there are situations where no XML language is available. When a new XML language needs to be developed then the following guidelines should be followed:

1. **XML Accessibility Guidelines**
   Produced by the World Wide Web Consortium (W3C)
   (Dardailler, Palmer, & McCathieNevile, 2002)
   (http://www.w3.org/TR/xag)

   This document provides guidelines for designing Extensible Markup Language (XML) formats and applications that take accessibility into account.

1.2.3 Accessibility Focussed Metadata and Information Architecture

When content is created for the web, it is important to add information about the content to aid search and retrieval. This content information is called metadata. Canadian Heritage

Information Network (CHIN) provides information about metadata standards on its web site (http://www.chin.gc.ca/English/Standards/metadata_intro.html). Several standards for metadata are available for different kinds of content, such as learning, multimedia and cultural artefacts. Some common metadata standards for learning technologies are Dublin Core (http://dublincore.org/), CanCore (http://www.cancore.ca/), ARIADNE (http://www.ariadne-eu.org/), CEN/ISS (http://www.cenorm.be/isss/), LTSC-LOM (http://ltsc.ieee.org/) and IMS (http://www.ims.org/). Although the standards are different, there has been a move towards interoperability so that the correspondences between them can be sorted and understood by different search and cataloguing systems.

An important aspect of accessibility in metadata is the specification of user preferences and the ability to transform content and content presentation to meet these preferences. In July 2003, IMS adopted an extended specification for accessibility elements of the Learner Information Package (ACCLIP (http://www.imsglobal.org/accessibility/index.cfm)) that control how a user wants information displayed. Other work has been carried out by the ATRC (http://www.utoronto.ca/atrc/) as part of the TILE Project (http://barrierfree.ca/tile/) to develop metadata standards for learning repositories which catalogue content to be presented in more than one modality, such as a caption for an audio track. Viewers with preset preferences in a Learner Information Package (LIP) would experience, for example, text alternatives to sound.

In developing accessible content for the web, care must be taken to include metadata that reflects content. In this way, individuals will be able to detect and utilize content that is appropriate to their control and display preferences.

1.2.4 Inclusive Usability Evaluation Methods

Meeting accessibility standards such as those outlined in 1.2, Existing Principles of Online Accessibility, is an important first step in presenting cultural content on the Internet. Although the two aspects are closely related, perfect accessibility does not guarantee perfect usability. Good design should include some consideration of usability, ideally from the start of a project. Furthermore, the design and development process should include ongoing usability evaluations. In addition to evaluating how well the design meets user expectations for functionality and engagement, usability evaluations should be assessments of how well the design meets the *Seven Principles of Universal Design*:

1. Equitable Use
2. Flexibility in Use
3. Simple and Intuitive to Use
4. Perceptible Information
5. Tolerance for Error
6. Low Physical Effort
7. Size and Space for Approach and Use

(NC State University & The Center for Universal Design, 1997, http://www.design.ncsu.edu:8120/cud/univ_design/principles/udprinciples.htm)

A number of usability evaluation methods have been developed and the reader is referred to the following Web resources for information on usability assessments shown in Table 1. The reader is cautioned, however, to be aware of the mismatch between the imagined user in accessible design guidelines and the imagined user in current usability evaluation methods. Evaluation methods tend to assume ability is a permanent and given state of the user rather than a temporary and chance state.

Resource Description	URL
Compilation of Usability Resources and References (Jeff Axup, 2002)	http://www.userdesign.com/usability.html
Comparison of Usability Evaluation Methods (UEMs) (based on Gray & Salzman, 1998)	http://www.userdesign.com/usability_uem.html
Efficient & Inexpensive UEMs (by Jacob Nielsen, 1994)	http://www.useit.com/papers/guerrilla_hci.html
HCI Web resources from Online Computer Library Centre (OCLC)	http://www.oclc.org/usability/resources/index.htm
Heuristic Evaluation (Jakob Nielsen)	http://www.useit.com/papers/heuristic/
Usability and Inspection Methods (James Hom, 1998)	http://jthom.best.vwh.net/usability/
Usability Evaluation Techniques (Napier U, School of Computing)	http://www.dcs.napier.ac.uk/marble/Usability/Evaluation.html
Usability Metrics (Napier U, School of Computing)	http://www.dcs.napier.ac.uk/marble/Usability/UsabilityMetrics.html

Table 1: Usability evaluation methods Web resources.

Usability evaluations should include users who use a variety of access modes (e.g. screen readers, screen magnifiers, switches and keyboard navigation) and a variety of skill levels (refers to skills required to engage the material and could include aspects such as comprehension level or previous learning and related experience). Furthermore, care should be taken to use evaluation methods that complement the evaluation participant. For example, a participant who communicates via sign language should not be asked to fill out questionnaires in the written form of a spoken language unless that individual is completely bilingual. Similarly, participants who have low vision should have appropriately sized text in all evaluation materials. Evaluators may benefit from sensitivity training that will better able them to build a good rapport with evaluation participants as well as have a better understanding of evaluation methods and materials that are appropriate for a given participant.

1.3 Key Benefits of Accessible Design for Online Cultural Content

For some time now, culture has been increasingly represented in various online formats including art museums, galleries, videos of dance productions, and art created strictly for online environments. An excellent resource for accessible real-world cultural events is *Design for Accessibility: A Cultural Administrator's Handbook* and it is freely available on the Internet at www.arts.gov/resources/Accessibility/DesignAccessibility.html. Some of the ideas in this handbook are relevant to new media displays or have ideas that are equally valuable for digital content as for real-world content. As cultural content moves onto the Internet, art forms previously unavailable have the potential to become accessible to people with sensory and physical disabilities. This potential is realised by capitalising on the capacity for the digital medium to provide increased user control over presentation formats and speeds, rapid presentation of multimedia content and capacity to transform content from one mode to another

In creating accessible online cultural content, a key challenge facing artists and designers is to create suitable multi-modal representations of that cultural object. There are three possible strategies for achieving this accessible modality translation:

1. To be as exhaustive in a work's description and its context as possible under any particular circumstance;
2. To reproduce an aesthetic experience consistent with the original work;
3. To acknowledge that different media are qualitatively different, and that languages are not the same, and instead create an original new work that carries some of the key feelings and formal values of the original.

It is generally agreed that artists and cultural producers choose a medium of expression for a variety of reasons :

Often, the choice of one medium over another is made because that medium is the one that the artist knows. The artist becomes an expert in a given medium and sees it as an instrument that is an extension of self. That said, in the last two centuries, multidisciplinary artists will choose a medium because it seems the most appropriate for a particular message. These artists are capable of both making art works across a wide range of disciplines, and collaborating with other technicians to put their artistic ideas into effect.

The concept of accessible online cultural content is in step with the skills of the multi-disciplinary or collaborative artist. The expression of artistic works in accessible ways provides exciting possibilities for artists to envision their online cultural works in new ways. While striving to keep "the message" intact, artists are asked to imagine their work from a fresh perspective, and in doing so find that accessibility inspires expression outside of their familiar medium. Providing routes for alternative access invites the artist to explore multiple forms of expression, which in turn, can only stretch and challenge the artistic imagination.

Another positive asset to working towards accessibility in online cultural works is the expanded opportunity to provide additional information about a work's context. Too often, artistic work is presented without context, even though we know that context can be a critical factor in understanding creative expression. For example, a particular photograph may reference important aspects of the history of photography through its subject matter and style. The photographer hopes that the viewer will find those cues within the visual language of the photograph because doing so greatly enriches the experience of viewing it. Realistically, however, only those with the relevant background knowledge are able to place, and therefore, appreciate the photograph in the intended context. Hence, one of the key benefits of creating access to cultural content, is having the opportunity to better represent those works within the particular contexts in which they are created, presented and interpreted. In transforming a work for accessibility, there is an opportunity to add additional information regarding the work's history, creation process and cultural significance that might not be feasible within other paradigms. Inclusion of context strengthens the value of "translation" for access; it becomes an enhancement for all audiences who might otherwise miss the artist's allusions.

Providing accessible equivalents for online cultural works offers many more challenges and benefits to our larger notions of culture and community than just these expressed above. The following sections will articulate several of these challenges and benefits and will provide a starting place for artists and designers to begin contemplating extending their notions of audience to include people with disabilities through extending and expanding their ideas connected with the creation of online cultural content. We intend these sections, like much of this document, to provoke as many questions as they answer, about creating Canadian online culture that includes us all.

1.3.1 Accessible Equivalents for Deliberately Challenging Interfaces

One of the key roles of artists in Western culture is to present ideas, images and experiences that provide us with new understandings of who we are as people, both emotionally and intellectually. Artists are often willing to confront that which is difficult or different and will articulate challenging ideas through various forms of expression. As a result of this role of 'provocateur,' artists are sometimes misunderstood within their own time, only to be highly regarded by subsequent generations who better understand artistic works that the artist's own contemporaries had viewed as difficult. "Artistic license" is the term used to express the right of an artist to create challenging work, regardless of the expectations of others.

At times, artists demonstrate artistic licence through the creation of works with complex interfaces that are intentionally difficult to access, and whose meaning is at least in part derived from this tension. A key challenge in making work accessible may be not to transform the work into something easier to access, but instead to create something in an alternate modality which captures the equivalent sense of artistic expression intended by a piece.

Artists who develop works that are intentionally challenging to access may want to consider offering their work in the range of alternate modalities outlined later in this document. It is clearly easier for an accessible alternative to be developed if the artist is considering multiple modalities as a means to providing access from the work's inception, and we encourage these artists to integrate the ideas espoused in this document into their own design practices.

Commissioning parties, in an effort to design inclusive collections, may also ask artists to consider placing their individual expression into a form that is accessible to the audience. Artists are often willing to create work for specific contexts and to balance pure expression with the communication of ideas. It is hoped that this discussion document may form a starting place for artists in planning and conceptualising such a piece, particularly those later sections related to modality transformations.

It is important that the freedom to develop pieces that are challenging be recognized as a key part of artistic freedom, and it is not the intention of this document to discourage such challenging and provocative developments. Instead, artists are encouraged to take their original work and try envisioning how the meaning associated with it can be successfully extended into different mediums.

1.3.2 The Role of Aesthetics

Aesthetics are the criteria used to judge the qualities of an artistic work including its formal and expressive qualities, its relationship to context, and its relationship to cultural history. For example the minimalist aesthetic is different from the folk art aesthetic however, in either case, there is a starting point for discussion/interpretation/understanding of the work that relates to how the aesthetic is a reaction to things such as other aesthetics and social change. The consideration of aesthetics can be thought of as both an issue related purely to considerations of representation and form and also as a feature of access itself. Considering aesthetics as connected to the production of accessible art works represents a challenge to artists and designers of online cultural content, as each media (video, music, text, images) has its own set of aesthetic values and concepts, which do not always move easily across forms.

When examining the intersections between aesthetic considerations and the translation among modalities of online cultural works, the key is to resist viewing the translation among modalities as a strictly functional operation. The artist or producer may carry out distinctly different conceptualization and design processes depending upon whether the work is being developed as accessible from the start, or whether a an existing work is being transformed into one that is accessible.

In the first case, the creator may be freer to explore alternate modalities and options for expressing their ideas across a wider range of media from the onset. It is possible, in this scenario, for those commissioning or producing the piece to work with designers to make the online cultural content and all of its elements stylistically consistent and coherent in relation to an aesthetic direction. Accessible art works should have formal and communicative values that are as aesthetically effective as other art works.

The latter situation, that of transforming an existing work into one that is accessible, requires a reasoned balance between the work's original expression and what can similarly be achieved in other media. Depending on how closely the designers of the transformation are working with the original designers, there can be negotiation of a change in aesthetic direction. It is important that this kind of negotiation occur with a shared sense of respect for the original design aesthetic. Linking old and new elements together can be challenging and requires that stylistic/aesthetic coherence be considered desirable and preferable. It may even be possible that the complexity of such an undertaking results in the decision not to renovate completely and restyle a particular online cultural piece, but instead to develop a new accessible alternative that captures the intention of the original. Either way, translated art works should continue to be aesthetically successful in the new medium.

Regardless of the specific tactics undertaken, deciding to consider questions of aesthetics in the context of creating accessible online culture is a worthwhile pursuit given that the online cultural content's "attractiveness" ultimately has value, insofar as it encourages a continued use and enjoyment of those materials by its viewing audience.

Finally, as it is wise to periodically update the aesthetic/style/design of any online cultural content in order to keep an interested audience, it is recommended that issues of accessibility of those materials be examined whenever an update of the content occurs.

1.3.3 Entertainment and Engagement Values

"Entertainment value" is a term used to describe the capacity to capture the attention of a popular audience and keep them watching. Entertainment value also describes emotional involvement with a cultural work. To make a work entertaining requires mobilizing the craft and artistry of a medium around content that appeals to audience members. It implies a total immersion in the experience, also known as the "suspension of disbelief," in which viewers knowingly set aside their notions of reality in order to engage with the alternative reality put forth by a work of art. When an art or media work has so succeeded at this creation of an alternate reality, it is referred to as having "suture" because it has been sewn together with perfect, invisible, seams.

Narrative forms are most often the vehicle for entertainment and, depending on the medium, narratives follow certain formulae to achieve this sense of seamless flow. In filmmaking, for instance, audio is often used to link scenes, and a familiar shot language is drawn upon to suggest transitions of time or space. It is important to note, however, that these media-based conventions can change as result of cultural and technological shifts. For example, early filmmakers were extremely literal in their depiction of time transitions, often even following a character or story line in real time. Today, time transitions are more inferred and less literal, and are achieved by drawing on a modern audiences' accumulated knowledge of film conventions.

Finding ways of indicating emotional cues, and rhythms is key to successfully transforming an online cultural work while staying true to its emotional content. Having a consistent way of indicating an emotion, point of view or shot, could potentially build a parallel language for accepted visual and audio conventions for viewers who might not be able to see or hear the content of such a piece. As a large part of a viewer's interest in a work is derived from the emotional content of the work, exploring alternate ways of making this information available to people with disabilities remains central to the challenge of creating accessible online cultural content.

Designers or artists responsible for transforming online cultural material into accessible alternatives need to pay attention to narrative conventions and processes and the various ways that a specific medium carries emotional impact and think about parallel strategies for alternative modalities. If the intention of an online piece is entertainment, then it is integral that the artists/designers will find a way to maintain the entertainment value associated with a piece when offering an accessible alternative.

The engagement value of an art work is also an important concept to consider when producing an accessible alterative. Engagement assumes a critical and media literate approach to cultural content and asks that the audience figure out the work without being fed all the answers by the artist. Engagement assumes the right for art works to be ambiguous, and requires that the viewer interact with a work in such a way as to have space to interpret the work, by picking up on sometimes subtle cues offered by the artist. This ability to unpack or disentangle how cultural products engage or seduce audiences is valuable piece of how people engage with cultural works. As such, it too is a factor that should be considered when both making inclusive works, or translating existing works into accessible alternatives. Artists and designers should be cautioned to not completely spell out all of a work's ambiguous qualities when making an accessible version, but instead to find creative ways to incorporate the same sense of engagement in the accessible version, by creatively exploring the medium and it's capacity to maintain the engagement value of the original.

1.3.4 Perspective: A Cultural and Technical Consideration

Like many elements of culture, perspective is both a technical element and an element that suggests a specific culture. In creating an accessible alternative to an online cultural work, particularly those that are primarily visual in nature, it is important to understand and consider the perspective of the piece from both a technical and cultural point of view.

Perspective, in a technical sense, means the visual devices used to create depth of field within a painting or photograph, and which allow subjects to be arranged for the viewer in a hierarchy within the picture plane

In the cultural sense, perspective is arranged and understood in terms common to a shared experience and is influenced by cultural references that derive meaning from a cultural understanding. For example, while the dominant arrangement of the picture plane in Western culture is to have a depth of field with images receding into the background, Australian Aboriginal sand paintings present images across the visual plane suggesting a sense of continuity of space and time, a concept deeply linked with the Australian Aboriginal worldview. Inuit paintings that incorporate transformation processes within a flat plane also use depth across a canvas to suggest that events happen within cultural notions of time.

In creating an accessible equivalent of an online cultural piece, it is valuable to reference the ways that perspective act within a given art work. The perspective of a work can impart a great deal of information about who created the work, what cultural communities and histories the work references and what the viewer's relationship to the work is intended to be. It is important that this information be treated as central to the creation of an accessible equivalent work.

In creating an accessible equivalent of a work, it may also be possible for designers and artists to explore other notions of perspective offered by a wide range of artistic communities. Artists and designers should be encouraged to explore the range of ways that perspective can be implied in alternate modalities.

1.3.5 Interpretation

This discussion considers interpretation through two lenses. The first considers the ways that specific languages are not neutral, but informed by their cultures. The second discusses the subjectivity that is fundamental to reading works of art.

When we take a word or concept from one language and explain it in another, we place it into a different cultural context. Cultures use language in different ways. Nicole Brossard (1987) made this point in her book *The Mauve Desert*, a story told through the "voice" of the French interpreter and then that of the author with quite different nuances. This use of characters' cultural voices illustrates the dynamics of interpretation. When works of art, literature, and music are interpreted for the deaf, these take hold some of their original meaning but are also forever transformed through language. The same is true in reverse. Deaf culture, created in sign language, would have a different meaning for the hearing. The following statement from the Sign Language Center's web site reinforces this point

> Deaf culture is the celebration of pride in the history of the traditions of the community. The knowledge of American Sign Language (ASL) is what truly hold the community together and how socialization is passed down to succeeding generations. As a visual language, social mores, and rules are affected. Eye contact and body language, for example, is key in effective communication with a group of communities visually. These tools are often overlooked communicating orally(Sign Language Center).

As the popular film states meaning can be "lost in translation," Yet at the same time, as the film also illustrates, new kinds of knowledge and culture emerges in the friction between original meaning and culture and translation. It is in this exciting context that art that crosses the boundaries of ability and disability must be considered. While culture is often the route to another's experience, language can be a difficult issue particularly when the lived culture behind the languages differ. Xose Roig, a professional translator has remarked, "the single most important aspect of being a translator is not the aspiration to master both languages, but rather both cultures" (Roig, 2004). These underlying differences and cultural nuances are an important part of any interpretation. It may be useful to consult with a professional translator who will be able to provide cultural insights or writings in the profession such as *Translation Journal* to gather information about approach when translation of a work is undertaken. Also, issues of translation have been tackled in the literary arena and information about translating literature may provide guidance in ways to manage interpretation of non-literary cultural objects.

The second issue in interpretation is subjectivity. There is often not one fixed meaning or interpretation for an artwork. The artist usually has an understanding of how they want the work interpreted, and over time, there may develop general agreement about the key readings of the art, but the audience, may have quite different and very subjective perspectives. One of the values of cultural works is that they open new experiences of the world for the audience, experiences that are not always easy to describe. This subjectivity makes it hard for the translator to explain a work of art. While, as the interpreter can describe the artwork as a physical entity, a description of the emotional qualities of a work is subjected to the interpreter's experience and engagement with the work. For example, when the National Gallery of Canada bought a Barnet Newman painting, some were delighted while others were offended.. Also, interpretations of art works change over time, as art is constantly developing new practices. Hence, the interpreter of an artwork needs to recognize that there is personal and historical subjectivity involved. Finally, the interpreter must take care that qualities of the artwork that are not literal are described in the interpretation process.

1.4 Moving Artwork Across Modalities

Movement across modalities refers to the re-articulation of the expression to a form that is perceived by a different sense from the original. For example, a film that is perceived fully in audio and visual modes could have modality translations that include audio descriptions of visual elements (described video) and text descriptions of audio elements (captions). In this way, the film may be fully perceived in just the audio or just the video mode without loss of significant information. Different media have different expressive qualities. Artists use a medium, or a combination of media, in order to create specific effects that are aesthetic, and in order to reference particular histories of practice related to a particular media. (For example, video art often references television conventions). The complexity of considering translating work into different mediums deepens when also considering that audience members, either because of disability or lack of exposure to a particular media, may not necessarily have the experience of the original medium. Information sheets and audio guides at galleries, for example, attempt to fill this gap between common knowledge and specialty knowledge that may alter understanding of the cultural object and this model may be a good starting point for modality translations.

The following sub-sections will explore aspects of cultural objects that require specific attention when modality translations are carried out: emotion, perspective presentation and cultural differences.

1.4.1 Translating Emotion Across Modalities

Finding expression for narrative intensity and for emotion is one of the most demanding aspects of creating alternative modality equivalents. Artists and designers are encouraged to be creative and explore both a range of options from within their own chosen media and others in order to arrive at creative solutions for the transmission of emotional content. Within an audio format, for example, adding simple devices like volume and directional shifts, or layered sound effects to suggest a particular physical setting are some simple ideas for getting at emotional content in a work. In the visual format, colour coding may be an effective strategy for expressing emotion.

Perhaps it is useful to look at an example of a translation of an art work in order to better elucidate these concepts. In this case, consider the elements that may be involved in order to express a sculpture (largely a visual modality) through sound (auditory modality).

Working with sound that moves through the dimension of the screen, with different physical volumes represented by changing qualities of audio might provide a non-sighted viewer with a sense of the movement of the sculpture. Working with a composer who can create a melody that feels like the sculpture and incorporates these technical suggestions could further enhance sound that could capture the sculpture's atmosphere.

Cross-disciplinary teams are needed for effective translation. Multimodality will always work better when incorporated into the original plan for the artwork. Also, whenever possible, the user community should be involved in the development of the work from the planning stage. This participatory design strategy will mean that an effective language emerges as the artwork develops that is satisfactory to the artist and the community.

Box 1: Accessibility from an Artist's Perspective *by Sara Diamond*

Artists will have varying reactions to the request that their work be accessible. Many of them will find the idea of representation of their idea in another medium to be a difficult yet intriguing challenge. Artists who have a range of disciplinary skills, or who share a curiosity about other media will best meet this challenge. The goal is to produce a professional quality artistic experience from the original concept through all media. This again reinforces a team-based approach, where artists with skills in different areas can help each other with multimodal transformation. Artists will need to know that they have access to resources to undertake this work.

In this document we stress the importance of participatory design involves the user from the beginning. Creative artists engage with their subjects, internalize the experience of them as much as they can and create from that basis. Recent success in community arts shows that these projects can be inspiring for those involved, for the artist and for the community, bringing out their skills.

However, most artists are not trained to work in this collaborative manner or if they are, it's in the hierarchy of film or television production. Participatory design requires that the artist revise his or her vision as the creative process emerges—this is not always possible and can also weaken the artwork. Many artists work in a solitary way. A system of "sketching" or "prototyping" and testing the perception of the work may be a viable strategy.

1.4.2 Provide Multiple Perspectives

The issue of the “adaptation perspective” is a critical one for content in the cultural domain. In the educational domain, this perspective is not such a problem, because the perspective of the adaptation can be primarily driven by the intended learning outcomes. For example, when a particular piece of artwork is described, with the intention of teaching a straightforward concept, such as a particular lighting technique, the description is likely to be fairly impersonal and may focus on the mechanical.

On the other hand, in the cultural domain, where the intention is to convey, at least to some extent, the aesthetic experience of the artwork, any description will depend to a far greater extent on the expertise, experience, and personal taste of the describer. A work will “mean” something different to each person who experiences it, so there is no truly “neutral” perspective. For people who must experience some or all aspects of a work through the lens of someone else’s reaction, the result will inevitably be subjectively influenced by the opinion of the describer.

A possible solution is to provide multiple, clearly labelled, perspectives for each work. The viewer can then pick and choose from the range of perspectives available in order to construct their own experience.

Potentially enlightening perspectives might be provided by:

- **Artist/designer/composer:** As the creator of the work, this is often an extremely valuable perspective, although the artist’s view should not necessarily be treated as “the last word” on what a piece of content “means.”

- **Curator:** Curators often work towards the development of a theme in shows that they produce, and choose artwork for particular reasons related to it. The curator’s perspective on why a particular piece appears in a show or gallery can be extremely enlightening for those accessing the content.

- **Other “experts” on the work:** Gallery cards/dossiers already exist for many works, and can be accessed as a source of perspective on a piece.

- **Contemporaries (other artists, friends, family members, etc.):** Mixed perspectives from individuals who know the artist or artists who will understand the influences of the artist can add a depth to the work not otherwise available.

- **Others:** This could include descriptions and reactions from members of the general public. There are many ways to do this, ranging from simple feedback systems, to iterative systems that allow visitors to interact with and modify each other’s submissions on what the work means to them.

We suggest providing the available perspectives in an “on demand” manner, controlled by the user, allowing the user to experience as much, or as little, as they choose.

1.4.3 Consider the Presentation Context

All curators consider the context in which art works will be presented when they create an exhibition and the presentation context lends meaning to the work itself. The web increasingly provides gallery spaces for web art and design. An online work may have been created specifically for a web based environment and may not have any “real life” equivalent; other online cultural objects represent art pieces that have been created for a physical environment.

In translating or providing descriptions for online cultural pieces, considering the presentation context is key. If the piece is being described, it may simply be enough to mention the works presentation context. If instead a work is being transformed into an entirely different modality, artists and designers should consider the meaning imparted by the original presentation context to the piece and determine ways of imparting some of that meaning in the transformed version of the work.

1.4.4 Considering Cultural Differences

Integrating cultural specificity into the process of transformation is important. Artists may want to work with others from their own cultural background when transforming or developing an expression in another modality. They may also consider consulting those within the community (for example, cultural elders) who are knowledgeable about cultural practices and conventions in order to develop a culturally sensitive, and accessible, transformation of their expression.

In those cases where someone other than the original artist is responsible for transforming a work into an alternative modality, it is integral that they consult the artist directly for guidance. Should it not be possible to contact the artist, guidance should be sought from members of the cultural community represented before moving ahead with modality transformations.

Beyond issues related to translating culturally situated online materials into accessible formats, the storage and presentation of these works in an online electronic archive is an example of another situation that requires cultural sensitivity.

Many Aboriginal communities throughout the world have different rituals concerning some of the major elements of life, death, birth and the afterlife. In an Australian Aboriginal community, for example, when a member of the tribe has died, all of their knowledge and all physical items are destroyed. The name of that person is no longer used in that community, and any other member with the same name has to change it out of respect for the deceased. This is a situation that requires that those responsible for presenting related materials online and/or transforming those materials into an alternate modality approach their roles armed with knowledge about acceptable community practices.

In cases such as the one described above, it is highly recommended that those who are responsible for cataloguing or posting culturally situated art works or artefacts online be knowledgeable about the cultures from which the works are derived. Consultation with the communities represented should be undertaken before attempting to make these works available in an online environment, and any concerns or issues that surface throughout the course of consultation should be adequately addressed on a case by case basis. Consensus concerning the presentation, use, transformations, descriptions and or any other manipulating of the works should be reached with representatives of the community before moving forward with making the material available.

Some existing online cataloguing projects have addressed the concerns related to hosting culturally sensitive works online through the addition of a Cultural Use Statement, the terms of which users must agree to before entering the site. A Cultural Use Statement may include the legal rights pertaining to intellectual property issues, but may also demand that the use or transformation of any of the online materials be approached from a position of respect for the culture represented. The Cultural Use Statement can also be used to offer any specific instructions the user may need in order to treat the materials in a respectful and sensitive fashion

Additional measures that may be undertaken to ensure that cultural content online is treated appropriately may include requiring that some materials be password protected, or require permission from the site's administrator to be viewed. It may also be possible to ensure that such materials may only be viewed directly from a remote server and not be available for download to a local computer.

1.4.5 Integrate with Workflow

In general, accessible design is far more difficult to achieve when the measures required for accessibility are left to the end of the design process. By that time, fundamental decisions have already been made about the format, style, and layout of the content that may have inadvertently introduced unnecessary barriers to access. Revisiting these earlier decisions often results in a cascade of further changes. As these large-scale changes may be impossible due to scheduling or budgetary constraints, well-intentioned goals of accessibility may be ignored.

Considerations:
When accessibility is integrated into the workflow of online content development from planning, to storyboarding and through coding, barriers to access are flagged earlier and can be addressed without requiring major changes. A positive side effect of this approach is that other quality assurance issues, such as general usability and platform independence, are often dealt with more easily when developers have taken greater care in the planning process.

1.5 Introduction to Online Content Modalities

This section explores further the concept of content modalities in the online medium. Translation between these modalities is discussed in section 1.6, Modality Translations. In order to understand modality translation, it is important to understand the properties of the modalities that are available. In general, sensory modalities derive from the full capabilities of the familiar human senses: sight, hearing, smell, taste, touch and the less well known sense of kinesthesis (the sense of movement and position of the body). However, in online settings the range and quality of sensory modalities available are reduced by the practical constraints on representation and output imposed by the online medium and the computer systems used to access that medium. The first senses excluded from the online context are smell and taste. Interfaces for these senses are an area of research, but no practical applications exist. The modalities that remain--visuals (see 1.5.1, Visual Modality), audio (see 1.5.2, Audio Modality) and haptics (see 1.5.3, Haptic Modality)--are affected to varying degrees by the online medium as explained in the sections that follow.

In addition to these sensory modalities, is a more abstract modality: language (see 1.5.4, Language Modality). This modality is not experienced directly; instead it is experienced via transformation into any of the three common online sensory modalities: visually as text or gesture, aurally as speech and tactilely as braille.

In the previous section, the more abstract aspects of cultural objects were discussed: emotion, perspective, cultural influence and artistic license. The following sections introduce each of the common computer sensory modes as well as the more concrete aspects of cultural objects that are conveyed through these modes.

1.5.1 Visual Modality

The *visuals modality* refers to content that is perceived by the sense of sight. Once content is perceived, the visual system determines information such as colour, shape contours, movement, and location in space. In addition to these very basic characteristics, a large number of higher-order visual properties have been identified in visual cultural objects. These properties include composition and layout. Language may also be present in the form of text or sign language images. When a visual object is translated to another modality, the information provided by these concrete properties of the visual object should be utilised by the translator. Further discussion of the translation of both simple and complex properties of the visual modality is available in section 1.6.2, Alternative Modality Equivalents of Visual Content. In the same way, the properties of visual objects may be useful for providing an alternative equivalent of an online cultural object. When visuals are utilized, the following should be considered:

- Linguistic information such as text must be read serially, which is slower and requires more attention than non-linguistic, graphical information. Graphical information can be communicated in a highly parallel fashion.
- Interference can occur if linguistic information in the audio and visuals does not match.

1.5.2 Audio Modality

The audio modality refers to content perceived by the sense of hearing.

Once audio content is perceived, the auditory system determines perceptual characteristics such as loudness, pitch, and sound localization. Language may be included in the form of speech or singing. When an audio object is translated to another modality, the information provided by these concrete properties of the audio object should be utilised by the translator. Further discussion of the translation of both simple and complex properties of the visual modality is available in section 1.6.3, Alternative Modality Equivalents of Audio Content. In the same way, the properties of audio objects may be useful for providing an alternative equivalent of an online cultural object. When audios are utilized, the following should be considered:

- Speech is well-suited to linguistic information and proceeds faster and requires less attention than reading.
- Speech is usually limited to one speaker at a time.
- Music is processed quickly and has a strong emotional component.
- Sound effects are processed quickly.
- There is a strong expectation that sound effects will be coupled with actions or events.

1.5.3 Haptic Modality

Note: The number of home users with access to haptic devices is relatively low at present, but may increase in the future.

The haptic modality refers to content perceived by the senses of touch (via sensors in the skin) and kinesthesis (via sensors in the muscles, tendons and joints). The sense of touch is sensitive to pressure (that may be further perceived as vibration and texture) and temperature. The sense of kinesthesis is sensitive to the movement and position in space of body parts. Information from this sense is vital for contextualizing the sense of touch;. without knowing the position of one's fingers, one would not be able to tell whether the pressure felt across the hands and palms was due to a ball or a flat surface. The haptic senses are perhaps the most "exploratory" of the senses, due to the necessity of reaching out and physically manipulating objects in order to form perceptions.

In online settings, these concrete aspects of the senses of touch and kinesthesis are made available by haptic devices that translate signals from the computer into physical forces. However, only the most sophisticated (and expensive) devices are able to convey them all.
Further discussion of the translation of both simple and complex properties of the visual modality is available in section 1.6.2, Alternative Modality Equivalents of Visual Content. In the same way, the properties of audio objects may be useful for providing an alternative equivalent of an online cultural object. When audio is utilized, the following should be considered:

- There is a strong expectation that haptics will be coupled with visual actions and sound effects.

- While haptics can convey certain types of information very well, the bandwidth of the haptic devices currently available are quite limited in comparison to the visual modality.
- Only Braille users will be able to access linguistic information via haptics.

1.5.3.1 Haptic Devices

Because haptic devices differ from each other far more than screens or speakers, it is useful to compare what is available:

Note: Haptic devices are designed to be held in the hands are often designed to serve a dual role, as producers of output forces (display) and detectors of user input forces (control).

- **Refreshable Braille Displays:**
  Devices that, in association with text-to-Braille transformation software, produce Braille output via bumps or pins that raise and lower in real-time as required. These devices may include directional keys for document navigation. Hard copies of Braille information can be created with Braille embossers.

- **3D full force feedback devices:**
  These devices incorporate specialized servomotors that resist or constrain hand motion. They allow users to experience virtual objects via physical traits such as resistance, solidity, or viscosity. Examples of such devices include the "dataglove" with an external articulating exoskeleton and boom-mounted stylus devices that provide six degree-of-freedom (x, y, z, yaw, pitch, and roll) interaction. These devices are generally expensive, but the realistic quality of their output makes them well-suited for applications such as surgical simulation, molecular research and immersive environments.

- **2D force feedback devices:**
  These devices are usually mice that are constrained both to a flat plane and by an arm governing left-right, near-far position. While such devices showed promise, they were commercially unsuccessful and are now hard to procure. Force feedback joysticks are relatively common, they are of very limited use outside of gaming applications.

- **Tactile feedback devices:**
  Tactile feedback devices are more affordable, and therefore more common. These devices usually vibrate to convey taps, bumps, boundaries and textures rather than move on their own or restrict movement. The prime audience for tactile feedback devices is gamers, but applications have and are being developed for blind or low vision users, or for those who learn best through tactile or kinaesthetic sensory channels.

- **Techniques and applications:**

Tactile desktop technologies, software that enables the haptic device to sense the desktop, such as TouchSense and the FEELit Desktop technologies by Immersion Corporation (http://www.immersion.com), allow users to feel their way across the computer "desktop" (locating and selecting icons and menus) or feel web content.

- **Hybrid Displays**:
  Displays such as Tactile-Visual screens and fingertip/pad or nail-mounted tactile displays have yet to progress beyond research labs

1.5.4 Language Modality

Language is useful for translation in that it bridges all three of the sensory modes common to the online environment; it can be represented visually, audibly and tactilely through automatic translation routines that function at the processor level. The first of these automated translation routines, from language to visuals, is familiar to anyone who has seen text on a computer screen. The second automatic translation routine, from language to audio requires text-to-speech (voice synthesis) software or hardware that produces a series of sounds that mimic human speech on the basis of the linguistic information. The third automatic translation routine, from language to haptics makes use of a text-to-Braille software that feed into a refreshable Braille display or Braille printing embosser machine.

The existence of these automatic translation routines makes the language modality a critical go-between for translations between the sensory modalities. When translating to or from the language modality, there are several characteristics that should be considered such as pacing, word choice, sentence length, tone, genre, rhyme and rhythm. The full nuances of written language are well documented in the literary field and we recommend that translators look to this body of work or collaborate with experts in this field to achieve an equivalent alternative of the original object.

Further discussion of the translation of the language is found in the relevant modality discussion of section 1.6, Modality Translations (e.g. text is discussed in the Visual Modality section). When language is involved in a modality translation, the following should be considered:

- Language mastery varies with age, order language was learned and learning disability
- Linguistic information is processed at different speeds depending on the sensory mode in which the information is delivered

1.6 Modality Translations

Modality translation refers to the creation of equivalent alternative representations of online cultural content. Modality translation is necessary remove barriers to access experienced by individuals who have sensory impairments. Modality translation widens the audience for any online culture object to include all individuals rather than limiting participation by ability to perceive information in any one modality. A further benefit to modality translation is that the audience can experience the cultural object in a variety of modes that may in turn enhance their understanding and experience of the cultural object. Artists too may find that their creativity is positively challenged and stretched by the creation of multiple access modes to their artistic expressions.

1.6.1 Twelve Possible Modality Translations

The three sensory modalities and the language modality can be inter-connected with each other by means of twelve potential *modality translations*[1].

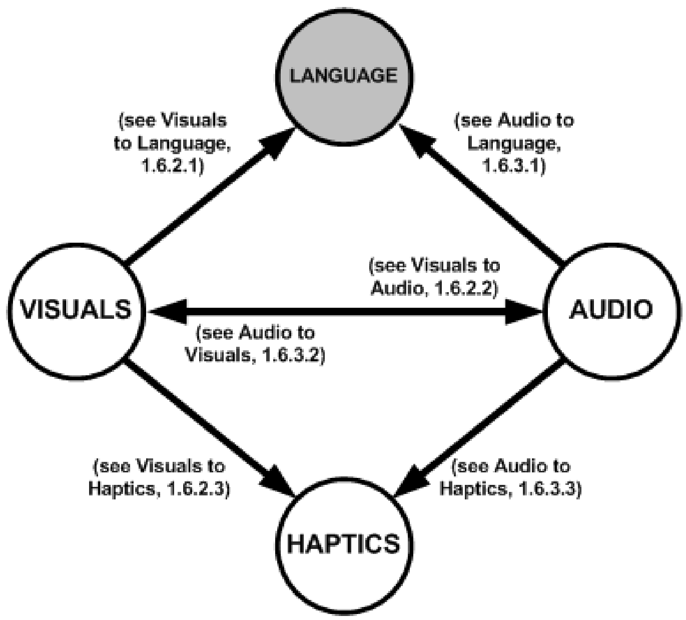


Figure 1: Potential Modality Translations.

[1] The term "translation" as used here can include relatively strict transformations (as between text and haptic Braille) as well as more interpretative translations (as between visuals and musical audio).

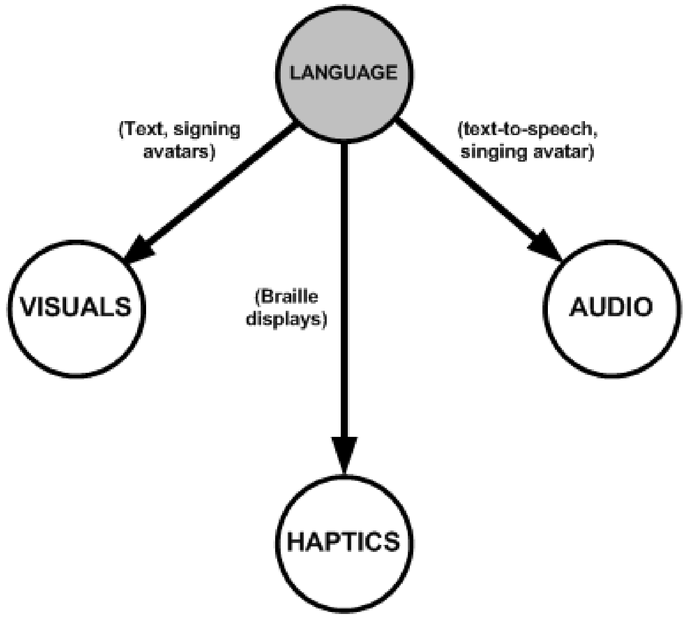


Figure 2: Language bridges all three computer modalities and is a flexible medium for modality translation.

Three of these potential modality translations are *from* language and require little explanation:

- **Language to Visuals:** Translated automatically and without information loss into rendered text, displayed on a monitor screen.
- **Language to Audio:** Transformed automatically and with little information loss into synthesized speech, played from a speaker.
- **Language to Haptics:** Transformed automatically and without information loss into Braille, displayed on a refreshable Braille display.

Three more of the potential modality transformations are *from* haptics and are not covered in depth in this document because little work has been done in these areas. However, translation of real-world haptic information is fairly common; some examples are briefly described here:

- **Haptics to Language:** Text that describes any haptic information of objects, including texture, weight and temperature.
- **Haptics to Visuals:** The texture of a real-world object is often apparent in visual images of that object.
- **Haptics to Audio:** When a real-world object is rubbed or dropped, the texture or weight of the object is often apparent in the resulting audio information.

The six remaining modality transformations are covered in greater depth in the sections that follow.

- Visuals to Language (see 1.6.2.1)

- Visuals to Audio (see 1.6.2.2)
- Visuals to Haptics (see 1.6.2.3)
- Audio to Language (see 1.6.3.1)
- Audio to Visuals (see 1.6.3.2)
- Audio to Haptics (see 1.6.3.3)

1.6.2 Alternative Modality Equivalents of Visual Content

The following sections provide an in depth discussion of the translation of visual content to language, audio and haptic modes.

1.6.2.1 Visuals to Language

Language as an alternative mode to visuals involves the use of text or speech to label or describe both static (e.g., photographs, paintings.) and dynamic (e.g., film, animation) visual information.

Visuals are a powerful communicative tool for ideas and emotions. However, though it is said that a picture is worth a thousand words, to individuals unable to see due to a visual disability or circumstance (e.g. a low bandwidth connection or an audio browser in an eyes free setting like a car), the picture has no worth at all.

A variety of methods exist for adding text descriptions of one kind or another to visual online content. Some of these methods are listed here, but the reader is referred to the WCAG v.1.0 (Chisolm et al., 1999) for more detail:

- For simple images, short animations, buttons or hot-spots, **short text labels** (e.g. <alt> text in HTML) can quickly provide the user with sense of the appearance and purpose of the visual element.

- For more complex images or simple animations and videos, the short text labels can be supplemented with longer **text descriptions** (e.g. <longdesc> text in HTML).

- For complex animations and videos it may sometimes be appropriate to provide **text transcripts** that combine captions (text format of audio) with text descriptions of visual information in the video (descriptions of the actions, body language, graphics, and scene changes of the video track).

- For geospatial information, please see Companion Document 5, Representations of Visual Geo-Spatial Information.

There are added benefits to including these kinds of text descriptions. For example, alt-text descriptions of images stored in a database can also serve as searchable meta-data to enable queries. Information such as the title of an artwork, artist and the year the artwork is produced can be included in the image tag as <alt> text, and a longer description of the visual media including colours, visual composition and emotions evoked can be described in a longer description, as in the case of <longdesc> text. Other keyword searches of long descriptions or text transcripts would also be possible opening up new ways of searching and studying visual media. For example, an individual interested in the lexicon of body language may search for examples of body language in film through a keyword search of the video description.

General Techniques for using Language to Describe Visuals

The technical guidelines for including language-based descriptions with images are well-established, however, the form of the language description of cultural objects is less clear. The intent of art or of the artist varies; for example, it may be to shock, to educate, to comment, to entertain, to vent, to explore or to express. This variety makes it more difficult to provide stock guidelines for how to create language alternatives for visuals that are in digital form. Some aspects to consider including in the description, however are:

- The appearance of the visual (e.g.: genre, dominant colours, object placement, textures, setting, use of light, use of movement).
- The context of the visual (e.g.: the discourse around the visual that helps us to understand it).
- Use of conventions (e.g. flashbacks in video or bare feet in a modern ballet video) or borrowing of conventions (e.g.: use of pose, countenance and background common to renaissance religious painting in a 20^{th} century painting).
- Meaning of the aesthetic experience of the visual to the alternative content creator.
- Essential elements of the visual (e.g.: the words of a letter shown in a movie, the brushstroke of a painting or a character listening in the shadows).
- Instructions to take the same pose of a subject to stimulate a sympathetic kinaesthetic sense.
- The different goal/interest/age of the viewer (e.g.: descriptions might vary for a curious visitor, for a student of art history, or for a young person.
- Any subtext in the visual (e.g.: emotions, mood, or autobiographical information).
- The most appropriate format of the description to mirror the visual (e.g.: pace, tone, emotion-laden text, poetry or prose, formal or informal diction).

The following will examine specific kinds of visuals and considerations for the making of equivalent text versions more closely. First photographic images will be discussed, then moving images and finally interactive images.

As with all art forms, a dialogueue and means for examining and understanding **photographic images** has developed. These conventions of discussion may be useful to the individual who creates text alternatives for photographs. For example, When providing descriptions of photographic still images, it is important to draw attention to the photograph's punctum and focal area. Punctum is that element in the photo that touches the viewer emotionally. For some, punctum is the key to understanding the essence of the photograph. Photographic images also give the appearance of depth through depth of field, as a photographic image always contains a foreground, field of view (range of focus) and background.

In the case of **moving images**, the additional element of time enables a narrative or story to be told to the viewer. The narrative unfolds through a series of images that reveal action within the image's frame over a period of time. The illusion of motion occurs either by the subject movement across the frame or by moving the camera's field of view over the subject. In either case, moving images direct the eye to a focal point—an image or symbol that is important to the development of the story. The added complexity of time and motion to a sequence of images may be confusing or cause difficulties for some viewers. Ways in which moving images can be enhanced or clarified by text descriptions are discussed in the following.

While it is more common to use speech to describe video (see 1.6.2.2 and Companion Document 3 Online Video Description), there are a few instances where it is useful to provide a text description of the visuals that appear on-screen. Text descriptions of only the essential visual elements usually appear with captions and are referred to as collated text transcripts. These descriptions might be used by a blind or deaf-blind viewer who is accessing the content via Braille display in an area where sound is either not appropriate (e.g. a library) or not possible (e.g. no speakers), or a viewer who needs help to understand the important visual elements in a video. Collated transcripts are usually brief to keep demands on the viewer reasonable. However, full text transcripts of the visual portion of a video could include such information as the set design, ambience, scenery and appearance of characters (i.e. costume and emotional expression). In this way, the full richness of the visual information can be accessed and read by a screen reader, viewed on a Braille display, or transformed in size to accommodate a person who has low vision.

Interactive images are hybrid images that can be displayed in either a browser or kiosk environment. These images serve a function, in addition to their aesthetic value. Interactive images include icons, buttons, thumbnail images, image maps and animated gifs.

In all of these instances (with the exception of animated gifs and animation sequences), the user is informed via a visual cue such as a button, how to perform an action such as "go to the next page." If a user is unable to see the button and the button is not accompanied by alternative text then the user will not have any information about the function of the button. Text that is embedded in images is not accessible to screen readers which only convey the electronic text used to encode the language modality; embedded text is an "image" to screen readers and cannot be translated.. Iconographic images such as button that looks like an envelope to represent "send email" present a similar challenge for screen readers. Icons rely on visual language (i.e. symbols) that represent the function of a button. Although icons are beneficial in that they transcend language barriers, they still require alternative text in order to remove vision barriers.

1.6.2.2 Visuals to Audio

Audio as an alternative to visuals involves the use of audio information (e.g., voice, music, sound effects,) to convey visual content.

There are several different ways that this transformation might be achieved, each suited to different situations:

- Video Description
- Music
- Sound Effects
- Automated Audio Display

The following sections discuss the translation of visual content to the audio modality.

Video Descriptions

Video descriptions (also called audio descriptions or described video) are small segments of recorded speech that are added to the audio track of video presentations in order to convey important information from the visual track. The CRTC (2004) defines this transformation in two ways:

1. Audio description involves the provision of basic voice-overs of textual or graphic information displayed on the screen. A broadcaster providing audio description will, for example, not simply display sports scores on the screen, but also read them aloud so that people who are visually impaired can receive the information.

2. Video description, or described video, consists of narrative descriptions of a program's key visual elements so that people who are visually impaired are able to form a mental picture of what is occurring on the screen.

In a traditional broadcast setting, video descriptions are added "post-production" (i.e. after the original program is fully edited) and are usually confined to pauses in the audio dialogue, to prevent interference. However, in video presentations in which the dialogue is nearly continuous, this arrangement can reduce the amount of information that can be conveyed via video description. In contrast, digital video has greater flexibility and in online settings it is sometimes possible to pause the video presentation in order to provide the user with sufficient or even additional information via extended video descriptions.

For more information on "post production" video description, see Companion Document 3, Online Video Description.

Live events present a major challenge to the production of video descriptions, because the real-time nature of the presentation leaves no time for the traditional process of identifying dialogue pauses and carefully crafting descriptive utterances to insert into those pauses. Instead, the describer is thrust into a role very similar to that of a sports colour commentator, who must describe the dynamic action on a playing field in real time. However, the describer's task is further complicated by two factors: first, the need to share the audio track with any dialogue from the live event and second, the potential lack of a guaranteed frame of reference as is the case with playing field action.

For more information on "live" video description, see 3.1.1, Real-Time Online Video Description.

Music

Music is an art form that relies on the creative abilities of a composer to apply the rules and idioms of a culture for its creation. As with many of the modality translations, any attempt to express a visual element with music would be subjective to the composer's creative abilities. Even so, consumers of a musical equivalent may gain understanding of the visual element being described if they share a common culture with the composer. If the visual element itself is artistic or abstract in nature experiencing an image through another art form may be desirable, even though the transformation is a subjective one.

There is a long history of describing visuals through music in the genre of program music, which was defined in the late 1800's by Dittersdorf (Randel & Apel, 1986) and flourished in the 19th century. Program music is "music that attempts to depict nonmusical ideas, images or events" (Randel & Apel, 1986, p.656). There are three ways that composers approach the creation of program music:

First, program music can be created to be expressive in nature. The composer creates music that is intended to invoke the mood or character of a visual element. In his "Enigma Variations," Elgar attempts to capture the emotional essence of a group of people (and a dog) who are closest to him. After stating the Theme, Elgar writes a series of variations titled with the initials of the people they are to represent. Starting with his wife Elgar portrays each individual often focusing on a specific defining attribute such as limited abilities on the piano, a laugh, love, or gentrification. Elgar also uses specific events to inspire the representation of some of the people in his variations, such as a conversation, an accidental fall into the River Wye, or a long journey. Of all the approaches to transforming visual elements with music this is the most subjective.

Secondly, the depictive approach to musical transformation sees the composer use imitation to describe visual objects. In a sort of musical verisimilitude the composer imitates real world sounds, motions, or events to represent the visual element. An example of this approach is by Mussorgsky who used music to describe the exhibition of a series of paintings by Victor Hartmann in his piece "Pictures at an Exhibition." Mussorgsky represents the act of walking from picture to picture at the exhibition in a series of "Promenade" movements, which vary as the mood of the viewer is affected by the paintings. For many of the paintings Mussorgsky suggests a complete animation of the picture in the music. For instance in the movement "Bydlo" for a picture of an Ox pulling a cart Mussorgsky varies the volume of the music to depict the Ox pulling the cart from a great distance, past the listener, and away over the horizon. Mussorgsky also uses suggested actions in the picture to inspire the music. For the picture titled "Baba Yaga," Mussorgsky uses rhythmic orchestral shots in the music to suggest the tamping of the giant mortar and pestle used by the witch of that Russian fairy tale.

The narrative approach is the third possibility for a composer attempting to transform a visual element using music. Relying on the music's progression from beginning to end the composer gives the impression of continuous story, argument, or emotional evolution. A verbal or textual narrative often accompanies this approach to program music. In Prokofiev's "Peter and the Wolf" a narrator explains the method by which the music tells the story and then tells the story in sync with the music. Prokofiev uses a specific motive and instrument to represent each character in the story. For example, Prokofiev uses the string section to represent Peter, the flute for the bird, the oboe for the duck, the clarinet for the cat, the bassoon for the grandfather, three French horns for the wolf and the timpani for the hunters. Additionally, the motive or musical germ of the music changes to represent the state of the character being represented as the story progresses.

The ability of a composer to transform a visual element or event into music continues to be the subject of much debate within the artistic community. Arguing against such a possibility Stravinsky (1936) said, "music is, by its very nature, powerless to express anything at all." Whether this is true or not, the people of western culture seem to have a large lexicon of musical associations that a composer can draw on to express visual elements in musical form.

Sound Effects

The visual channel's great capacity for information transfer has made it the predominant focus in the design of both human-computer interfaces and the content they convey. Strongly neglected are non-visual, aural representations of web content. Ours ears guide us in everyday activities outside of computer use; often, the ears lead the eyes. It would benefit us to focus more on how they can aid us in better comprehending the increasing complexity of information we are exposed to while using computers.

Sound effects are *non-speech* audio (whether real-world or abstract, synthetic or recorded) that may also convey relevant information. One difficulty in applying a sound effect to an image is that sound is temporal, with no dimensions, in contrast to the two dimensional nature of the static image, which has colours, shapes, and patterns. These dimensional differences mean that the individual transforming visual content must consider the following:

- Type:
  - real-world (aural metaphor)
  - abstract
- Polarities (low to high, slow to fast) of:
  - Pitch
  - Loudness
  - Tempo
  - Repetition
- Location (stereo spatialization)

Real-world sounds tend to work well for representing objects, as the sound itself effectively prompts the user to identify or recall the associated object. Abstract sounds include beeps, whistles and buzzes made by computer applications to convey information; they may even be musical. Sound effects are effective for giving cause and effect related feedback to users, indicating the state of an element, or sonifying data about an object being represented. While abstract sounds can be used to identify objects, a user may quickly be overwhelmed in attempting to memorize a large list of objects represented with abstract sounds.

With comparatively little in the way of research findings or guidelines for effective visual to sound transformation techniques, designers have generally used experience and “common sense” to guide their work. Unfortunately, many such design choices taken under close observation have proven counterproductive for the users. Surveying listeners’ perspectives in context may serve as the best guide. Ensure validity by testing any sound design, and avoid the error of relying solely on intuition. The results of audio perception experiments have frequently yielded surprises, flying in the face of researchers’ predictions.

Walker and Kramer (1996) have shown that while listeners agree pitch *increases* in the representation of greater *value,* pitch might actually *decrease* in representing greater *size.* Visually impaired users may show a contrasting preference for the *type* of mapping. Walker (2002), asked blind and sighted subjects to choose preferred sound effects to represent increasing monetary amounts. Blind subjects opted for a reality-based model, where the increasing dollar value was represented by sounds of decreasing pitch. In this case the sound represented the sound made by stacks of money dropped on a table-top (“tap”, “plop”, “thud”) and dollars were represented concretely as being heavier in the perceptual model of the blind subjects. Sighted subjects, also expected sounds to change with values but did not expect a real mapping of the sound to dollar value; instead, abstract sounds were preferred.

Iterative evaluations throughout the development phase that gather feedback from target users is recommended for any user interface that includes sound mappings. Through recurring testing during development, one can avoid a variety of pitfalls, including the phenomenon in which a worst performing cue at outset changes, with practice, to become the preferred and most effective one (Walker & Lindsay, 2003). Also, the developer should keep aesthetics in mind because they have been shown to play a vital role (Bussemakers & de Haan, 2002). For example, sounds that are harmonically rich, yet elegantly simple and emotionally pleasing help to enrich the user experience
(Thornton, Kolb, Gemperle, & Hirsch, 2003).

Discussion, research findings and design considerations regarding durations, polarities and parameters of sound effects are further provided in Companion Document 5, Representations of Visual Geo-Spatial Information as is the topic of spatialized sound. Panning is a feature of electronic sound authored for spatialized display systems, such as stereo or surround speakers. Combined with proper equalization (the determination dominant frequencies in a sound), the inclusion of audio based three dimensional location cues (for an object) or a soundscape can aptly contribute to user comprehension of information.

While web content creators can easily have sound effects triggered with visuals, there is a specific non-proprietary graphic format we recommend for its accessibility features. Scalar Vector Graphics (SVG) is an open standard newly being adapted by conventional web browsers. Currently, a user would be required to have an SVG viewer plug-in associated with their browser to experience the graphic, and Adobe SVG viewer is the only SVG browser to support sound (*wav* and *au* formats). Support for sound had not been included in the W3C guidelines, but a user of Adobe SVG viewer can hear sound implemented in the Adobe specified way. The W3C plans to add this support in future versions of SVG.

The general procedure for adding these accessible graphics to HTML with an associated sound is as follows:

1. *Construct the SVG*
Any text editor can be used. Third-party software, such as Adobe Illustrator 10 or 11 (www.adobe.com/products/illustrator), can greatly speed up development.

Code Sample 1: An SVG that renders a blue rectangle.

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 20010904//EN"
  "http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd">
<svg xmlns="http://www.w3.org/2000/svg"
    xmlns:xlink=http://www.w3.org/1999/xlink
    xmlns:a=http://ns.adobe.com/AdobeSVGViewerExtensions/3.0 >
                <rect width="150" height="50" x="20" y="20" rx="10"
                        fill="blue" stroke="#933" stroke-width="5"/>
                              </svg>
```

2. *Add Labels and Group for Interactivity*
All interactive elements need to be grouped, which creates a group tag in the SVG document; in Illustrator, use the layers palette. Each element identifiable by screen readers needs a label, which will in turn be spoken.
Code Sample 2: Adding a title and group (shown in bold)

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 20010904//EN"
  "http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd">
<svg xmlns="http://www.w3.org/2000/svg"
    xmlns:xlink=http://www.w3.org/1999/xlink
    xmlns:a=http://ns.adobe.com/AdobeSVGViewerExtensions/3.0 >
                        <title>SVG audio example</title>
                                <g id="mybutton">
                  <rect width="150" height="50" x="20" y="20" rx="10"
                        fill="blue" stroke="#933" stroke-width="5"/>
                <text x="95" y="55" text-anchor="middle" font-size="30"
                              fill="#933">Press Me</text>
                                    </g>
                              </svg>
```

3. Add Sound Effect

Before the opening of the group tag <g> insert an audio anchor <a:audio> with the attribute xlink:href set to the file name of the sound to be played. The event handler must also be set in the audio anchor with the "begin" attribute.

Code Sample 3: Adding audio anchor (shown in bold)

```
…
<title>SVG audio example</title>
<a:audio xlink:href="button.wav" begin="mybutton.mouseover"/>
<g id="mybutton">
<rect width="150" height="50" x="20" y="20" rx="10"
fill="blue" stroke="#933" stroke-width="5"/>
<text x="95" y="55" text-anchor="middle" font-size="30"
fill="#933">Press Me</text>
</g>
…
```

4. Incorporate SVG into HTML

Use an EMBED tag to place the SVG graphic in a webpage.

Code Sample 4: HTML with embedded SVG. The previous samples could provide "rectangleWithSound.svg".

```
<HTML>
<EMBED NAME="rectangle" SRC="rectangleWithSound.svg"
WIDTH="700" HEIGHT="400">
</BODY>
</HTML>
```

The addition of tactility to SVG is discussed in 1.6.2.3 Visuals to Haptics. Further details on SVG and code samples for providing visual to audio and tactile mappings are available in Companion Document 5, Representations of Visual Geo-Spatial Information where descriptions and references are made to resources and guidelines for the optimal design of accessible geo-spatial content.

Automated auditory displays

Auditory displays are technologies that *automatically* provide auditory information (synthesized speech, sound effects, etc.) about particular non-auditory events that they are able to detect. A well known auditory display is the Geiger counter, which is a device that responds to increasing levels of environmental radiation by producing an increasingly frequent clicking sound. Auditory displays are common in specialized environments such as surgery theatres and aircraft cockpits.

Screen reader software programs (see www.utoronto.ca/atrc/reference/tech/scread.html for more information) are examples of speech-based auditory displays commonly used by people with visual disabilities, but these programs are limited to synthesized speech displays of electronic text.

Only during the last several years have more general-purpose auditory displays for visual information become available. One example of a non-speech auditory display is the "vOICe", pronounced "vee oh-I-see-ee" (http://www.seeingwithsound.com). The vOICe is an application that allows blind users to hear an aural representation of any computer graphic. The graphic might be a picture or a graph from the user's computer or it might be an image captured by video camera worn by the user and processed by portable computer equipment. Of course, any kind of real-life on-the-fly navigation with such equipment entails a steep learning curve.

The soundscape produced by the vOICe is derived through three forms of greyscale image to sound mappings:

- Left and Right are mapped to *stereo sound*. The sound pans from left to right at a default rate of one second for an entire image.
- Dark and Light are mapped to *loudness*. Brighter areas are louder and black areas are silent.
- Up and Down are mapped to *pitch*. A bright spots that are higher in an image sound higher pitched than the same bright spot lower in an image.

The application has a graphical user interface sound representation feature that transforms the region around the desktop cursor into sound, and an "Internet Sonification Browser" for web graphics that may be used in conjunction with a screen reader. The software has also been extended to web3D (virtual reality modeling language) content browsers, so that users can hear their way through a navigable, web-based 3D environment. Artists or translators should consider including the vOICe sound output with an image (see http://www.seeingwithsound.com/voice.htm for more information) or incorporating the output into the work.

1.6.2.3 Visuals to Haptics

Haptics, as an alternative to visuals, involve the use of tactile displays to convey visual content.

For most sighted people, it is hard to conceive of the world around them in any other way than by forming an internal visual construct. Maps and cartographic drawings have been used widely throughout the ages to record and convey to others the lay of the land, watery divides, and hazards unseen. Alternative means of grasping geographical information have also arisen. For instance, the Inuit have utilized bone carvings as tactile maps of coastlines. These tactile maps enable "in-the-mitt information-access" for kayakers - a method more practical in inclement and low vision conditions.

A spread in use of affordable tactile devices and correlate software for encoding tactile information in web content (as discussed in 1.5.3.1 Haptic Devices) has made interactivity on the Web a more sensual, yet unrefined reality. Some researchers have observed that touch is the "reality sense", in that objects touched are discerned as more real that what is seen or heard (Taylor, Lederman, & Gibson, 1973). The ability to feel the face of information might be considered a valuable inclusion to a user's future experience of surfing the Web.

Research literature on haptics generally distinguish between two types of tactile representation: *object shape* and *texture.* Stephen Wall, in his University of Edinburgh thesis "An investigation of temporal and spatial limitations of haptic interfaces" (www.dcs.gla.ac.uk/~steven), refers to a "haptic spectrum" where *geometric primitives* making up the overall shape of an object can best be described as low frequency, high amplitude signals, available as kinesthetic cues. Such cues would necessitate a haptic device with force feedback. Conversely, *surface properties* of objects, such as textures, can be expressed as high frequency, low amplitude signals, more suited as cutaneous cues, including vibration.

General acceptance and use of specialized control and display devices have a strong correlation to cost, and affordable haptic devices lack the sophisticated mechanisms providing robust force feedback. While SensAble Technologies (www.sensable.com) currently provides the PHANTOM (Personal Haptic iNTerface Mechanism) with high fidelity force feedback at a cost of over ten thousand dollars, an affordable variant has been promised in the near future. In the meantime, Immersion Corporation (www.immersion.com), formed in 1993 to bring haptics out of the research lab and into the market with over 200 patents in tactile feedback, has an array of devices on offer. Like Sensable, the higher-end is only accessible, dollar wise, to the research labs and corporate realm. Immersion's affordable force feedback devices are no longer commercially available, leaving only basic tactile feedback trackballs and mice (currently including the HP Force Feedback Web Mouse, Kensington Orbit 3D, Belkin Nostromo n30, Logitech iFeel Mouse, and Saitek Touch Force Optical Mouse) that incorporate TouchSense Technology:

> The term **tactile feedback** is used to describe devices that play high-fidelity tactile

sensations, but generally won't move or inhibit the movement of either the device or the hand holding it. A tactile feedback device can play a wide variety of distinguishable taps, textures, and vibration effects to communicate with the user and greatly enrich the computing or gaming experience (www.immersion.com/developer/technology/devices).

Immersion Technology’s Web Designer and Studio are TouchSense applications (www.immersion.com/developer/technology/tools) available to web content designers for a fee. They allow for creation and editing of physical effect parameters (such as duration, gain, magnitude, phase, waveform) and deployment of Immersion Force Resource (IFR) effects in desktop applications (using programming languages which access ActiveX controls, such as Visual Basic or Director’s Lingo) and web contexts (using Javascript functions). An introduction to fundamentals document is available at www.immersion.com/developer/downloads/ImmFundamentals/HTML/ImmFundamentals.htm

Our recommendation for associating haptic effects (using Immersion TouchSense) with web-based visual information entails the non-proprietary graphic format Scalar Vector Graphics (SVG). The user would be required to have an SVG viewer plug-in associated with their browser, such as the Adobe SVG Viewer (www.adobe.com/svg/viewer/install), and the Immersion Web Plug-in (www.immersion.com/plugins). The general procedure[2] for adding accessible graphics to HTML (in this context, a map image having both audio and tactility using Immersion-licensed TouchSense devices) is as follows:

1. *Construct the SVG*
   Any text editor can be used. Third-party software, such as Adobe Illustrator 10 or 11, can greatly speed up development.

2. *Organize Objects in Layers for Interactivity*
   Group all elements in appropriate layers and label both elements and layers. For example, in a map setting, put symbols for train station, bus station, and airport into one group, primary, secondary and tertiary roads into a second group, mountains and bodies of water forming two more groups. All interactive elements need to be grouped, which creates a group tag in the SVG document; in Illustrator, use the layers palette. Each element identifiable by screen readers needs a label, which will in turn be spoken.

3. Assign Haptic and Sound Effects
   AssigTn JavaScript functions to layers. Associate each haptic effect with a sound effect, which makes map features more evident to people with visual impairment. Each category of map elements gets assigned a separate haptic effect with javascript, e.g. one for roads, one for built-up areas, one for water and occasionally

[2] Steps two and three are adapted from *SVG Maps for People with Visual Impairment* (www.svgopen.org/2003/papers/svgmappingforpeoplewithvisualimpairments)

one for borders. Javascript functions to detect user input and trigger the events; in Illustrator, use the interactivity palette. Otherwise, for those versed in javascript, code can be added to the SVG with a simple text editor.

4. *Incorporate SVG and haptic effects into HTML*
Use an EMBED tag to place the SVG graphic in a webpage. An object reference to the Immersion Web Plug-in must be made, along with javascript referring to the tactile effects.

5. *Add both Immersion Web plug-in and user environment detection to HTML*
Use an OBJECT tag and javascript to detect presence of the Immersion plug-in associated with users' web browser. If non-existent, user is referred to plug-in site. (Code supplied in 8.7)

Further details on SVG and code samples for providing visual to tactile mappings are discussed in *8.7 Representations of Visual Geo-Spatial Information*, where descriptions and references are made to resources and guidelines for the optimal design of accessible geo-spatial content. Aside from SVG for 2D content, the *Representations of Visual Geo-Spatial Information* delves into Web3D (Virtual Reality Modeling Language (VRML)/Extensible3D (X3D)) for web-based 3D content. *Modality complimentarity*, or how signals in varying sensory modalities reinforce or bootstrap each other to convey a message, is stressed in anticipation of the next generation of multimodal systems.

1.6.3 Alternative Modality Equivalents of Audio Content

The following sections provide an in depth discussion of the translation of audio content (speech, music and sound effects) to language, visual and haptic modes.

1.6.3.1 Audio to Language

Language as an alternative to audio involves the use of words to describe or label sound information (voice, music, sound effects)

There are several different ways that this transformation might be achieved, each suited to different situations:

- Transcription of dialogue (Captions)
- Description of Music
- Description of Sound Effects

These translations from audio to text are discussed further in the following sections.

Transcription of dialogue (Captions)
Audio dialogue is typically accompanied by a video although it may also appear on its own as in a web radio broadcast. When accompanied by a video, dialogue is usually represented in visual form as captions. Other options for text description of dialogue is full text

transcripts. However, spoken words are more than just the words alone; other qualities such as tone, pitch, rhythm, hesitation and even silences add a great deal of meaning to dialogue.

For this reason, dialogue transcribers and captioners are encouraged to consider how these paralinguistic aspects of language may be conveyed. One option for captioners that may have applicable techniques for transcribers is to create enhanced captions. Enhanced captions are text captions that include other information to convey paralinguistics. For example, a caption that is enhanced might have a word appear in a different font style, size or colour to create emphasis or convey mood. Other options are inclusion of images in an enhanced caption to convey paralinguistic information that is not otherwise available. These options are discussed further in 1.6.3.2 Audio to Visuals and in detail in Companion Document 2 Online Enhanced Captioning.

Description of Music

"Music is universal, but its meaning is not." This simple truth posses a great challenge to the person who takes on the task of describing music. A medium, such as music, that is totally ubiquitous among all cultures has great potential to be the ultimate tool of translation. However, due to its abstract nature meaning in music and even its significance vary greatly from culture to culture and from person to person. While music is an abstract medium, it is always created in a cultural context. It is this cultural context that can be used to create a complete and useful description of a piece of music.

Using text to describe a piece of music creates an additional challenge because the experiences of listening to music and reading text are very different. A listener of music's first reaction is visceral or emotive; the reader of text's first reaction is intellectual. This perceptual difference makes the effective textual description of music difficult.

One could go about describing physical characteristics of a piece of music, such as texture, timbre, or tempo. While this information may be of use to a performer or composer of music, a listener does not understand a piece of music according to its physical characteristics. The describer could also approach the task of describing music from the more emotive point of view. This approach is closer to the way a listener perceives music, but the emotion or mood that a piece of music invoke is culturally subjective and may even differ from individual to individual within a culture.

The describer of music must identify these culturally subjective elements in order to communicate the music-culture of a piece to the reader. Music-culture has four components that describe how people interact with music. These components can be used build a description of a piece of music that adequately communicates the nature of the described music. The four components of a music-culture are:

1) Ideas about music (why)
2) Social organization (who)
3) Repertories (what)
4) Material culture (how)

By organizing a description of a piece of music according to the components of music-culture the describer will not only give a complete description, but also a description that is easily sorted into a layered system. A layered system will allow the reader to get a description of the aspect of the music relevant to them.

The way a culture thinks about music has a great impact on the meaning that a piece of music carries. Different cultures even differ on what music is. For instance the Kaluli of Papua New Guinea believe that bird sounds are music and the music they make represents feelings of sadness. In some cultures music has exclusive religious meaning, while in other cultures there is not even a word for music. In western culture it is generally thought that music is "humanly organized sound" (Blacking, 1973). Individuals within a culture may also have differing ideas about what music is. While one person understands music as the product of popular icons, another may have primarily experienced music in a religious context, and yet another may associate music with elevator rides on the way to work. Considering these differences, the describer of a piece of music will want to include information about the culture the piece of music was written for and in what context it was to be performed. In addition to explaining the ideas that influenced the creation of the piece, the writer will want to explain the context in which the music is being used (assuming the description is of a piece of music that exists in a larger context and the cultural context that it is embedded in differs from that in which it was created.)

Who wrote the piece of piece of music and their social position has a surprising effect on the listener. For instance a television spot for world aid uses John Lennon's "Happy Christmas (War is Over)." The fact that John Lennon was one of the dominant pop icons of the 1970's, that he was very involved in the anti war effort in the later part of his career and that he was assassinated all have an effect on the listener who was a part of that culture. Cat Stevens' conversion to Islam was very relevant to the meaning in much of his music. This event becomes especially potent when one considers attitudes toward music in strict Muslim sects, where music is considered an earthly distraction. While John Lennon and Cat Stevens had the status of popular icons in western culture musicians in some other world-cultures have a low social status, and in other cultures music is a strictly communal event and the identity of a composer is not acknowledged.

Repertories:
This part of music-culture is about the music itself as opposed to cultural definitions of music, or who wrote the music. Every culture has a group of musical performances that we call a repertory. These performances are identifiable by the members of the culture as being their own, and share a set of defining attributes. This is the most obvious layer of a piece of music's music-culture and is a natural place to begin describing a piece of music. There are six types of attributes associated with a music's repertory. They are:

1) Style
2) Genres
3) Texts
4) Composition
5) Transmission
6) Activity

The stylistic attributes of a piece of music include pitch elements (mode, melody, harmony, tuning systems), time elements (rhythms, meter), timbre elements (voice quality, instrumental tone colour), and sound intensity (volume) (Titon, 1996). As noted earlier, a listener does not observe these aspects of a piece of music in detail, so generalizations about stylistic elements are desirable (unless the description is for composers or performers.)

A culture's repertory of music is divided into units called genres. The genre of a piece of music can act as very valuable information to the reader of a description as it contains much about the context in which the piece was written (lullaby, hymn, rock and roll, jazz, dance, etc.)

The text (or words) of a song is the easiest information to communicate to the reader about a piece of music as a modality translation already exists in the form of written text. Similarly, the actual musical notation for the piece may be a useful translation to visual for individuals with musical training.

The method by which a piece of music enters the repertory may also have bearing on a complete description of a piece of music. For instance, if the piece of music belongs to a classical tradition the music will follow a strict set of rules and will have been created by a person trained in the art form the music represents. On the other hand, if a piece of music hails from a folk tradition it will follow a much simpler set of rules and may evolve and change from performance to performance. Improvisation may also be a part of a pieces composition. In this case the performer becomes involved in the creation of the music and depending on the genre may have an additional set of rules governing the improvisation.

The transmission a piece of music has describes the method by which a performer knows how to perform the music, whether by oral tradition or by a notation system. While this aspect of a piece of music is primarily of interest to composers and performers a notation system is a ready-made translation of a piece of music that can be incorporated in the description to give the reader an idea of the shape of the melody, character of the harmony, and the number of sounds involved. It should be noted that western style music notation was designed to give performance instructions to performers with maximum efficiency, and therefore does a poor job of depicting the actual sound in a visual manner. "Piano Scroll" notation is an example of system of music notation that is intended to represent sound.

The performance of music in almost invariably accompanied by actions on the part of the people involved in the performance (both audience and performers, assuming such a distinction exists). Some questions a musical description will answer with respect to this topic are: Is the performer animated in some way? Is the audience dancing? Does the performer engage in the ritual destruction of his/her instrument? How are the audience and performs dressed?

The final component of music-culture that a description of a piece of music are the materials or tools that are used in the performance of the music. No one alive today has actually heard any performance before 1870. However, using information about the tools used in those performances we are able to recreate the music that was made in times before recording technology was invented.

Description of Sound Effects

Sound effects refer to audio that is neither speech nor music. Sound effects usually serve one of three functions: 1) draw attention or emphasize, 2) provide ambience and 3) convey information over time.

An attention-drawing sound effect is most often used to provide confirmation of some action a user takes (reinforcing cause and effect) or to emphasize an event. One characteristic of these types of sound effects is that they are short, being designed to deliver their effect with maximum efficiency. Generally, these kinds of sound effects have two forms: those that imitate real world sounds or those that are abstract in nature.

In the case of a real world sound effect it is important for the description to reference what real world thing is being imitated, include the duration of the effect and the intention of the effect.

Sound effects that are abstract include all sound effects that do not refer to any real world object and who may have music attributes. Naturally these sound effects are more difficult to describe. However, the description should include the physical characteristics of the sound effect such as timbre, texture, rhythm and pitch. Real world sound effects often suggest their intended effect in the thing it imitates. Abstract sound effects do not share this feature, so it is very important that the intention of the sound effect also be fully described.

Ambience sound effects provide ambience for whatever context the user is in. Like other sound effects, ambient sound effects can be real world in nature (the sound of a busy train station or machine noises) or be abstract (pulsing hum.) Typically ambient sound effects are longer in duration. When describing ambient sound effects it is important to include the intended atmospheric effect. While describing the physical characteristics of this type of sound effect include how constant the effect is answering these questions: what is the total duration? Does the effect change over time? Does interaction by the user cause the sound effect to change?

A third category of sound effects includes those that are intended to convey information over time. A description of this type of sound effect must fist describe what information is being conveyed and the nature of the information being represented. This sort of sound effect may be used to inform the user of the changing state of a progress bar or the sonification of a dynamic data set, such as the number of active process a computer has running. A description of such an effect must address whether the information is progressive, random, interrupted or continuous. The description of an informative sound effect must then include how the physical characteristics of the sound change with the information.

1.6.3.2 Audio to Visuals

Visuals as an alternative to audio involves the use of graphics or animation to transform audio (voice, music, sound effects) into visual information.

"Visual for audio" covers the following:

- ASL Translation
- Affective component of enhanced captions
- Visual displays of music or sound effects

The following sections discuss these translations from the audio modality to the visual modality further.

ASL Translation of Speech

While captions are a good visual representation of speech, a better alternative for many deaf individuals is a sign language interpretation. Extensive guidelines on provision of ASL interpretation of audio may be read in Companion Document 4 Remote Real-Time ASL Interpretation.

Graphical Component of Enhanced Captions

Captions are effective in conveying the words of the speech component of audio. Other aspects of the audio, however, are often lost in standard captions. For example, tone and pitch of the voice are lost in the caption process. Also lost are non-speech elements such as music, environment sounds and even noticeable silences or pauses. Enhanced captions may provide a way to include these important audio elements in this modality translation. Enhancements may be simple such as changes to the style of the caption such as font type, size or colour. Other enhancements may be emoticons or graphics that indicate mood or kind of music. An extended discussion of enhanced captioning is available in Companion Document 2 Online Enhanced Captioning.

Visual Displays of Music or Sound Effects

It is very important to include alternative modes of crucial information such as interface warnings or alerts. For example, a sound that confirms an event such as sending a file should have a corresponding visual confirmation. An artist may incorporate sound into a new media piece such that sound occurs when the cursor moves over a certain area of an image. A viewer without benefit of speakers or unable to hear due to the environment or physical factors would not be able to participate fully in the work. However, if the artist includes other modes of access such as a visual translation of the music then the viewer who cannot hear is better able to participate fully in the work. Visual representations could be static or animated representation and may be speech or abstract. Defining optimal ways to convey non-speech audio through visual means is an area open to exploration and experimentation. This avenue of exploration may provide exciting and interesting challenges for artists interested in expanding their audience as well as their audience's experience of their work.

1.6.3.3 Audio to Haptics

Haptics, as an alternative to audio, involves the use of tactile displays or controls to convey audio information.

Tactile output is not unusual in the form of dedicated device prompts, as in the case of a vibrating pager. While affordable tactile devices (see 1.5.3.1 Haptic Devices) have increased the use of tactile prompting on personal computers, their current software and drivers neglect the strong potential for tactile alternatives to audio alarms ("You have mail!"), particularly when applications visually obscure one another. Very little research or recommendation exists for optimum haptic representation of sound. However, software developer kits for haptic devices do allow to those with patience and programming capability to tap into this largely unrealized potential.

Some experimentation in musical expression has led to the development of specialized devices and even suits, such as that developed by Eric Gunther at MIT. Gunther's haptic suit "facilitates the perception of musically-structured spatio-temporal patterns of tactile vibration on the body's surface, with concurrent patterns of audible vibrations as musical accompaniment" (Purcell, 2002, p.36). Media Lab Europe's Palpable Machines Research Group (www.mle.ie/~ian/palpable) is prominent in its pursuit of audio haptics applied to the creative realms of story telling, music, dance and broadcast media.

In our everyday experiences, tactile sensations are often accompanied by sound; think of how a subway platform resonates in conjunction with an approaching, unseen train. In sensory modality transformation research, the connection between sound and touch have received strong attention.

1.6.4 Accessible Collaborative Tools

While many of the above solutions consider accessibility from the view point of participating in available online culture, it is also critical that the collaborative process of building culture also be accessible. Following are descriptions of two tools that enable individuals to collaborate over the web via completely accessible interfaces developed by the Adaptive Technology Resource Centre: A-Chat and A-Communicator.

1.6.4.1 A-Chat: Accessible Online Chat Tool

Available at: http://achat.atrc.utoronto.ca, A-Chat is a basic browser-based chat program designed to accommodate users with differing access needs. This software serves as a research platform to enable investigation of ways in which common chat features might be made more accessible.

Key Features

- Implemented in PERL.
- Administrator Settings include:
    - Language
    - Transcript

    - Message/Chat/User ID lifespans
    - Public or private user list
    - Hands-up mode in which a moderator controls the "floor"
- User can login to an existing account or automatically start a new one.
- Accessibility related preference options are provided:
    - Option to play sound effect when a new message arrives.
    - Various refresh options, including manual
    - Message ordering
    - Show only new messages
    - Font size, font face
    - Colour scheme
    - Navigation aids on or off
    - Language settings
- Participant list with names linked to participant message history.
- Screen reader friendly page layout.

Screenshots:
The following figures illustrate how individual preferences are set in A-Chat to allow customization of the appearance of the interface and the content.

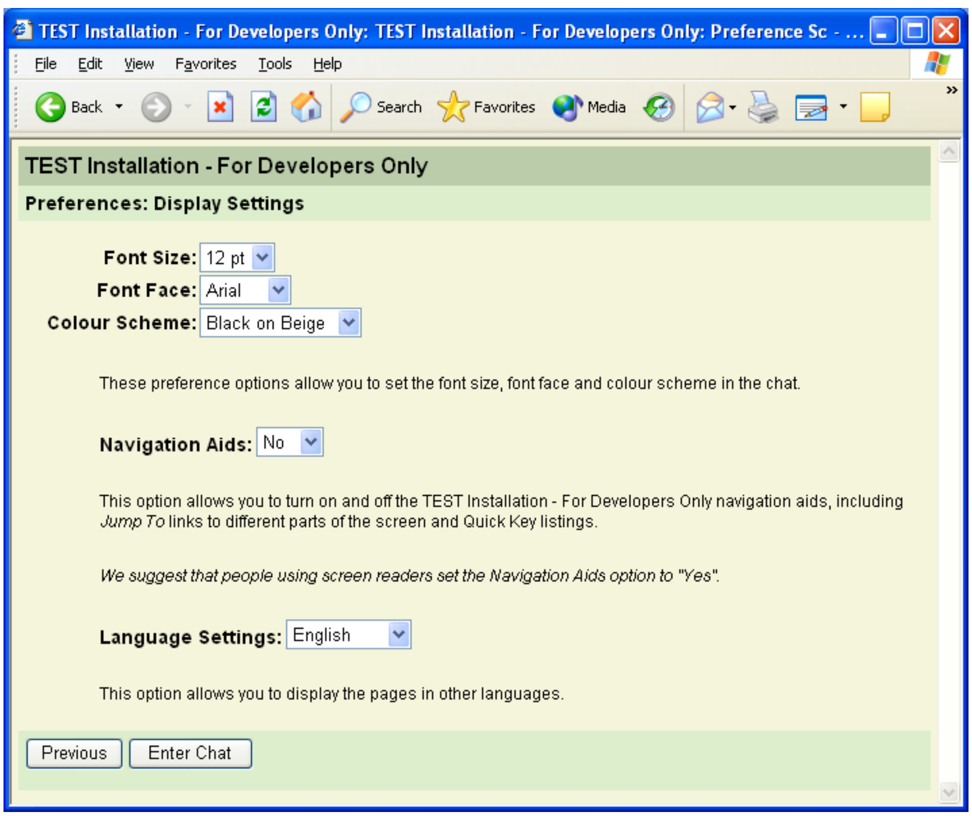


Figure 3: This screenshot shows one of the A-Chat user preference screens.

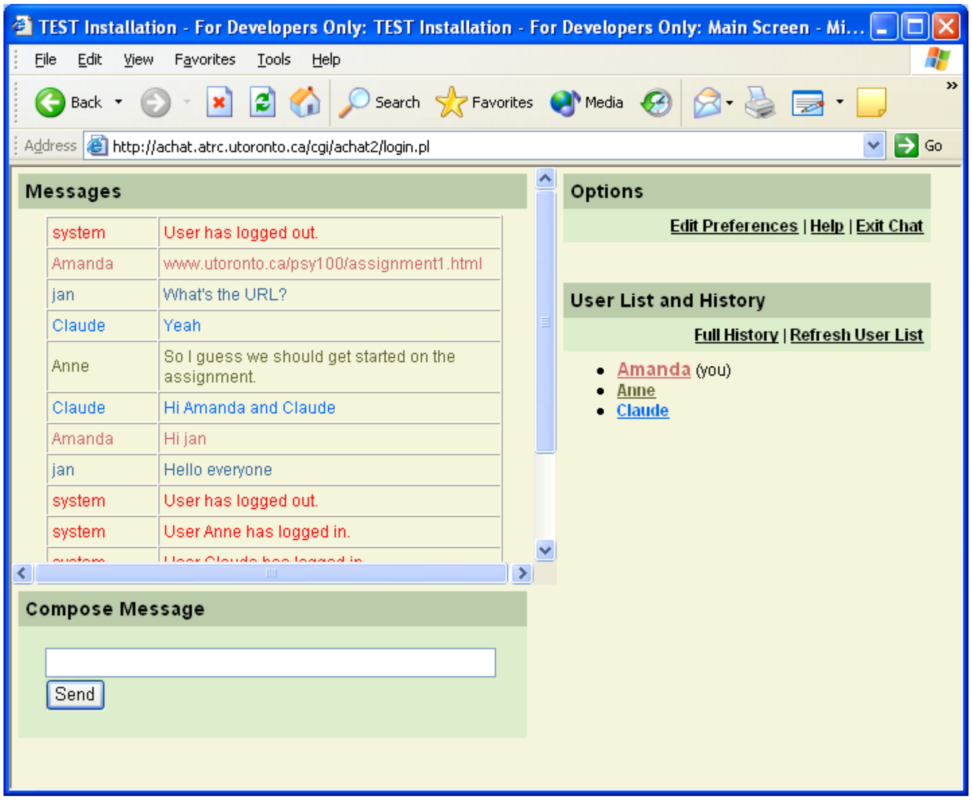


Figure 4:This screenshot shows the main A-Chat screen. The message area in on the top-left, the compose message area under it and the ootions and user list and history areas on the right.

Suggested Use of A-Chat:
A-Chat can be used to facilitate any number of text-based collaborative activities, including lectures, story-telling, tutorials, discussion groups, and other meetings. Administrative options allow the chat session to be customized according to the level of privacy and user participation desired.

The tool does not currently implement any type of warning system to let the other participants know when someone is still composing a message, so users of the system should remember that some users will take longer to compose and send a message than others. There are several methods for minimizing the chance of conversations changing topic before a person can contribute their message. These include:

- allowing the use of abbreviations
- leaving a reasonable period or getting participant assent before closing a subject
- encouraging users to split long messages into two, with any preliminary messages ending with "…" to make it clear than more is on the way.
- controlling the "floor" either by making use of some informal arrangement or using the more formal (and restrictive) A-Chat hand-raise mode.

A-Chat is capable of producing text transcripts (see administrative options) in the form of HTML documents. Some groups may find it useful to post these files following a session in order to facilitate understanding by users who may not have been able to follow all that was said.

1.6.4.2 A-Comm: Client-Side Chat/Whiteboard Tool

Available at: http://acomm.atrc.utoronto.ca, A-Communicator (A-Comm) is a basic client-side Instant Messaging (IM) and Whiteboard tool developed as a research platform for investigating peer description and keyboard accessible drawing. A-Comm utilizes an object-oriented drawing environment, implemented with SVG. This software serves as a research platform to enable investigation of some design features that might ameliorate the accessibility challenges posed by the display and collaborative control of graphics in whiteboards.

Key Features:

- User can login to an existing Jabber account or sign up for a new one.
- Roster provides "presence" information for contacts.
- Users can participate in private or group chat/whiteboard sessions.
- Text chat window and optional whiteboard
- Participant list on demand.
- A variety of whiteboard drawing tools are available
    - Pencil
    - Line tool
    - Rectangle tool
    - Ellipse tool
    - Line colour, fill colour, line thickness
- Peer description allows any participant to describe the drawn objects. Anti-collision feature ensures one describer per object.
- Some accessibility related options are provided:
    - Option to play sound effect when a new message arrives.
    - Keyboard drawing checkbox
    - Manual refresh checkbox
    - Show only new checkbox
    - Newest message first checkbox
- Keyboard enabled drawing lets users use some features of the whiteboard without a mouse.
- Save and open palette feature lets users create and describe pictures beforehand.

Screenshot:

The following figure illustrates peer description of images in the collaborative whiteboard space.

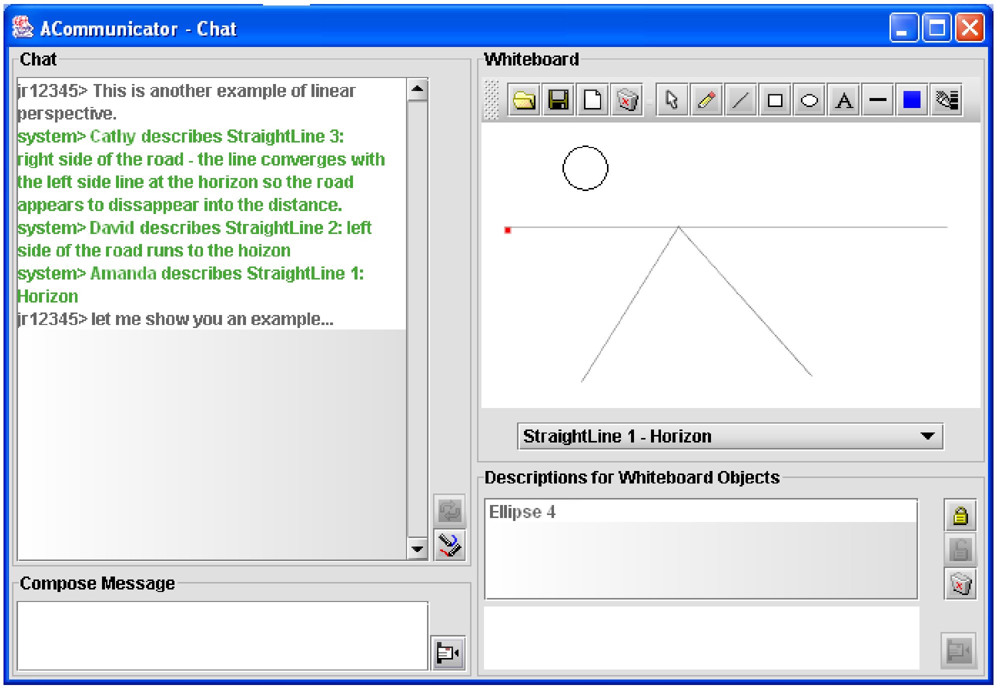


Figure 5: An example of a teacher-led lesson using drawings to demonstrate a concept, three students ("Amanda", "David" and "Cathy") submit peer descriptions. The most recently drawn shape, an ellipse has not yet been described.

Suggested Description Strategies:

The peer description tool allows descriptions to be added to as many or as few drawing objects as describers wish. Describers may decide to describe every drawing object on the screen or they may choose to ignore most of the objects and focus on adding higher level descriptions of what is being drawn to just a few choice drawing objects (in the future, object grouping will make this distinction less important).

The benefit of describing every object is that users with visual disabilities will have an in depth picture of the different objects being added to the whiteboard. However, the drawback is that the chat may become cluttered with very repetitive descriptions of low level details (e.g. describing every tiger stripe, etc.), which prevents the description reader from "seeing the forest for the trees". On the other hand, placing high level descriptions on just a few objects gives the user a better overview, but risks problems if the objects holding the high level description are deleted or re-purposed as part of a different drawing.

The best guidance is probably to use a combination of the two approaches. Use low level descriptions, sparingly, to get across to the description reader the process used to create the picture, but avoid overly repetitive description of repeating features. At the same time remember to use the description of some of the objects for higher level descriptions of the composition as a whole.

In all cases, be careful with your describing. A-Comm does not currently have a method for revising descriptions once they are submitted.

Using Pre-Authored Drawing Objects:
A-Comm allows users to create and re-load batches of pre-authored (and pre-described) drawing objects in the form of SVG files. Additional pre-authored SVG files are also available from the A-Comm website. Note: A-Comm does not support all features of SVG so do not attempt to load SVG files created by other tools.

This feature can be used by anyone on the chat and causes the drawing objects to be placed in addition to any objects already on the whiteboard. Any descriptions of objects in the SVG file will be displayed in the chat. Undescribed drawing objects will be available for peer description. There is also nothing to prevent this feature from being used by people with visual disabilities.

This feature may be useful for the following:

- “Slide shows” or other planned presentations
- Standardized drawings (e.g. flow charts, organizational charts, etc.)
- Drawings that include frequently used objects (e.g. company logos, arrows, etc.)

For more information on description writing see 1.6.2.1, Visuals to Language, Companion Document 2, Online Enhanced Captioning and Companion Document 3, Online Video Description.

1.6.5 Ensuring Accessible Control

Even systems that are keyboard operable should be designed to take into account the accessibility of any point-and-click interface components, because some users will prefer to continue using a mouse their having reduced accuracy. A variety of adapted mouse devices exist to facilitate control, including joysticks, trackballs, eye gaze and head tracking systems. Some of these devices include separate switches or allow selections to be made by dwelling on a location.

While WCAG does not include any explicit requirements related to mouse adaptation, developers can better accommodate the users of these devices, and improve the general accessibility of their interfaces, by avoiding very small clickable controls and by providing some space between controls to reduce the probability of error.

1.6.5.1 Keyboard Accessibility

The most important factor determining the accessibility of the controls in a given piece of on-line content is the keyboard operability of those controls. This term refers to the degree to which a user can operate the interface of the content using only the keyboard, instead of the mouse or other pointing device. Operating the interface involves more than just being able to press all the same buttons with the keyboard as with the mouse. It also includes ensuring that the user can move between controls, select controls without necessarily activating them, and in the case of more complex controls such as those that incorporate drag-and-drop, ensuring that all the same functionality is available by alternate keyboard operable methods. Keyboard operability should be enhanced by considering the number of keystrokes required to perform an action. Generally, the number of keystrokes required to perform the most popular actions should be minimized. This minimization can be achieved by providing shortcut keys as well as a predictable TAB order that places the most popular actions early in the order.

Keyboard operability is important because in the process of implementing it, content is freed from the visual modality of point-and-click. Once the control actions are modality independent, they can be controlled by devices as different as conventional keyboards, alternative keyboards, onscreen keyboards, and voice recognition systems.

Obviously, certain highly visual/spatial activities, such as drawing, lend themselves strongly to operation with a mouse and present special challenges when designing keyboard accessible systems. However, it is possible to implement a rudimentary drawing facility, parallel to a mouse-driven system, that uses keyboard commands that enable drawing objects to be added, moved, rotated, and resized.

For more information on keyboard accessibility, see checkpoints 9.2, 9.3, 9.4 and 9.5 in WCAG 1.0. Links to techniques are included.

1.6.5.2 Mouse Adaptation

Even systems that are keyboard operable should be designed to take into account the accessibility of any point-and-click interface components, because some users will prefer to continue using a mouse their having reduced accuracy. A variety of adapted mouse devices exist to facilitate control, including joysticks, trackballs, eye gaze and head tracking systems. Some of these devices include separate switches or allow selections to be made by dwelling on a location.

While WCAG does not include any explicit requirements related to mouse adaptation, developers can better accommodate the users of these devices, and improve the general accessibility of their interfaces, by avoiding very small clickable controls and by providing some space between controls to reduce the probability of error.

1.6.5.3 Voice Recognition

Voice recognition input systems allow a user's voice to be used as an input device. Voice recognition may be used to dictate text or give commands, such as opening application programs, pulling down menus, or saving work. Modern systems allow a user to dictate text fluently into the computer at rates of up to 160 words per minute. However, while these systems do give the user a fairly high degree of control, they are not yet completely hands-free.

For online cultural content developers considering implementing voice recognition systems, even simplified ones, it is also important to keep in mind that some people are unable to speak or to speak clearly to their computer due to disability or circumstance (e.g. no microphone, poor quality audio line, etc.). Therefore, it is recommended that systems that can be controlled by voice recognition also be operable via the keyboard (see Keyboard Accessibility,1.6.5.1).

1.6.5.4 Gestural Interfaces

Gestural interfaces allow the user to control a system with movements of their body. This input is made possible by the deployment of cameras and special recognition software. Different systems can be setup to detect different gestures, such as full body movements, movements of the hands, or movements of the face only. Research and development is underway on more advanced systems to recognize sign language.

As was the case with voice recognition, online cultural content developers considering implementing gestural interfaces systems should keep in mind that some people are unable to gesture or gesture clearly to their computer due to disability or circumstance. Therefore, it is recommended that systems that can be controlled by gestural interface also be operable via the keyboard (see Keyboard Accessibility,1.6.5.1).

1.7 Conclusion

These documents outline possible techniques and considerations for creating inclusive online cultural content. The discussion began with an outline of existing accessibility principles for online content and then move towards an exploration of benefits of accessible content. We have seen how accessible content provides not only a wider audience to the artist but, also provides the audience with multiple ways to access, experience and understand cultural content. The key route for creating accessible content is to consider multiple modalities of presentation since barriers to access are chiefly inability to access the given content mode.

Within the broadcast medium, standards for video specific modality translations exist. These translation are closed captions and described video. These translations, however, are limited and do not fully express the original content. The digital medium provides room for expansion of these translations to enable greater expression and audience control. Some exploration and discussion of enhanced captioning and video description are found in this document and in detail in the following companion documents. Furthermore, there is room for translations that do not necessarily use prose form or language. Sound, motion, images and tactile sensations all have the potential to convey meaning. Artists skilled in multiple mediums may find the suggestion to create a work across multiple modalities an interesting challenge that stretches their creativity and expressive skills. Other artists may need to collaborate with others to achieve inclusive content. Collaborations should enhance the expressive ability of the artist rather than stifle it. The concept of modality translation has been opened up for further discussion and exploration. By no means have the optimum ways to translate existing art or to convey meaning in new pieces been determined. The artist will be able to help lead expressive innovation by working towards inclusive work either alone, in collaboration with other artists or in consultation with a diverse audience. Modality translation in the digital medium presents an exciting opportunity for artists, developers and curators alike to achieve new levels of expression and inclusion.

Certainly, there is a need to recognize that ability is a fluid state. When considering cultural content, accessibility may have as much to do with prior knowledge of the individual as with level of sensory perception. Furthermore, as individuals age, they often experience some decline in sight, hearing and mobility. Thus cultural content that is initially accessible to one individual may through chance or aging become inaccessible. Multiple modes of expression provide opportunity for all individuals regardless of ability to approach cultural content in several ways. These modes will allow individuals with different learning preferences, different abilities and different backgrounds to experience the same content through different modes. The artist will have achieved a wider audience and potentially will have greater opportunity to have an impact on that audience.

Next steps in this movement towards multi-modal cultural content will be to further explore concepts of modality translation. Suggestions and guidelines for practice in this document and the companion documents should be implemented and outlined further. These documents are the beginning of the discussion and we need to move towards models of practice that will meet the needs of a diverse and expanded view of the audience.

Acknowledgements

Canadian Network for Inclusive Cultural Exchange (CNICE)

Creating Accessible Online Cultural Content Discussion Document Series

CNICE Partners
Adaptive Technology Resource Centre
Banff New Media Institute
Canadian Abilities Foundation
Canadian Centre on Disability Studies
Canadian Cultural Society of the Deaf
Canadian Hearing Society
Canadian Learning Television
Centre for Learning Technology
Centre for Research in Disability Studies
Digital Frog International
marblemedia
National Broadcast Reading Service

Editors and Contributors:
Michael Boyce, Sara Diamond, Deborah Fels, Elle Gadsby, Rob Harvie, Wendy Porch, Lake Porter, Jan Richards, Vera Roberts, Marc Rosen, Nancy Sicchia, Charles Silverman, Jutta Treviranus

This project was made possible through the funding by the New Media Research Networks Fund of the Canadian Heritage Canadian Culture Online Programme.

Canadian Heritage Patrimoine canadien

2 Online Enhanced Captioning

2.1 The Current State of Captioning and Descriptive Video on the Web

Video or audio streams and files, available over the web as part of many rich media presentations, are rarely presented in alternative formats. While the broadcast industry has addressed both captioning for people who are Deaf or Hard of Hearing and video description, which is the spoken description of visual content for people who are blind, the web's "audioscape" remains largely inaccessible.

While it is technically feasible, and some guidelines have been generated over the last several years, there are few current examples of captioning or descriptive video service (DVS) over the Web. Most of the available content is limited to simple captioning of a few sample content pieces, is usually offered by groups that focus on disability issues. A major reason for this dearth of captioned and DVS content is the lack of legislative requirements for these services. This is in sharp contrast to legislative requirements for broadcast captioning, and lately, start-up requirements for DVS.

MarbleMedia is one of the few new media companies to provide captioning of streaming video (www.deafplanet.com). The following example is from MarbleMedia's web site where they are currently providing short captioned clips from their TVO production, DeafPlanet. Deaf children make up a fair portion of the audience for a show that is geared to Deaf and Hearing children and their families.

Figure 6: Example of captioned streamed video.

For streamed video that isn't directed towards a disability group, the one mainstream exception regarding Web-based captioning is PBS. A number of its shows are provided on the Web with options for closed captioning. The following example from PBS (www.pbs.org) is just one of many available streaming videos that provide captioning.

Figure 7: Sample of captioned PBS video.

In each of the examples, captioning can be toggled off and on by selecting the appropriate button next to the video. Providing the caption controls next to the video itself gives the user a clear indication that captioning is available.

One of the only, and by default, best examples of rich media to make accessible culture is from the WGBH/NCAM (http://broadband.wgbh.org/quicktime/mpeg4/). In addition to captioning, in both English and Spanish, annotations, called "enhancements" here, are offered in different sections of the video stream.

Figure 8: Example of annotations with captions.

Other than a few instances of demonstrations that show examples of DVS, we have not found any instances of DVS on the Web.

2.2 There's No One Standard for Captioning and Description on the Web

There are a variety of media players available for the web, as well as a number of different approaches for creating web pages. As a result, there is no one approach for providing captioning and description over the web. Captions can be transmitted directly onto a Web page using conventional web tools, such as HTML, Javascript, and CGI, through a particular media tool such as Flash, one of the three major media players (Quicktime, Real, and Windows Media), or even by means of a chat service. (See the NCAM site at http://ncam.wgbh.org/richmedia/ for an exhaustive list of approaches and techniques for creating accessible rich media).

2.3 Style Guides for Captioning and Describing On the Web

While WGBH/NCAM and others offer a recommended guide to caption styles (http://main.wgbh.org/wgbh/pages/mag/services/captioning/faq/sugg-styles-conv-faq.html). Of particular value is a discussion on sentence breaks for making readable captions.

A similar online guide for video description is available as well from WGBH/NCAM (http://main.wgbh.org/wgbh/pages/mag/resources/archive/dvs_guide_06-24-2003.html).

These documents were created for broadcast and as a result are shaped by the technical limitations of the broadcast medium

2.4 Tools for Captioning and Describing on the Web

While a number of tools exist for broadcast captioning and description, only a few tools directly support adding captions or description to media for use on the Web. Video editors and software-based audio recording studios provide ways to add additional tracks that can be used to generate captions and description,, but these tools not well designed for such tasks, and as a result adding captioning and described video is far more difficult and time-consuming than necessary.

The most robust of the available tools currently on the market, is Magpie, which was developed by WGBH/NCAM (http://ncam.wgbh.org/Webaccess/magpie/#v2). Magpie lets authors produce captions and descriptions for different Web formats and media players. Magpie allows authors to add SMIL, SAMI, or QuickTime style captions and description.

The CNICE project has developed a tool for captioning and description that has let us begin to explore and demonstrate a number of "enhanced features" for captioning and description. The tool is tentatively called CapScribe, and like Magpie, provides both captioning and description. Currently, CapScribe provides support for adding text styles, graphics, and support for a second video window. CapScribe lets authors create SMIL files for QuickTime.

2.5 Possibilities for Captions and Describing on the Web

Traditional broadcast captioning provides a single mono-spaced font, size, style, and colour, usually white, appearing over a black background. We are restricted to the area of the broadcast screen itself, so all captions must be placed on the actual video. The broadcast is always time-based, with captions appearing and disappearing for the next set of captions. There's usually enough time to provide verbatim captioning, but often additional information such as speaker pitch and intonation, music, and background sounds, is not adequately conveyed if at all. The most notable change occurring with this format the past forty years has been the shift to mixed case captioning several years ago.

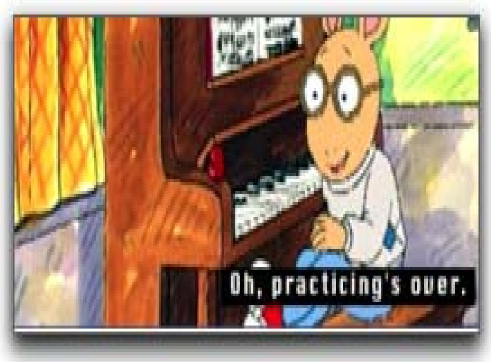


Figure 9: Example of traditional appearance, mixed case captioning.

Descriptive video is relatively new. There are several broadcast programs that provide descriptive video. DVS is limited by the amount of space between the spoken text that the describer has available for to provide descriptions of various visual information happening on the screen. Needless to say, only a small subset of visual information can be imparted to the listener under this constraint.

2.5.1 Captioning and the Web

There are no limits to how caption text can be presented on the web. All three media players support a full suite of text styles. Layout of captions is not limited to being placed over the actual video, but may be positioned off video when appropriate. Due to the small frame size of videos that are streamed on the web, it may make more sense to place text outside of the video, itself. In addition, text presented in different fonts, colours, and style can be used to identify speaker, content, emphasis, and even emotions. In the clip below, different fonts and colours are employed and positioned to the right of the video.

Figure 10: Example of captions with enhanced font styles.

Authors of original video content may choose to enhance the emotive tone of their subjects. In the following clips from the music video "Boy Like You" director Erica Shallow added enhanced captions to highlight the vocal styles of the different performers. Erica felt that having style control over her captions allowed her to capture the emotions and the energy that would otherwise go missing, and that standard captions would "flatten" the experience for the Deaf viewer.

Figure 11: Example of enhanced caption font style used to convey tone and energy.

Users can be presented with a menu of choices for how to view and interact with various content. One menu selection may offer content with basic captioning or description, while another provides a rich selection of captioning and or descriptive video experiences. Multiple language selections are most obvious, other options include, short and long descriptions, edited and verbatim captioning, or even captioning with graphics or animation.

A second video window option may provide sign language interpretation, or even mime or creative dance as a way to convey music and or sound effects. In the following example, both French and English captions are provided. This kind of flexibility provides much opportunity for letting users set different language preferences and yet enjoy the same experience.

Figure 12: Example of multilingual captions.

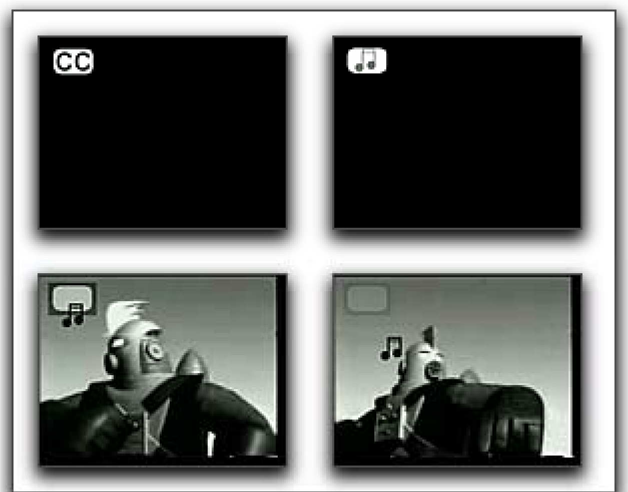
Graphics can be employed to convey feeling or even movement. Marblemedia's "Il Menu," a comical opera, was enhanced with comic art speech bubbles and icons. Using speech bubbles, words can have shapes around them, and be shaped themselves, to match the tones and affect that are produced by the speaker.

Figure 13: Example of graphics to convey language and sounds.

Graphics can be used to create dramatic effects that come closer in reproducing the impact of the original audio than plain text. The piano keys icon is marked by a red "x", indicating that the music has stopped. For the listener, the pausing of the music is quite unexpected; the "x" effect produces a similar sudden surprise visually for the caption user. Conventional captioning with text would have employed the words "Silence" or "Music stopped" which may not achieve the same effect that of the crossed out icon.

Figure 14: Example of graphics to convey silences and pauses.

In the following example, an animation was added to the opening scene of the TV series Rocket Science. The “CC” or closed caption symbol becomes a musical note which then moves and weaves through the show’s opening animated sequence, both letting the user know about the music and giving a sense of the music itself.



.

Figure 15: Example of animation to convey sound elements of video

Interactivity can play a large role, well beyond turning the captions on and off. Interactive features can provide a means of getting more information than there is room for while the video is playing. For example, an iconic symbol can be used to indicate that additional information is available. Clicking on the appropriate button stops the video and shows the additional information. Interactivity can offer the user a better description of the music, background noise, or other information that would take up too much screen space.

Video description can overcome its broadcast time limitations on the web. For the user who needs descriptive video, interactivity offers the option to stop the action for an extended description where appropriate.

2.5.2 Description and the Web

The use of the web's natural interactivity allows for a number of enhancements for descriptive video. WGBH/NCAM has already begun using an approach called "extended description" as way to make addition time for descriptions. Extended description works by freezing the video and playing the description where there isn't sufficient time to adequately describe visual information. Currently, these extended descriptions are at the discretion of the author of the description and they happen automatically for the end user.

On the Web, with enhanced DVS, different description tracks can be made available to the user. A description track that focuses on what people are wearing, facial expressions, or landscape detail would serve to add missing components that sighted users take for granted.

2.6 Taking Time to Add Access To The Web

We've worked with a number of new media developers over the past several years. Our sense has been that the lack of access is more a lack of knowledge and tools than willingness. We presented several workshops on Access, New Media and Conversion Modalities at the Banff Interactive Screen Workshop in 2003. Participants began to appreciate that access was something that went beyond people with disabilities and could be potentially helpful to all.

For example, if the video stream suffers from poor audio quality , then with the additions of captions, the audio becomes intelligible—there is no ambiguity about what is being said. In a library or classroom environment, it may not be appropriate to view web content with audio on, in these settings, the captioning enables the content to be viewed without disrupting others. Described video is similarly useful and becomes a talking movie or tv show that you take on the road with you since the descriptions explain the action.

New media developer David Bastedo made the follow comment after attending one of our workshops:

> There's a lot more to accessibility than just the disabled and while I think creating content specifically for the disabled is great, I think that there is more that should go into it than that.

Paul Ortchanian, another workshop participant, adds the following observation that seems at once representative and hopeful.

> I was the type of person who was looking at these issues as being, kind of like, to be ignored, and now I'm seeing all of this value from it. I think that we should actually, as digital designers start looking at these things as actual benefits to what we are doing and just you know using them to go further

2.7 The CapScribe Tool

The lack of a fast, easy-to-use tool for the creation of accessible media motivated us to build a prototype tool in order to explore features that involved adding text, graphics, audio (for description) and additional video to make accessible clips. The tool, called CapScribe, outputs a QuickTime SMIL file, a small text file for managing captions, graphics description, and video files. As SMIL files are quite small, it is possible to offer a number of SMIL-based choices on a single Web page.

CapScribe combines captioning and descriptions so that new media developers will have access to a single tool to do both.

Its features include:

- A flexible text editor for adding captions with styles.
- A graphics library and importer for adding graphics
- A second video option for providing an additional video to accompany the primary video
- A recording studio for adding video descriptions
- A layout screen for designing media placement
- Output to SMIL or Quicktime formats for playback.

The following example is the main screen of the editor. It's set to the Caption option. A preview window lets the user look at the SMIL generated output.

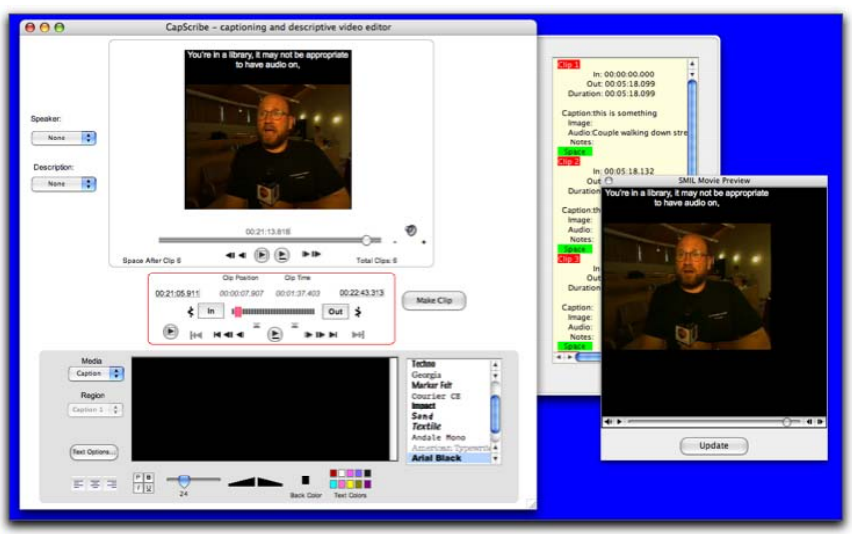


Figure 16: CapScribe Tool

The CapScribe editor prototype will be made available for public use shortly after the conclusion of the CNICE project.

2.8 Online Captioning of Flash

Closed captioning for the web affords greater flexibility for the development of captions than that currently found in broadcast environments. This document section will provide a brief overview of current captioning practices as they relate to broadcast environments. Next follows a discussion of captioning techniques for web based Flash content. The section close with the introduction of a new tool for captioning Flash developed for the Canadian Network for Inclusive Cultural Exchange (CNICE) Project by marblemedia, a partner in the CNICE project.

Television and film producers have long used a relatively rigid method of closed captioning in which captions are primarily created offsite by experienced captioners using specialized equipment. Captions are then added to the Vertical Blanking Interval (VBI), in line 21 of the video signal. Limited by the resources available in the captioning decoders in set-top boxes, broadcast captioning is limited in the size, font and placement of the captions. Broadcast captions typically appear as white text on a black background. Techniques and guidelines for analogue television broadcast captioning have been created with these limitations in mind.

Web captioning differs remarkably from traditional broadcast captioning due to it's very lack of limitations. Where broadcast captioning is static in nature, captioning for the web is a much more creative process. In developing the captions, the developer and the captioner (who may actually be the same person) work together as a team. Together they are better able to manipulate the captioning process to meet the needs of the target audience from both a design and usability perspective. Most importantly, however web captioning allows the option for customization of the captions by the audience members.

For the purposes of this discussion, we will use the marblemedia project www.deafplanet.com as a case study, as it required extensive closed-captioning in the Flash environment. As a result, a Flash captioning tool was specifically designed to meet the needs of this site.

2.8.1 The Flash Captioning Tool

Developed by marblemedia, www.deafplanet.com is a companion site to a television series geared primarily to Deaf children. Both the television show *Deaf Planet* and its web site appear entirely in American Sign Language (ASL). As a result, www.deafplanet.com had specific design and usability requirements with respect to the video provided on the site. All episodes of the show are available on the site in video format, and due to the use of ASL, it was absolutely necessary that the video be of high enough quality that the ASL, and specifically the fingerspelling that forms part of the language, be easily visible and readable. Providing this video in a sufficiently high quality while not becoming "bandwidth excessive" could only occur through the use of Flash video. At the time of development there was no Flash captioning device available, so a tool was developed in-house by marblemedia to specifically to meet this need.

Developing captions in Flash allows the developers a great deal of control over the creative process. Unlike other web platforms, for example HTML, Flash reacts in a predictable way within a variety of browser (Internet Explorer, Netscape Navigator) and platform environments (PC, Mac). Furthermore, the Flash plug-in has an incredibly high market penetration; currently estimated to be in the area of 95%. In effect, the developers are able to control the captioning environment, by determining the placement, font size, and colour of the captions, without sacrificing ensuring user control.

2.8.2 Using the marblemedia Flash Captioning tool

In order to use the marblemedia Flash captioning tool, the video must first be converted into compressed Flash format and an audio script file completed. The audio script file must then be converted into a plain text format with any formatting or additional information stripped out. XML tags are added to the text that tell the captioning tool and website that the lines in the text file are captioned lines. Each spoken line of text is followed by a blank line (a space character), which then clears dialogue from the screen during times of inactivity. At this point in the development, the raw XML file of tagged alternating spoken dialogue and blank lines is complete, but does not contain any time coding information.

These files are now ready to be loaded into the Flash captioning tool. When the video that is to be captioned is loaded into the Flash captioning tool it will use file name assumptions to automatically find the captioning file. The tool allows the video to be played with regular start, stop, pause, rewind and fast forward controls. It also allows the user to navigate to specific frames within the video. By highlighting the text, the captioner can attach text to a specific frame. As the captioning tool only notes an in frame for the caption, blank lines of text are used to clear text off the screen during times of inactivity. The Flash captioning tool does not alter the video file, but merely adds frame information to the captioning text file. The captions, as developed by the captioner, are now visible upon playback of the Flash video in a web-based environment.

2.8.3 Techniques, Usability Issues and Further Considerations

Any form of video on the Internet invariably raises the issue of bandwidth and download speeds. The Flash captioning tool has successfully navigated some of these issues, but others remain. One of the key difficulties encountered in streaming video is the stalling that occurs due either to slow download speeds or bandwidth shortages. By adding tags to link the captioning file to the frame, the captions developed by this Flash captioning tool will never speed ahead of the video (should the download stall), nor lag behind. Therefore, when using the Flash captioning tool correctly, the captions and video should always be synchronized.

A method of alleviating user bandwidth difficulties is to allow users to choose between streams of video, with a user on a lower bandwidth able to choose a smaller video file than higher bandwidth users. The size of the Flash captions will not shrink with the size of the video stream viewed, ensuring that users who view a smaller bandwidth video stream still have captions that are readable.

The Flash captioning tool does encounter some difficulty in video streams that are compressed by decreasing the video's frame rate. As the captions are linked to the frames, a decrease in the frame rate means that the video and captions may no longer be in sync. Therefore, separate captioning files need to be created for video files that change in frame rate, even if they are the same video. If the frame rate remains the same, the same captioning file can be used. It is important to note that the captions created by the Flash captioning tool are contained in a small text file, and therefore demand relatively little in the area of user system resources and bandwidth.

At the current time, the Flash captioning tool only allows the user to turn the captions on or off. However, it would be relatively easy to adapt the Flash captioning tool to allow for greater user input. For example, captions created currently appear at the bottom of the Flash video. (Positioning the captions here, as opposed to on top of the video as in typical broadcast captioning, ensures that subtitles are not covered up.) A different user interface could easily allow the user to move the captions around their screen, and increase the font type, size or colour of the text. These kinds of user options would be especially useful in addressing access issues for people with visual impairments, as well, who may require larger text or captions with a greater contrast between text and background colors.

Furthermore, since turning the captioning "on" merely draws upon a text file, it could also be possible to use the Flash captioning tool to caption the video in multiple languages through the creation of multiple caption text files. Web users could then select which language they would prefer to view the captions in. Additionally, hyperlinks could be built directly into the captions to allow the user access to related information on web sites or in web based references.

One of the drawbacks of the current Flash captioning tool is that it does not allow the captioner to caption multiple areas of the screen at the same time. While this could be built in, it does have the disadvantage of making the final captioned product less flexible for the end user, as it may require the captioner to be more rigid in the positioning of the caption boxes.

2.8.4 Summary

It is difficult to apply the standards of broadcast captioning to web captioning, (specifically in the case of Flash captioning) as the development and viewing environments for the web differ radically from that of the broadcast environment. Where broadcast captioning exists in a very static environment with little user input, web captioning has the ability to meet a variety of accessibility needs and usability requirements. The developer and captioner can have a great deal more control over the artistic process of captioning the video than that available to broadcast captioners. With minimal expansion to the Flash captioning tool, there also exists an opportunity to make the captioning process interactive for the end user, by providing options regarding text style and positioning that traditional broadcast captions do not allow.

Box 2: An ASL Content Developer's Perspective ***by Elle Gadsby***

On deafplanet.com we worked from a unique perspective. Typically web production and television production occur in two separate realms, with little communication between the two. Any television or film product that is utilized on the web is often merely streamed on the web, and the film production crew may not take into account the boundaries of streaming on the web. Throughout deafplanet.com our web team worked extremely closely with our television team, and all times decisions about the television production were made with an eye to how it would work on the web. Captioning for the web was merely an extension of this cross-communication between our web and television teams.

Captioning for American Sign Language (ASL) video streamed on the web has been an extremely positive experience for us as developers (www.deafplanet.com), as it has given us a great deal more control over the captioning process, and therefore the entire artistic expression and allowed us to fully capitalize on our unique integrated production process. Typically, traditional broadcast captioning is added to the video signal after all editing takes place, and once added cannot be altered. The process primarily occurs offsite as the last step in the post-production process before delivery to the broadcaster (or to VHS, DVD, etc). Unfortunately, this leads to several consequences, the most important of which is that it is difficult to correct any errors that are found and that it is difficult for the director or producer to maintain creative control over the captioning process.

Captioning for web video has allowed us to maintain a great deal more control, primarily because of the nature of the captioning itself. Our captioning file remains as a separate file from the video stream at all times. This means that any errors that are discovered can easily be corrected, at any point. More importantly, in-house captioning for the web has given more creative control over the final captioned ASL content. Through the traditional broadcast captioning process the director has a great deal of difficulty in maintaining control over the artistic nature expressed by the captioning, such as captioning any important environmental sounds, how long a caption remains on the screen, placement, etc. By captioning in-house, and due to the flexible nature of the captioning tool itself, we are able to edit the captioning in much the same manner as the video itself was edited. The director can review the video with captioning, make any necessary changes or alterations, add in or delete captions and then review the video again. Ultimately, this treatment has allowed us the creative control over the captioning process so that the final captioned video is a true and accurate reflection of the artist's original vision for the project.

Acknowledgements

Canadian Network for Inclusive Cultural Exchange (CNICE)

Creating Accessible Online Cultural Content Discussion Document Series

CNICE Partners
Adaptive Technology Resource Centre
Banff New Media Institute
Canadian Abilities Foundation
Canadian Centre on Disability Studies
Canadian Cultural Society of the Deaf
Canadian Hearing Society
Canadian Learning Television
Centre for Learning Technology
Centre for Research in Disability Studies
Digital Frog International
marblemedia
National Broadcast Reading Service

Editors and Contributors:
Michael Boyce, Sara Diamond, Deborah Fels, Elle Gadsby, Rob Harvie, Wendy Porch, Lake Porter, Jan Richards, Vera Roberts, Marc Rosen, Nancy Sicchia, Charles Silverman, Jutta Treviranus

This project was made possible through the funding by the New Media Research Networks Fund of the Canadian Heritage Canadian Culture Online Programme.

Canadian Heritage Patrimoine canadien

3 Online Video Description

3.1 Introduction to Video Description

The general purpose of describing in this context is to enable viewers who are vision-impaired to follow the program and make sense of the material presented. This purpose is accomplished by conveying verbally essential information that the program conveys visually.

How you serve this general purpose is influenced by the type (e.g., entertainment, information, education, or demonstration). (Why have the majority of viewers tuned in to this program?) and purpose(s) of the particular work.

Following are guidelines to follow when describing video:

- Understand the type and purpose(s) of the video. So you can determine what visually conveyed information is germane to that purpose, and what's not germane.
- Understand the material presented. So you can make sense of what you see, and convey that sense to the learner.
- Understand what someone who can't see needs to know to follow the program and make sense of the material presented. So you can determine what needs to be conveyed, and what doesn't—because it's apparent from the audio, or because it's not helpful (enough).
- Observe closely and accurately.
- Speak simply, clearly and concisely. So what you say is easy to understand the first time it's heard.
- Respect the audio. Description is intended to complement the audio, not compete with it and create confusion.
- Be an objective and positive observer. Your purpose is not to convey your personal feelings about the program or to point out what you perceive to be its shortcomings.
- Disappear. Good description directs attention to the presentation, not to itself.

3.1.1 Real-Time Online Video Description

Access to video content (on-line and television) for people who are low vision or blind has only recently received considerable attention. The demand for this technology is growing as the low vision community becomes a stronger advocate for access to digital technologies and legislation is put in place to meet those demands. Television and film are important cultural experiences that help to shape and define society. Being able to share and participate in these experiences on an equal level is a crucial element in providing social and cultural accessibility for people with disabilities.

Video description (DVS), used by people with low vision to access television and digital video content is a relatively new technology. It provides spoken descriptions of the visual content (costumes, settings, facial expressions, action, etc.) of the video interjected in the audio pauses of the original soundtrack (CRTC, 2004). These descriptions may not be synchronized with the occurrence of the visual information due to the timing and availability of the audio pauses (e.g., the description may occur sometime before or after the visual information is presented). It is slowly becoming a more accepted and standard practice in the broadcast industry. Broadcasters are particularly concerned with the increased expense and timelines to air required to broadcast media adhering to new mandatory DVS requirements of regulatory agencies such as the FCC and CRTC (FCC reference, CRTC, 2004).

Using the SAP channel on stereo televisions, major television outlets in the US are required to provide 50 hours of video description per quarter (Crawford, 2003), and only recently, in Canada, provision of video description is a condition of the licensing agreement (broadcasters operating in large markets are required to provide 2 hours per week of described video description for priority programming) (CRTC, 2004).

The standard practice for adding video description is to outsource the finished production product to specialty service providers. The quality of the resulting work is often lacking in several respects. First, describers have no first hand knowledge of the original intentions of the writers, directors, and actors. As a consequence they will make a series of choices that will affect the viewer’s understanding of the original content. For the describer, knowing what to describe from a rich visual scene is crucial to communicating a general understanding of the story and its more subtle underpinnings. In addition, descriptive video is especially challenging when there is not enough silent space in the audio track to convey what is taking place on the screen.

The process of post-production video description is similar to a production process of the video content itself—a script is written, a narrator is hired and the video descriptions produced and edited (Office of Communication, 2004). This process can be very time consuming and expensive. The outcomes are not necessarily optimal because of the choices that were made during the process and the lack of sufficient audio pauses in the original content. There is, however, no standard process for and very few examples of real-time DVS for live broadcasts or webcasts (Media Access Group, 2004). Real-time video description for broadcast television is essentially non-existent.

Video description of on-line content is produced using a post-production model. However, there is very little of it available on-line and only limited presence in guideline and recommendation documents. The Web Content Accessibility Guidelines (WCAG) guidelines do recommend having video description for on-line video content (WAI, 2003). However, it appears that considerations are being made to move this from checkpoint 1.2 to a best practices section due to the popularity of webcams, and live webcasts. This move seems to point to the difficulty and lack of process for providing live DVS.

The benefits of DVS for users have been detailed in recent studies (Schmeidler & Kirchner, 2001). They showed that DVS significantly improves a person's comprehension of the video content, particularly when information is presented visually with no accompanying dialogue or explanation (e.g., nature programs). Other benefits highlighted in these studies include: being more comfortable speaking with sighted people about the same program, not having to rely on informal descriptions provided by friends or family, and participating in shared cultural experiences on a more equal level. However, these results have been studied for DVS content that has been carefully crafted and produced, and is error free.

Live DVS for theatre performance (Weeks, 2002) and play-by-play commentary for live sporting events. The process for theatre description involves using a well-trained volunteer describer. This individual will spend time preparing before the performance by attending dress rehearsals or early performances, and reading the script. The process for play-by-play commentary involves describing what is happening in the sporting event as it is happening. The sportscaster must have detailed knowledge of the game in order to produce accurate commentary. This type of real-time commentary does not often provide what someone or something looks like but rather describes the action.

Description of the actual performance often involves an audio preview for listeners just prior to the beginning of the performance. Strict rules regarding what describers are allowed to describe are followed. For example, describers are trained to provide "objective" descriptions and avoid opinions or interpretations of the performance (the use of emotive words such as joy, or sadness tends to be limited). Training for this process involves approximately a five day intensive workshop (Audio Description Associates, 2004). The Audio Description Associates (2004) report that a sense of timing concentration and observation are very important skills for all DVS describers so that they can provide adequate descriptions in the time available.

Live description is prone to higher error rates than post-production description due to the fact that it is delivered live and cannot be reviewed or edited. However, there is little research on rates of error for theatre describers (or what constitutes adequate description), and acceptance ratings on the DVS quality by people who are blind or low vision.
Real-time captioning produced for people who are deaf or hard of hearing may provide some insight into possible error rates and user quality expectations. Service providers claim to provide high accuracy rates (e.g., NCI claims a 98% accuracy rate; NCI, 2004), and there are speed and accuracy performance requirements specified in training regimes for real-time captioners (see Media Access Group, 2003 as an example). However, real-time

captioning is a very different process than DVS as it is a typed verbatim reproduction of spoken dialogue (Clark, 2003). DVS is an interpretation and synthesis of visual information into spoken form requiring different and potentially more involved cognitive processes.

We present an innovative technology, LiveDescribe, that has been developed to allow a single individual to analyze content and provide video description in near real-time.

3.1.2 System Description

LiveDescribe has been designed to allow near real-time video description for broadcast or web content. This process requires a different approach to video description than the conventional processes because a describer will not have time to prepare a script or even preview much of the material ahead of time. New tools are required to facilitate this type of video description that involves detection and presentation of non-dialog, low volume occurrences in the sound track and to allow the description to extend beyond the limits of those occurrences. Extending the limits must be under describer control and have minimal effect on the video presentation.

The LiveDescribe! system thus performs two main functions:

1) notifies the describer of upcoming silent or low volume occurrences and of the length of time of those occurrences in the video. These occurrences offer an opportunity for inserting a description in real time. In LiveDescribe! silent or low volume occurrences are presented as a series of blocks (indicated in blue) on a white timeline that plays along with the video material (see Figure 17).

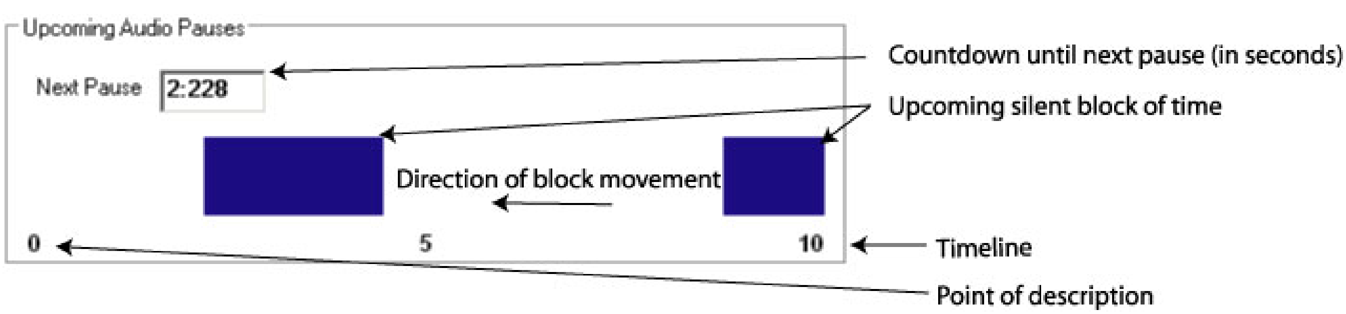


Figure 17: Silence indicators that move horizontally as the video progresses

2) automatically pauses the video (extending its running length) to accommodate descriptions that exceed the available silence period. To accommodate descriptions longer than the detected silent periods, the system pauses the broadcasted video so that the describer can finish her/his description. Professional video describers have noted the importance of this feature to provide further and richer descriptions. It also allows describers to be less accurate with the timing of the description and the silence periods, particularly important when there is little or no planning process as in the live description situation. However, video extension cannot be implemented

when video is constrained by external time demands such as with television programs.

We predict that the task of describing live will impose a high cognitive demand on the video describer. The video describer must attend to two important main activities: 1) the occurrence and length of silence or low volume, non-dialog periods indicated by the silence period indicator; and 2) the content that must be described. The describer must decide what is important to describe, determine the best or most appropriate aspects of a scene to describe, and then insert the description. All of these tasks must be accomplished as the video content is playing. Describers may have only a few seconds to a few minutes of dialog time to prepare for describing in the silence periods. Research must be carried out to determine how many silence periods can be used and at what frequency they can occur before the describer cannot cope or makes too many errors.

System ease of use is critical in supporting these describer activities because of the high cognitive load imposed by the two main describing activities. We want to minimize any additional cognitive load imposed by the system's user interface. The system is thus designed using the concept of a single push button microphone giving describer the ability to control the entire process through one button. The button is pressed once to start the broadcasting of the describer's voice and then pressed again (released) to stop the broadcast (see Figure 18).

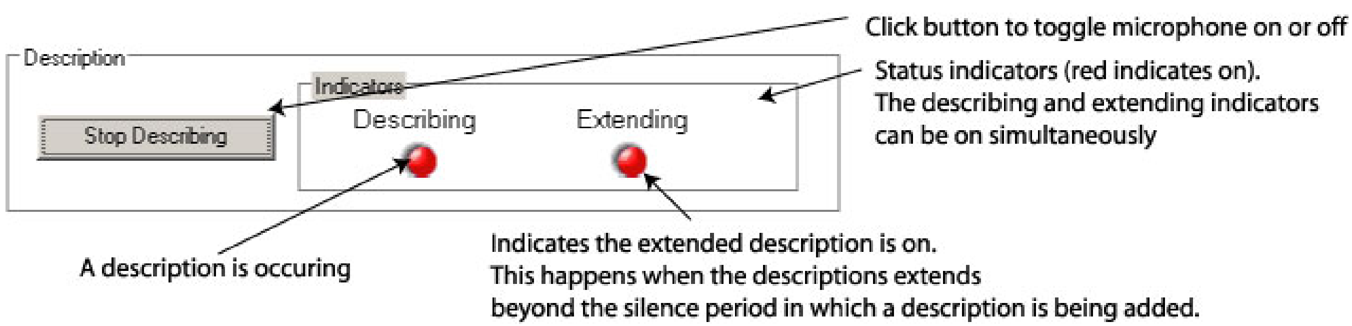


Figure 18: Describer control and status indicators.

During broadcasting of the describer's voice, the system monitors the sound level in the source video. If there is sound and the describer has not stopped the description, the source video is automatically paused by inserting frames in it and extended description can continue. Once the describer stops his description, the video continues from the point where frames were inserted. Visually these events appear as a video pause which may be disconcerting for a sighted viewer. Again, research is required to determine the tolerance levels of vision impaired and sighted user pairs for this type of video behaviour.

3.1.3 Procedure for describing a live stream

The procedure for describing a live broadcast or web stream involves three main steps. First, the describer must first define the video source. The video source is displayed in a separate window that occupies much of the main user interface area (see Figure 19). Currently the system can only work with digital media, specifically in Apple QuickTime format. Once displayed, there is a delay while the system looks ahead and analyzes the upcoming portion of the video for silent periods. The amount of look ahead time is defined by the describer and we estimate that this would typically be within the range of 10 to 60 seconds.

For example, if the describer wishes to see the silent periods for the next 30 seconds, the system will delay the processing time by 30 seconds. The describer must wait for 30 seconds before beginning the description process and the video broadcast will be delayed by 30 seconds as well. This short delay means that the system is actually *near* real-time because of the delay introduced by this "look ahead" time. However, it provides the describer with at least 30 seconds of advanced notice of and opportunity for preparation for upcoming silence periods. Whether or not this is enough preparation time depends on the visual complexity of the content and describer knowledge of the content style. For example, if this is a live sporting event, the describer may have experience with describing that sport as well as advanced knowledge of the sport itself such as the rules, team colours, players, and venue appearance reducing the potential preparation time required for producing appropriate descriptions.

The second step involves determining which silence periods will be used for description. Below the video area in the interface LiveDescribe! is the timeline that shows the time and the length of the upcoming audio pauses. These pauses are indicated as blue rectangular bars. The width of each bar indicates the length of time of the silence or low volume, non-dialog period. A countdown timer is provided to show the describer exactly how many seconds are remaining before the next silence period (see Figure 19).

The next step is for a describer to insert and broadcast descriptions with the video material. The describer accomplishes this task by "pressing" the Describe button. As seen in Figure 19, this button is located below the timeline. Status indicators are located beside this button to provide feedback to the describer on which description process is active. When the describer clicks on the button to broadcast a description, the "Describing" status light appears to show the system's current state. When the button is clicked again, the light disappears indicating that the describer's voice is no longer being broadcast. The "Extending" status light appears when the describer continues to describe into a non-silent part of the video. In this case, the system is still broadcasting the description with the original source video paused. When the button is clicked again, the light disappears, the describers voice is no longer broadcast, and the video continues to play.

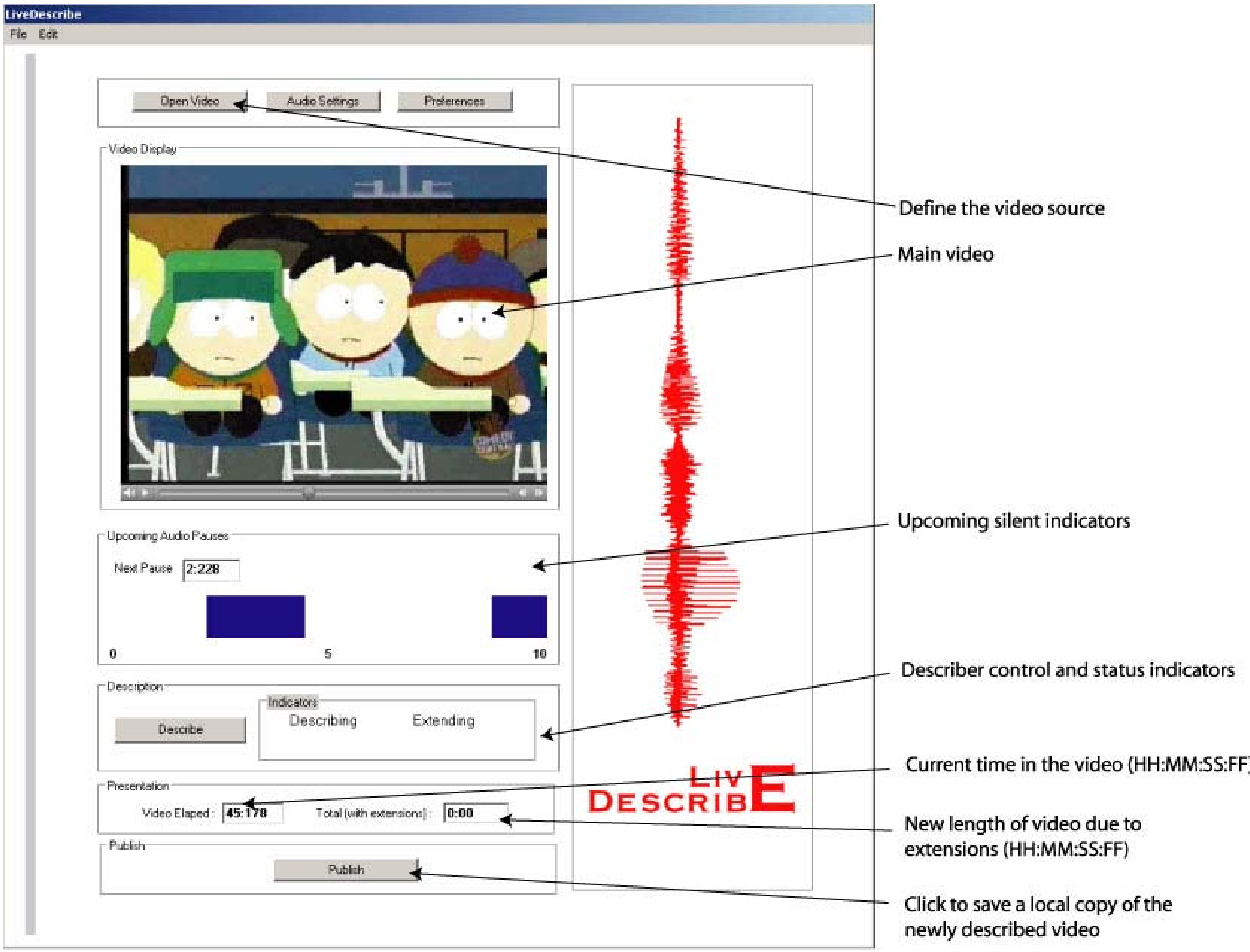


Figure 19: Describer interface.

There is publish option located below the description controls. This option will allow the user to save a copy of the newly described video including the video extensions.

There are two counters located at the bottom of the screen to provide a general overview of movie status. These indicators show how much of the video has elapsed, and the new length of the video due to any extensions.

3.1.4 Preferences

It is important to provide describers with control over their environment particularly when the task of real-time describing can be so intensive. As is shown in Figure 20, the describer can modify the following viewing and analyzing options: 1) amount of time LiveDescribe! will "look ahead" and analyze the silence periods in the video; 2) the volume threshold; 3) the accuracy level; 4) the pause length; and 5) the auto extend.

The "look ahead" time is defined as the amount of the video (in seconds) and the corresponding silence periods that the system must detect, analyse and display for the describer. The longer the "look ahead" time, the less real-time the system becomes. However, it also provides the describer with some preparation time. The describer must therefore balance preparation time with the need for timeliness. The maximum "look ahead" time is equal to the entire length of the pre-processed video.

The volume threshold option defines the level of sound (in decibels) in the video that is considered "silence". The higher the threshold option, the fewer silence spaces will be found. If the volume threshold is too low, the system may offer silent periods that contain low volume speech (e.g., whispers). While this is one simple method for detecting and displaying the "silence" portions of the sound, there is no guarantee of excluding competing dialog depending on the threshold set. Further research is required to develop algorithms that detect the presence or absence of speech (rather than only volume levels). This would ensure that description is not provided over content dialog.

The accuracy setting (in percentage) defines the level of volume variation allowed within a silent period. For example, if the system has detected a pause and then there is a short-duration cough, or a door closing during the silent period, the describer can set the accuracy setting below 100% so that the system will ignore lower volume, short duration sounds and report the entire segment as silent.

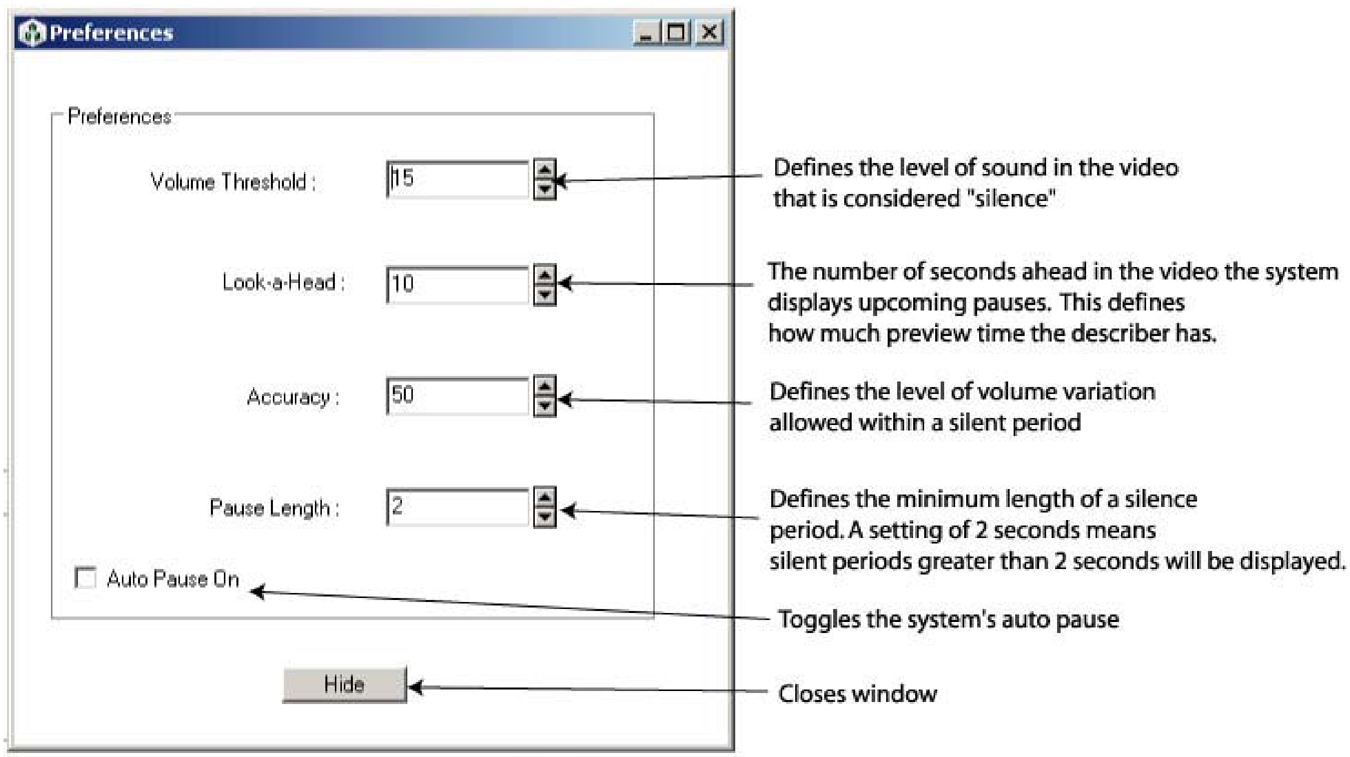


Figure 20: Preferences screen.

The pause length (in seconds) defines the minimum length of time considered as a pause. If the length is set to two seconds, the system will only show pauses that are longer than two seconds, ignoring silences periods that are less than two seconds in length.

The auto-extend option allows the extension feature to be activated/deactivated (the extension feature allows the video to be paused if there is description during a non-silent period). If auto-extend is selected the system will automatically extend the video when non-silent periods are detected and the describer is describing. The default position for auto-extend is “on”.

3.1.5 Technical Difficulties

Multimedia files are inherently very large (e.g., video files can easily occupy 30MB per minute of video content). Full screen uncompressed video requires powerful and expensive hardware to manipulate and play. For example, each frame of 640X480 uncompressed full colour video using a resolution of 24 bits per pixel would require 922 KB of disk space. If the video was to be played at a frame rate of 30 frames per second, it would require 221 MB of storage space per second (a 30 second video would occupy approximately 6.6 GB of hard drive space). The difficulties imposed by such a large file requirement are obvious. Large multimedia files require a significant amount of memory and processing power. Even compressed video files can be very large. Algorithms designed to manipulate these large files must be specially designed to accommodate them. For example, traditional backup procedures, such as those seen in sophisticated word processing packages, cannot be used as copying a 6.6 GB file would require a significant amount of system time and likely interfere with other processing.

This issue becomes even more significant when dealing with a time sensitive operations such as describing a live video stream. Any live descriptive system must at least keep up with the minimum demands of broadcasting a continuous stream of video. A viewing audience would not be forgiving of a video stream that is fragmented or distorted due to computer processing constraints of large files.

Bandwidth and network traffic is another difficulty associated with the high demands of multimedia data. For a technology such as live description to be successful, users must have access to a network delivery system capable of distributing high quality audio and video. The television medium would not be appropriate since live description as described here would require an extension of the programs over time, a requirement obviously not possible in television. The next most appropriate medium would be the Internet. It is possible using existing network technology to deliver broadcast quality video to a wide audience. However the technology is expensive and not widely available to mass consumer markets. The high-speed networks required for multimedia content delivery are currently limited to large cities where high population density warrants the installation of these networks. Users in rural settings with little or no Internet access would be excluded.

There are some technical solutions to better management of large multimedia files. These include:

1. dividing large files into smaller, more manageable chunks of data. This chunking could significantly reduce the processing requirements of the data, but would of course use additional system resources to divide and then reassemble the data chunks.
2. using sophisticated data compression methods to reduce the file size. Advances in QuickTime and MPEG technologies are very promising in this area. However, data is still lost and we need more research to determine whether this is acceptable to viewers.
3. using "bigger and better" hardware is probably the most effective and direct way to address with the problems associated with large multimedia files. However, it is a very expensive option that will always be in the future.

Traditional video descriptions are currently performed within the controlled environment of a recording studio. This type of environment is best suited for obtaining the highest quality recording of the describers voice as it is a permanent facility dedicated to recording high quality audio. Recording studios have access to high quality microphone and recording technology. For LiveDescribe! it is envisioned that the describer could work alone and perhaps even from home or a broadcast studio without the benefit of a traditional recording studio. Many broadcast and home-based centres are not equipped to produce audio content (just as many recording studios are not setup up for Internet broadcasting). These equipment differences can lead to a noticeable decrease in the overall quality of the audio production including the describer's voice, and room noise as less sophisticated recording equipment and space would be available. Microphones and soundcards available in most computer systems do not meet broadcast quality audio standards and could not be used for broadcast quality describing. Again, research is required to determine audio recording standards that would be acceptable to viewers in a live description situation.

3.2 Discussion and Recommendations

There are currently no standards of practice or even many examples of live description for television or for the digital media. Much more experience and research is required in order to make adequate and comprehensive recommendations and best practice guidelines. However, we have made many observations and identified key areas for further work.

Similar to live captioning, we predict that live description will be more error prone than content described after production. With post production description, content is carefully reviewed, and scripts and describers are prepared in advance. Experts in the description process and a CNICE partner evaluated LiveDescribe! and suggest that "With description you have to first understand what you are seeing, then determine if this is relevant to what this program is about, and if it is important to the audience understanding it. Then finally it all has to be put into words" (Rosen, 2004). The evaluators suggest that some preparation is required even for live description. This preparation could take the form of production scripts and/or a priori knowledge of the subject matter. For example, a describer could have knowledgeable about a particular sport. It may then be much more feasible for that person to provide relatively effective descriptions than a person who knows nothing about it.

In LiveDescribe!, there is no opportunity for review, writing scripts or correction in a live description situation so any preparation must be carried out before the actual event occurs, similar to how theatre description is accomplished. One important challenge in designing live description processes and tools involves determining the quantity and types of errors (e.g., omission, mis-descriptions) produced by describers for a variety of live events and timeframes, and the level and quantity of errors that would be acceptable to viewers. It would also be important to determine the level of cognitive load and fatigue rates imposed by the live description process and the LiveDescribe! tools.

Unlike closed captioning, quantifying and qualifying errors contained in a video description is a difficult and subjective task. It involves subjective judgments on whether the description of a particular scene or series of events provides a sufficiently accurate and complete “picture” to users of the visual information. For closed captioning, errors can be readily identified by incorrect spelling, missing words, or words incorrectly matched to the dialogue of the show. However, determining whether a description is correct or sufficient involves a personal judgment.

Guidelines developed for post production as seen in the section on providing description for Internet Programming in the main document point to several quality guidelines. They also point to the difficulty describers have in providing descriptions in real-time of unfamiliar content. However, there is much more research to be carried out to determine quality and quantity limitations, levels of subject knowledge and content familiarity that is required in order to successfully achieve live description, and the cognitive load imposed by live describing tasks.

In conclusion, we have developed a tool to allow descriptions to be generated for near live video description and carried out limited evaluation. Allowing describers to extend descriptions beyond the identified silence period and recommending describers have some content knowledge a priori makes live video description a possible reality. There is still considerable research required to determine acceptable quality parameters, efficient methods for describers and ways to reduce the high cognitive load and possible high fatigue rates imposed on describers by live description tasks.

ACKNOWLEDGMENTS
We thank the National Broadcast Reader Service (NBRS) for all of their assistance in this project. Funding for this project is provided by Canadian Heritage. The authors also wish to gratefully acknowledge the University of Toronto and Ryerson University.

Acknowledgements

Canadian Network for Inclusive Cultural Exchange (CNICE)

Creating Accessible Online Cultural Content Discussion Document Series

CNICE Partners
Adaptive Technology Resource Centre
Banff New Media Institute
Canadian Abilities Foundation
Canadian Centre on Disability Studies
Canadian Cultural Society of the Deaf
Canadian Hearing Society
Canadian Learning Television
Centre for Learning Technology
Centre for Research in Disability Studies
Digital Frog International
marblemedia
National Broadcast Reading Service

Editors and Contributors:
Michael Boyce, Sara Diamond, Deborah Fels, Elle Gadsby, Rob Harvie, Wendy Porch, Lake Porter, Jan Richards, Vera Roberts, Marc Rosen, Nancy Sicchia, Charles Silverman, Jutta Treviranus

This project was made possible through the funding by the New Media Research Networks Fund of the Canadian Heritage Canadian Culture Online Programme.

Canadian Heritage Patrimoine canadien

4 Remote Real-Time ASL Interpretation

4.1 Introduction

Video communication or video conferencing is becoming a much more commonly used and effective means of interpersonal communication (Finn, Sellen & Wilbur, 1997) such as for distance learning, business meetings and social communication. As hardware becomes more available and less expensive, and software, signal processing and compression technologies become more stable and efficient, there is an increasing interest and experimentation with the technology by the general public and by business.

Video communication technology is designed to support real-time communication between one or more users when long distances separate them. Video-mediated communications, as well as other forms of remote interpersonal communication (e.g., traditional telephony, email, and audio-only communication) are distinctive by the nature of the medium from which they are constituted (Olson, Olson & Meader, 1997). Each form of communication has particular strengths and weaknesses. The goal in using a particular communication medium is not to replicate identically that which can be easily achieved when interlocutors are face-to-face. Rather, it is to find ways to provide participants with a means to achieve all the interactions that are necessary to complete the defined task in a productive and efficient manner. We need to understand the strengths and weaknesses of each medium of communication in order optimize its use and find ways in which to overcome their weaknesses.

One of the strengths of video conferencing technology is the opportunity it offers people with disabilities to communicate with each other, with service providers and with business associates without having to travel. One important benefit relates to sign language users, who can communicate in their own language (using their own cultural expressions and dynamics) with each other and with people who are hearing and at a distance.

ASL speakers living in remote communities can now have access to and participate in Deaf culture where Deaf cultural events are often limited to urban settings with a large population of ASL speakers. Other telecommunications technologies such as the telephone, TTY devices and synchronous chat use text and so do not allow sign language users to communicate in their first language. For people who use sign as their primary language, text based expressions of a spoken language like English should be a viewed as second language and as such is a significantly less expressive and intuitive language for sign language users. Through video communication technology, sign language users can now have access to a more equitable means of communication. Sign language users have been experimenting with video mediated communication for some time and many lessons have been learned to mitigate some difficulties encountered with video conferencing technologies.

In order to ensure that people with disabilities are adequately served by video conferencing technologies and remote meeting procedures we must ensure that the technology and procedures are inclusive and accessible. Guidelines that have been established for video conferencing situations may not be inclusive particularly for users who are deaf. Researchers at Gallaudet University have devised some guidelines to assist sign language users in experiencing more effective one-on-one video mediated signed communication (Williams, 2002) based on the findings and recommendations of the above research. However, these guidelines only apply in one-on-one situations with small field-of-view cameras; they have not been extrapolated to one-to-many situations with high end technology or where there is an interpretation need – a potentially common application.

While video conferencing technologies offer exciting opportunities to support people with disabilities, there is a dearth of research, case studies and best practice literature to support the procurement, installation, management and operation of inclusive video conferencing services. This report provides a set of guidelines and best practice statements that will assist organizations and individuals in establishing accessible video conferencing.

In this document, there is a focus on guidelines for users who are sign language users or who are hard of hearing. These two groups of people have the greatest variety of unique needs for video conferencing. Lack of accessibility therefore has the greatest impact on them. In this document, we provide an overview of the technology, a description of remote sign language interpretation issues and requirements, small case studies and user reports. Specific access issues are identified and discussed followed by guidelines and recommendations to address these issues. Many of these recommendations are based on our experiences since 2003 using high-end video conferencing for remote interpretation.

4.2 Technology Overview

Before addressing the unique aspects of video communication for people with disabilities, we will provide a brief review of common video communication technologies. There is a standard and common set of hardware technologies and configurations for video conferencing regardless of how the system is used and who the users may be. First, video conferencing relies on having network connectivity so that video and audio signals can be transmitted in real-time over a distance (often over long distances).

There are two main types of network transmission technologies used for video conferencing, Integrated Services Digital Network (ISDN) and Internet Protocols (IP). ISDN, introduced in 1984, is designed to allow fast digital point-to-point connections over the public telephone network (Total Access Networks, 2004). Video communication signal processing and transmission are guided by the International Telecommunication's Union (ITU) H.320 video standards (Polycom, 2001). Guaranteed and consistent quality of service is provided by ISDN as the signal does not fluctuate with network availability because it is a continuous feed and direct connection. Common transmission speeds for ISDN used in video conferencing applications range from 128 kilobits per second (kbps) to 384 kbps. These transmission speeds allow audio-video signals to be consistently transmitted at near broadcast quality (broadcast quality video transmission is 29.95 frames per second (fps)). The cost of this service is based on a monthly line charge (e.g., for 128 kbps service, two 64 kbps lines are required) plus "on air" charges per minute. Video conferencing is "on-air" as soon as a connection is made and is only disconnected when the video conference is complete.

IP videoconferencing involves using Internet Protocols and technologies to process and transmit live video and audio signals. Video conferencing using IP protocols is governed by the ITU H.323 video standard (Polycom, 2001).

Internet protocols (IP) require that data signals are divided into small data packets and routed through various available networks rather than through the continuous feed, direct point-to-point connection available with ISDN. The IP video conferencing signals must share the network with all of the other Internet traffic resulting in inconsistent and fluctuating quality of the video and audio signals (ranging from 2 to 29.95 fps). As a result, high-speed Internet connectivity is required to have effective IP-based video conferencing. Much research and development effort has been placed in developing technical solutions for improving the quality of service for IP video conferencing. Some of this research that has met with some success includes better compression and signal processing techniques (Muresan, et al., 2002), and ways of assigning transmission priorities to video and audio signals (Babich & Vitez, 2000).

Gatekeeper technology is a network device that provides addressing service for H.323 (Internet-based) videoconference clients. It may also be configured to impose network bandwidth restrictions and to allow or disallow a call. Registration by the videoconference client usually takes place when the client is started; the address of the gatekeeper is put into the client's configuration. Use of a gatekeeper allows a videoconference device to "dial" another device using the videoconference address rather than an IP address (which could be changed by DHCP). Gatekeeper services might include bandwidth and call management. Bandwidth controls the number of H.323 terminals permitted simultaneous access to a LAN. Call Management maintains a list of active H.323 calls. This information indicates when a called terminal is busy, and provides information for the bandwidth management function. One or more gatekeepers may reside anywhere on the network, fully integrated into another networking device or operating as a standalone software application on a desktop computer.

Costs for IP-based video conferencing can be significantly lower than ISDN and are mostly related to the speed or bandwidth of connectivity rather than the length of time the video conference is in session. Cable modem or Digital Subscriber Line (DSL) connectivity is generally available and relatively inexpensive, and would be considered as the minimum bandwidth required. Similar to ISDN cost structure, the cost of this service is also based on a monthly rate plus "on-air" per minute charge for use. However, these costs are considerably less than ISDN because of the shared nature of IP-based networks. High-speed networks, and/or fibre-based Ethernets only improve the quality and reliability of video conferencing but costs are significantly increased.

A multipoint videoconference allows more than one site to connect at the same time. A multipoint videoconference involving 3 or more sites is possible through the use of a bridge or multipoint control unit (MCU). Some pre-configured systems such as the Polycom FX have built-in bridges which allow you to connect to multiple sites. Third party services such as Bell Canada bridge services can be rented on an hourly basis.

The video communication system itself consists of two subsystems, one at each end of the network connection. Each subsystem is composed of at least one video camera with optional zoom controls, microphones (desktop or wireless), speakers, a small preview screen (picture-in-picture capability), and monitors or large screen televisions. These subsystems can be PC-based such as iVisit and Netmeeting setups or can be dedicated hardware such as a PolyCom™ ViaVideo® II. Many sources are available to describe the characteristics of the various hardware/software options (e.g., Video Development Initiative's Video Conferencing Cookbook, (VIDe, 2004) is a good source for general detailed information about video conferencing hardware).

There are many different types of subsystem hardware that range in quality and cost. The simplest and least costly hardware is a webcam (at $50.00) and PC-microphone. Often this type of hardware is "plug and play" technology that is directly accepted by the computer. As a result this type of setup can be installed directly and used immediately with software such as NetMeeting or iVisit. However, this type of hardware has few controls and adjustments. For example, a simple webcam may be limited to focus control (there are not lighting, motion or zoom controls).

The audio system for video conferencing consists of some combination of audio headset, telephone handset, microphones, speakers, and digitising devices (hardware and software). One of the most traditional microphones in video conferencing is the lavaliere microphone, which is a miniature microphone that you clip onto the clothing of the person speaking. Wearing a lavaliere microphone reduces the feedback noise that is picked up by the other type of microphones. A second common type of microphone is the room microphone, which is a unidirectional boundary microphone. These microphones lie on the surface of a conference table or desk. They detect speech with a clear, natural sound. This type of microphone is specially designed to filter out room acoustics – much more so than a conventional microphone on a desk stand. A third microphone type that is often used with desktop video conferencing is the stick microphone or microphone built into the camera. Such microphones lack good audio quality and can effectively shut down a video conferencing. These types of microphones are the least expensive audio solutions but they are also the lowest quality. There is no adjustment in the echo or gain features for these microphones.

As the quality of the camera and microphone setup increases there is a corresponding increase in functionality, controls and the cost of the hardware. For example, a top of the line camera may cost $2,500 but will have a considerable number of functions such as zoom, pan and tilt controls, back light, and automatic tracking. In addition, the optics system in these types of cameras is of a much higher quality than typical webcams. These types of cameras are typically used with systems that allow remote control of camera functions so that the remote participant can control the zoom, pan and tilt functions of the remote camera.

4.2.1 Connecting IP technologies with ISDN

IP video conferencing systems and ISDN systems can be connected together using a gateway communication system. A gateway offers the most flexible link between ISDN (H.320) and IP (H.323) videoconferencing standards and delivers full interoperability between ISDN and IP endpoints. The Gateway integrates seamlessly to provide H.323 management and to control network capacity. For instance, you can conduct conference calls seamlessly from any endpoint to any other endpoint – regardless of system type or network configuration.

4.2.2 Bandwidth recommendations

Videoconferencing requires a large amount of data to be transmitted in a short amount of time. The recommended minimum amount of bandwidth is 384kbps for a typical business quality videoconference.

The main consideration for any video conference is how many frames per second are being transmitted. A good quality video conference requires 30 frames per second video transmission.

Any data rate higher than or equal to 384 kbps will support a video screen update of 30 frames per second, equivalent to VCR playback quality television.

Any data rate lower than 384 kbps will support a video screen update of 15 frames per second or less, which is still usable, but will appear slightly jerky under rapid motion.

4.2.3 Applications

A typical video conferencing interface is based on a telephone metaphor where remote participants are connected using telephone or telephone-like (IP address) numbers. Terminology such as dialing, hanging up, answer, busy, and address book is used to refer to the connecting tasks accomplished by users with the system. Unique aspects of the interface include video displays (local and remote windows) where the local window often appears as a picture-in-picture window in the main video window. Settings menus are often available to customise the look and placement of the video windows, and to adjust various technical properties such as compression quality or levels, video and audio settings and controls, and file management. On the high-end video conference interfaces there are also specialized camera settings for remote and local camera controls and automatic speaker detection.

Video conferencing can involve more than just the real-time transmission of audio and video images. Applications such as file and application sharing, logging audio and video, capturing and sending of static images, simultaneous text chat can be incorporated into video conferencing software. While all of these applications can be very useful in supporting work group needs, they can also interfere with communication strategies. In addition, appropriate access to all of these applications must be included for people with disabilities.

4.2.4 Physical space/Room technologies

Although the number of hardware providers of high end video conferencing equipment is relatively small and system configurations are limited, it is important to carry out a needs and task analysis, (i.e. what is the purpose of the video conferencing system in the organization and for what tasks will it be used) and an environmental survey. These surveys are useful in determining the system requirements, the human resources required to manage the system and an appropriate room or room modifications that might be required. The results of these analyses will have cost, management and timeline implications. For example, if lighting renovations are required the project budget will increase.

4.2.5 Environmental Considerations

In order to have effective video communication, the hardware must be properly configured and housed in an appropriate environment. There are numerous technical guidelines published to assist organizations and individuals in proper environmental assessments and setup for various configurations of video conferencing hardware (see Polycom, 2004; McAteer, 2000; and Brightline, 2002 to name a few) but few of these guidelines address the special considerations required to accommodate people with special needs, particularly sign language users and remote interpreters.

4.2.6 Future Considerations

Some industry pundits (Lee, T., 200X) suggest that instant messaging and mobile cell phone technologies will assist in the acceptance of personal video conferencing. Already, cell phones incorporate instant messaging capability with image processing and video capabilities. As the next generation enters the workforce they will already be exposed and using instant messaging and video conferencing., “These kids are ready for video conferencing because they were brought up in front of cameras their whole life.” (Brandofino, M. Glowpoint Inc. page #.)

4.3 Technology issues related to accessibility

Video conferencing technology mostly involves audio and video communication mediated through computing and network systems. There is some, although considerable less, interaction with computer software required to carry out a video conference. Accessibility by people with disabilities involves access to communication. For people who have no communication disabilities, video conferencing remains accessible. For example, for people who are blind or have low vision, the video conference becomes an audio-only conference. All of the standard audio conference issues such as, ensuring software applications are accessible to screenreaders, and the need to have visual materials made available prior to the conference and readable by a person who is blind are relevant here. However, none of these issues is unique to video conferencing and guidelines for inclusive audio conferencing, and access to images and other visual materials are available from other sources. An example of web guidelines are the Web Accessibility Guidelines of the W3C, (W3C, 2004).

For people who are keyboard users, there are also very few issues related to video conferencing applications that are different from other software applications. One unique issue is the accessibility of camera controls and pre-set buttons. These controls are often available through a remote control or button panels and/or software buttons. Remote control settings may need to loaded into specialized assistive technologies such as an environmental control unit in order to this person to access the camera controls. Where controls are provided through software, keyboard access is required to allow use of these controls.

The people with the highest need for access solutions to video conferencing then are people who are deaf or hard of hearing and who are non-speaking. The remainder of this report will focus on access issues and guidelines to ensure that there is access to audio and video communication for people who are deaf or hard of hearing.

4.3.1 Video conferencing and use with sign language interpreters for people who are deaf.

One exciting opportunity offered by video conferencing technology is that of supporting people who are sign language users in accessing sign language interpreter services; services that can be particularly difficult to obtain in geographically remote locations. Sign language interpretation is required to mediate communication between deaf and hearing people. When sign language users are in geographically remote locations, it is now feasible for them to have access to interpreter services using video mediated communication technologies. However, there are important considerations and differences to address. In this section, we discuss the unique issues that arise when remote interpretation is required and provide amendments to technical and use guidelines to account for these special needs. These recommendations are based on our experiences since 2003 using high-end video conferencing for remote interpretation.

We will also provide a brief introduction to sign language interpretation and video remote interpreting as a sub-specialty within sign language interpreting.

4.3.1.1 Sign Language Interpretation

Sign language interpretation is required when people who are deaf must interact with people who are hearing such as in business meetings, for court, and for accessing social and medical services. It is essential for providing equal access to these activities and services for people who are deaf, and in many western countries it is required through legislative initiatives (for example, see the Americans with Disabilities Act, US Department of Justice, 2003).

American Sign Language (ASL) is the most prevalent sign language used in North America although it is not the only one (e.g., in Quebec, Canada Langue des Signes Québécoise is used). ASL, like other sign languages, is a visual-spatial language without much grammatical similarity to English (Stokeo, 2001). It is considered a linguistically complete, natural language system where the elements of the language are not equivalent to vowels and consonants of written languages, and it is not a translation of English. Vocabulary and meaning of concepts in ASL are expressed using a series of hand gestures, facial gestures such as eyebrow motion and lip-mouth movements, and body movements that change in time and space. This series of gestures cannot easily be represented by a single written or spoken word. Fingerspelling is used to spell out a word in alphabetic characters that has no sign. For example, there is no unique sign for email and it is therefore represented as the hand gestures for an “e”, an “m”, an “a”, an “i” and an “l” separately.

In order for communication with hearing people to be effective in any situation, the person who is deaf must be able to completely express herself, and must also have complete access to the meeting and communication activities. Interpreters are required to provide a translation of what is said as well as broker any linguistic clarification required and cultural differences experienced between the two parties (Avery, 2001). In face-to-face situations with inexperienced participants, communication can be awkward and difficult. Common practices of turn-taking, participation, maintaining a common understanding and access are noticeably different from meetings that occur between all hearing participants or all sign language participants. For example, common cues to indicate a speaker is ready to relinquish the floor to another speaker are different between hearing and sign language users.

Hearing people use a variety of verbal and non-verbal signals such as eye gaze, asking a question of another person, and head turns to manage turn-taking (Preece et al., 2002) in a group environment. All of these signals are very subtle and learned and participants are generally unaware of when, where, how and the speed with which such tactics are used. Sign language users tend to use more obvious hand gestures such as pointing, and waving their hands to gain attention and fewer non-verbal cues.

In an interpreted setting, the flow of communication tends to be slower due to the translation process. In addition, the sign language user’s visual attention is with the interpreter and not with others in the meeting. As such, they cannot attend to subtle turn-taking gestures, such as shifts in eye gaze employed by hearing individuals, in order to know that the speaking floor is available. Turn-taking must therefore be slowed down and made more obvious or formal so that sign language users have a chance to keep up and participate.

Other difficulties that arise in meetings between deaf and hearing participants include difficulties maintaining a common level of understanding because of the use of two different languages to discuss common subjects, and difficulties with access to private or "whispered" conversations. These types of difficulties are not that different from those experienced during meetings between participants using two different spoken languages. However, the role of the interpreter can be very different. The sign language interpreter often must interject cultural and emotional interpretations, request clarification, and interrupt the meeting flow to allow the deaf person a chance to speak.

The role of the meeting chair in these kinds of interpreted settings is very important as management of the meeting is ultimately his responsibility. The meeting chair must be aware of the differences and needs of all meeting participants and take extra steps to ensure that all participants are included and involved.

4.3.1.2 Video remote interpreting

Video remote interpreting in general is identified by interpreting service providers as an enhancement to the existing service and requires specialised training and different terms of use. For example, the Association of International Conference Interpreters suggests that spoken language remote interpreters should only work for three hours per day and that video interpreting is significantly more fatiguing over a 30-minute turn than conventional face-to-face interpreting (Moser-Mercer, 2003). These restrictions and issues apply to sign language video remote interpreters and will likely be amplified because of the physical nature of sign languages.

Video remote interpreting is considered a specialization within the sign language interpreting field. It is defined as the use of video conferencing technology to provide interpreter services where a participant or the sign language interpreter is located at a geographically different location. Many of the difficulties evident in face-to-face meetings are amplified for video remote interpreting and new ones arise that reflect the weaknesses of video conferencing in general.

There are numerous organizations throughout the world that provide video remote interpreting services (My Video Interpreter, 2004; SignTalk, 2004; and Community Access Network, 2004 are three examples) but there is no standardized training protocol or guidelines to overcome some of the difficulties. In addition, there are few best practice examples that demonstrate ways to use the strengths of the technology and overcome some of the weaknesses.

4.3.1.3 Challenges of Video Remote Interpreting

Many challenges arise when video remote interpreting is employed. Some of these challenges relate directly to the quality of the camera hardware and the bandwidth of the video conferencing system. For example, remote interpreters must consciously adjust their natural signing space to accommodate the camera's field of view. They cannot sign outside of the area captured by the camera. Many high-end cameras have zoom controls that can adjust this field of view by zooming the lens either wider or closer. However, a wider field of view that shows more of an interpreter's natural signing space also captures more of the background and surrounding scenery that can be distracting for the person who is deaf. Inexpensive webcams do not have an adjustable field of view and sign language users often must move farther away from the camera so that more of their signing space can be seen or they must restrict their signing to the area around their faces.

Fingerspelling tends to be very fast-paced and not well articulated in face-to-face situations. Only a very high bandwidth video conferencing system will not become pixilated and be effective for fast fingerspelling. Fingerspelling therefore must be slowed down and done closer to the interpreter's body. However, when fingerspelling slows down so does the rate of communication between deaf and hearing interlocutors. There is thus a significant impact on the potential for misunderstandings and for missed opportunities to turn take.

One important aspect of hardware technologies that has a large impact on the success of sign language use is the camera view angle or field of view. This is defined as the viewable area or scene that can be seen through the camera (Segal, 2004) and it is a function of the focal length of the camera lens. For example, a wide angle lens with a short focal length has a very large field of view or area of the scene that can be seen through the camera lens. Zooming in the camera increases the focal length and decreases the field of view to a much smaller area of the scene.

Remote interpreting removes the chance to develop rapport with consumers, and for that reason has been met with some resistance on the part of sign language interpreters.

4.3.1.4 Considerations for video remote interpreting

With video remote interpreting there can be three possible interpreter locations:

1) The interpreter is remote from both parties (hearing person and deaf person physically located together);
2) The interpreter is physically located with the person who is deaf and the hearing participant(s) are remote; or
3) The interpreter is physically located with the hearing participant(s) and the person is deaf is remote.

Each scenario requires unique considerations regarding the behaviour and perception of the interpreter/deaf person pair. However, regardless of scenarios, one aspect remains constant; the interpreter and deaf person must have constant eye contact and must be able to see each other's signs at all times. A breach of eye contact indicates that communication has been severed.
Video conferencing technology is designed to support communication between one or more remote users.

Eye contact/gaze
Chen et al (2003) have suggested that eye contact is very important in communication of any kind. They indicate that when a speaker is looking to the left, right or upward, the recipient believes that the message is not meant for him. If the eye gaze is slightly downward, although not preferable, the recipient believes that the communication is

intended for them. This could explain why television newscasters read from a teleprompter that is positioned slightly below the front view camera. Although the newscaster's gaze is slightly downward, viewers still believe that the newscaster is speaking directly to them and that the message is meant for them. The same is true in video-mediated communication.

Cameras are usually positioned above the main viewing screen (the screen that shows the video images of the remote interlocutors) meaning that eye gaze is slightly downward (people are looking at the viewing screen and not the camera). However, large viewing screens or small field of view cameras create a large gap between the camera and the positions of a person's gaze (while they are looking at the viewing screen). It appears that a person's gaze is significantly downwards (not slightly downwards) and can be very disconcerting particularly if the interpreter is remote from the person who is deaf. When the interpreter is remote the deaf person has no other connection or means of communication with the interpreter other than through the video conferencing system. Maintaining eye contact (even artificially) is crucial for deaf people using video conferencing in this situation.

The interpreter can adjust their position to the camera by sitting far enough back from the camera or zooming the camera out so as to appear to be having eye contact with the person who is deaf. However, doing so also increases the amount of background scenery that the person who is deaf must contend with and screen out particularly with low level or even lighting conditions. Spot lighting can be used to emphasize the interpreter and de-emphasize the scenery but this must be carefully planned and orchestrated; something that is not normally part of a video conference setup.

When the person who is deaf is physically located with the interpreter, maintaining eye contact between the interpreter and the person who is deaf is relatively straight forward. The interpreter can sit next to the video display so that the person who is deaf is always looking toward the monitor. However, the person who is deaf may still appear to the remote participants as though she is looking in another direction and may be disregarded by the other interlocutors because it seems as though she is not participating due to the mis-interpreted eye gaze cue. In addition, the interpreter cannot see the remote participants and may miss the non-verbal cues for turn-taking and other important meeting activities.

In face-to-face situations, interpreters know that if they make eye contact with the hearing speaker, the speaker assumes he is talking to the interpreter and loses his connection with the deaf participant. It is an automatic human behaviour that people tend to look at the person who is speaking (verbally) and thus make eye contact. Experienced interpreters usually avoid eye contact with the hearing person as a non-verbal reminder to that person that he should direct his remarks to the deaf person. Remote interpreting can interfere with this practice particularly when the person who is deaf is remote from the interpreter (interpreter is physically present with the hearing participants or remote to all participants).

When the interpreter is remote to all parties, the interpreter must look at the viewing screen to watch the person who is deaf. To all parties this seems as though the remote interpreter is now looking more directly at them and the viewing screen (and hence the interpreter) becomes the centre of unwanted focus. In addition, the opportunity for the interpreter to use eye gaze as a non-verbal cue to indicate that hearing participants should relate to the person who is deaf is considerably reduced. Careful attention to seating plans is one way to alleviate some of these difficulties.

Seating

In a video conferencing setting, seating should always be an important consideration because the remote participant has considerably less presence and prominence than those at the local site. For remote interpreting situations much more thought must be given to the position of each participant, and whether the interpreter is sitting or standing (community interpreters may be more accustomed to standing while interpreting and thus may prefer to stand. However, in general it is preferable for the interpreter to be seated). Having tools such as notepads or laptops can also be useful for tracking what is being said or presented. Table 1 shows the suggested seating arrangements for the three different interpreter locations. Note for all situations, flowers, water bottles, computer screens and other items generally located on tables should be removed to reduce the visual clutter in the camera's field of view.

Location	Seating for deaf person	Seating for hearing person(s)	Seating for interpreter
Hearing person and deaf person at same site, interpreter is remote	Seated across from viewing screen showing interpreter and in front of camera.	Beside viewing screen so deaf person can see interpreter and hearing person(s) together.	In front of own viewing screen and camera. Will be able to see deaf person.
Hearing person and interpreter at same site, deaf person is remote.	Seated in front of own viewing screen and camera.	Beside interpreter so deaf person can see both parties. Interpreter should direct eye gaze towards deaf person at all times (while signing and voicing)	In front of viewing screen and camera.
Deaf person and interpreter at same site, hearing person is remote	Seated in front of own viewing screen and camera.	In front of own viewing screen and camera.	Beside viewing screen and visible to deaf person. May not be visible to hearing person.
All people are remote	Seated in front of own viewing screen and camera.	In front of own viewing screen and camera	In front of own viewing screen and camera

Table 2: Suggested seating arrangements for all participants.

Environmental and technical issues

There are environmental and technical solutions that can be optimised for people using sign language. When identifying the location(s) for video conferencing, considerations such as physical environment (e.g. room size, lighting, acoustics room setup and furniture) and uses of video conferencing in that environment are important to optimise and renovate if needed. Table 2 summarises some of the important technology considerations when hearing and deaf people are participating in a video conferencing together.

Technical or equipment solutions are also possible and in combination with environmental adjustments can greatly assist in optimising a video conference for people who are deaf and using interpretation. For example, video transmission frame rate of greater than 15 frames per second is critical. Having consistent, high quality image transmission rates available through dedicated high bandwidth networks such as ISDN or high speed IP-based networks can provide this.

When using video conferencing with people who are deaf, audio is less important (although it is still relatively important for the interpreter), and the need for camera controls (remote and local) becomes more important. A personal microphone such as a lavaliere or lapel microphone for the interpreter rather than a high-end room microphone can be used. In the situation where the interpreter is with the hearing participants a room microphone and a personal microphone may be required. This also means that intelligent audio signal processing (in the form of an intelligent hardware or software mixer) must be available to switch between the various audio sources (e.g., interpreter voicing and meeting participants).

Technology	Hearing People	Deaf People
Audio	• High quality audio is higher priority than high quality video.	• May not benefit from the audio. • May not understand the importance of having the microphone "muted" or "on" at their site. • When using voice-activated audio, the camera will focus on the interpreter voicing and not on the deaf person doing the signing.
Video	• Tolerate low quality video • May want to continue meeting even with poor video	• Cannot tolerate poor quality video. • Will notice more minute changes in frame rate and video quality. • Will want to cancel the meeting if video quality is unsatisfactory.
Placement of the microphone	• Microphone should be placed near person talking. • Most conference microphones are sensitive enough to capture all audio when placed in room center.	• *Microphone should be placed next to the interpreter who will be voicing for the deaf person.*
Lighting and	• www.effectivemeetings.c	• *Be aware of appropriate colours to*

Technology	Hearing People	Deaf People
Background	om provides recommendations for appropriate clothing • An office divider in a neutral colour makes an appropriate backdrop. • Avoid sitting in front of visually noisy scenery.	*wear that contrast with skin.* • An office divider in a neutral colour makes an appropriate backdrop. • Avoid sitting in front of visually noisy logos or murals.
The View of the Room	• Usually satisfied with a broad view of the room and rely on the audio to know who is talking • Camera presets are helpful but not critical.	• *May prefer a large view of the room to start, and then to zoom in when it is time to look at a specific person's signing (whether it be the interpreter or a deaf participant)* • *Use of presets more critical.*

Table 3: Summary of technology considerations for video conferences involved hearing and deaf participants.

The environment must have minimal "visual noise" such as windows, wall coverings, as any thing or person that moves that will distract users. These types of visual distractions can disrupt the whole flow of communication in a video communication session that includes people who are deaf. Camera controls such as zoom can be used to adjust the image to eliminate some of these distractions. However, adjusting the image may not be appropriate in all situations (e.g., in a many-to-many conference where the remote participant needs to see all of other participants). Cameras with controls tend to be more expensive, high end cameras.

Detailed Examples of Visual Noise

- Windows can cause significant visual noise. Outside activities seen through the window can be distracting, and lighting from windows can cause difficulties for cameras that automatically adjust for lighting conditions. Cameras pointed at users sitting in front of windows will be flooded with background light and the transmitted image of the user will appear only as a dark object in front of a well-lit window.
- Window coverings such as curtains that sway, or those that do not completely cover the window can be distracting.
- Objects typically found on tabletops during meetings, like water bottles, and laptops can also contribute to visual noise.
- Tables with highly glossed surfaces cause light to reflect off the table and into the camera. Cameras that automatically adjust brightness settings will then adjust for the camera to accommodate this reflected light and most other images in the surroundings, including the people will appear dark.

Other physical factors to consider

Tables: Tables with a matte finish are recommended in order to reduce additional glare**.** Depending on the size of the sessions, tables should be portable (eg: wheels on legs, or detachable legs) to allow optimal set up. A large meeting could be accommodated by a series of tables and technologies set up as shown in Figure 21.

Objects on Table: Minimize the amount of items visible on the table. Flowers, water bottles, computers and other objects disrupt the field of vision and should be placed elsewhere.

Curtains: Solid curtains are recommended to allow the least amount of visual distraction. Vertical shades, however, are not recommended, as they allow sunlight into the room and cause difficulties for cameras with automatic brightness compensation.

Lights: Adequate room lighting ensures productive communication. The most optimal lighting is one that illuminates the users face and hands directly. However, this type of lighting is likely to cause discomfort to the user.
Also, bright overhead lighting minimizes shadows on the face and hands. Brightline Inc. (Brightline, 2002) provides recommendations and shows optimal lighting arrangements for video conferencing. Lighting and lighting placement are a function of room size, person location, wall and ceiling reflectance.

Chairs: The chairs should not be “squeaky” when rocked back and forth. Casters on casters are most appropriate for tiled floors to prevent scraping sounds when the chairs are moved. The ideal floor covering is carpet with under padding to minimize the chair moving sounds.

Room Colour: The walls in the video conferencing room should be painted a solid neutral colour (blue and greens are also acceptable). Walls with patterns or pictures are visually distracting.

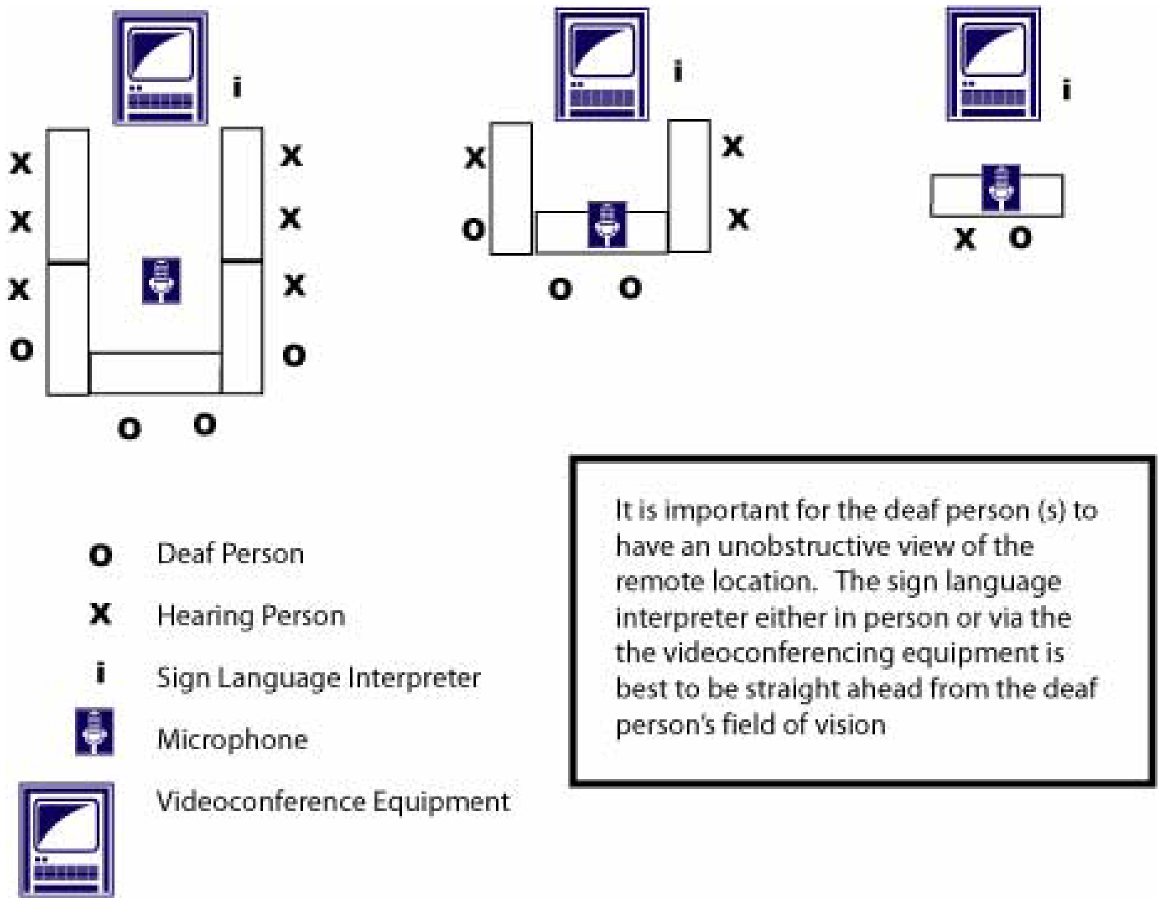


Figure 21: Different furniture and placement configurations.

Turn-taking

Another critically important consideration is that of turn-taking. Because of the difficulties in producing and understanding non-verbal cues during any video conference, turn taking becomes much more cumbersome in general. Interlocutors constantly miss turn-taking cues resulting in communication errors such as overlapping each other or interrupting, having long moments of silence, and taking control of the floor for lengthy periods of time. When someone who is deaf is added to the video conference these difficulties and errors become elevated.

It is often the role of the interpreter to mitigate turn-taking but it may be a much more difficult task for a remote interpreter because they too may miss the turn-taking cues. At the Canadian Hearing Society, attempts to solve this particular difficulty involve employing a conference manager at one site. This person is not the meeting chairperson and is only responsible for managing all of the technology as well as maintaining a formal speaker's list (by monitoring people's desire to speak). This approach is the most successful approach to date but it not the most cost-effective method because it requires another person to facilitate the meeting.

Fels et al. (2000) investigated the use of technology such as lights (flashing, spinning, etc), and a waving hand as a way of improving turn-taking for video conferencing. They found that having a waving hand mechanism activated by a remote participant was very successful at gaining the attention of all participants. While this may be appropriate for classroom settings or even meeting settings, a waving hand may not be acceptable or appropriate in all situations. Further study of this type of approach may provide acceptable solutions.

A third solution to this problem is to use an electronic token to request a turn to speak. This is similar to a physical meeting token or formal speaker list that is commonly used to formalize turn-taking in face-to-face meetings. A user would request the token indicating that she wants a turn to speak, have her turn and then release the token when finished. There must be an override mechanism, perhaps controlled by the meeting chairperson, so that a person does not take complete control of the meeting or the floor. The electronic token can keep circulating until the communication session is complete. There has been little research on the effectiveness of this approach to formalised turn-taking and the acceptance by hearing and deaf people of such a method for video conferenced meetings.

Confidentiality

There are important and unique confidentiality and ownership issues arise with video conferencing. For example, questions such as whether recording a video conference violates confidentiality rules, who owns the archive, and who can access the archive remain unanswered. Similar to audio recording practices, permission to record the audio/visual proceedings should always be sought from participants. This permission should address the ownership and access issues. Legal advice is warranted if archiving video conferences is a normal procedure for an organisation.

Special Considerations for managing multipoint or multi-application conferencing

The user will have the opportunity to connect such equipment as a document projector and/or a scan converter that allow people to present paper-based visual materials through the video conferencing.

Video conferencing equipment such as Polycom allows the user to switch between different presentation technologies while in session. For example, the user can switch and allow the users at the remote locations to see a PowerPoint presentation or videotape on the full screen while presenting. For the hearing presenter/participant, there are no barriers as the hearing presenter can continue to speak and the users can see the PowerPoint presentation simultaneously. For deaf users deaf or hard of hearing users must read the PowerPoint presentation first and then return to the presenter or the interpreter being on the screen. This can cause disruptions (and hence delays) in the flow of the presentation and requires considerable mental effort on the part of the deaf participant who must then remember what was on the visual display.

The presenter should be aware of the difficulties of presenting additional visual materials simultaneously with discussion and provide visual materials to participants before the video conferencing or prepare the presentation to account for participant's needs to keep switching the view screen between the visual material and the interpreter or speaker. For example, the presenter should constantly repeat verbally points on visual material being discussed or allow time for participants to take notes or copy the material on the slides.

Displaying a PowerPoint presentation on the same image as the presenter is worse because it can be very difficult for the users at the remote locations to see the PowerPoint presentation (it is usually too small or the projection screen in the local location is poorly lit).

Other factors

The need to constantly monitor the auditory "goings-on" of the physical environment and then decide what is important to communicate to the person who is deaf is an important secondary task for the interpreter. For example, an interpreter might ignore a pencil that falls on the ground because it does not have any consequence for the communication transaction in a face-to-face meeting. However, when the interpreter is remote the physical context is limited by what can be viewed though the camera. It is difficult for the interpreter to determine whether an unseen audio event such as a door shutting outside the view of the camera is important or not. The interpreter must be hyper-vigilant and constantly assess the importance of non-speech and unseen audio events increasing the already high cognitive load and corresponding fatigue levels for the interpreter. Frequent breaks (e.g., every 30 minutes) or more than one interpreter may be required to accommodate these increased cognitive demands.

One final consideration for video conferencing situations is that they can include multiple applications. Many video conferencing technologies allow you to incorporate presentations (e.g. PowerPoint, videotapes (VCR, DVD), a drawing application to share work among all participants and other visual mediums) If only one viewing screen is available, the video conferencing software allocates the viewing priority to the application. The hearing participants can talk over the display and hear what is being said. However, in situations where the interpreter is located with the hearing participants or remote from all participants, the deaf person cannot participate because the interpreter's video is replaced by the application images. In this situation, the use of two viewing screens, one dedicated to the interpreter and the second one for other images is required, or the shared work must be paper-based. Users might have the opportunity to connect such equipment as "Elmo" and or a scan converter to allow user to use different technologies through the videoconferencing to provide more visual information.

Videoconferencing equipment such as that available through Polycom allows the user to switch between the mediums while a session is happening. For example, the user can switch and allow the users at the remote locations to see a PowerPoint presentation or videotape on the full screen while presenting. For the hearing presenter, this results in no barriers as the hearing presenter can continue to speak and the users can see the PowerPoint presentation. For the deaf users it requires more time as the presenter must allow the deaf or hard of hearing users to read the PowerPoint presentation and then go back to the presenter or the interpreter being on the screen. Using a PowerPoint presentation with the presenter through the videoconferencing camera results in less than optimal viewing because it is very difficult for the users at the remote locations to see the PowerPoint presentation. Ultimately the ideal setting would be allowing sufficient viewing times between the mediums to allow the users to read the content. Ideally it would be an excellent idea to send copies of your presentation (e.g. PowerPoint) to the remote users and interpreters. Table 3 summarises the most common behavioural, etiquette and communication issues experienced by hearing and deaf people during video conference sessions.

Issue	Hearing People	Deaf People
Language choice of meetings and issues of language power	• Hearing people who know sign language still communicate in spoken language and use sign language interpreters.	• Culturally deaf people are still expected to be bilingual and follow written (English) documents shared in the meeting.
Introductions	• Often introduce themselves each time they speak as many female voices sound alike.	• It is obvious who is signing, but if deaf people announce themselves before signing, it will make their participation in the meeting equal to their hearing counterparts.
Eye Contact	• May tolerate the lack of eye contact with the people at the far site, provided there is good audio	• An important part of the visual language and severing eye contact means the communication has ceased. • Sitting back from the viewing screen (approximately 244 cm or 8' from a 32" monitor) simulates the most natural eye contact.
Chairperson (deaf or hearing)	• Must ensure that all people are included as equal participants	• Must ensure that all participants are included as equal participants
Use of peripherals	• Can keep talking while PowerPoint slides are being shown	• When interpreters and deaf people are showing PowerPoint slides in a videoconference, there must be pause between each slide and then interpreter can resume watching the signing. • Communication will not be accessible if the hearing person continues to talk over the PowerPoint slides.

Table 4: Behavioural, communication and etiquette issues.

4.3.1.5 Considerations for Remote Sign Language Interpreters

Preparation

An interpreter with many years of interpreting experience will be reasonably comfortable in most interpreting settings. However, the addition of remote interpreting makes every interpreting situation potentially new and anxiety provoking.

Some of the concerns that interpreters will have include:

1. Interpreters want assurances that the people at the remote site will understand their needs and be able to meet them.
2. Whether remote interpreting equipment will be placed in the correct position.
3. Whether people at the remote site will be familiar with how to use an interpreter and the particular dynamics of using a visual interpreter remotely?

To ease some of these concerns, remote interpreters require as much information as possible prior to the interpreting event. Information about the participants, the meeting purpose, agenda items for discussion, intended outcomes of the assignment, and any presentation materials should be provided at least one day ahead of the scheduled video conference.

During the video conference

If deaf people are meeting an interpreter face-to-face for the first time, they tend to ask personal questions such as whether the interpreter's parents are deaf, or where and why they learned sign language to establish a rapport with the interpreter. A remote interpreter may not have this same opportunity to interact with the deaf consumers before the interpreting session, and may be unable to develop that important rapport with those deaf consumers.

In addition, there can be regional differences in sign language that may cause communication difficulties or misunderstandings. For example, a Canadian deaf person may be assigned an American remote interpreter who will use subtle differences in sign language vocabulary. The may cause difficulties for the deaf person in understanding what is being said or expressing herself clearly. It is important to schedule introductory time between the remote interpreter and a deaf consumer ahead of the meeting (at least 10 minutes) to allow these individuals to establish some rapport, and acknowledge and accommodate for regional differences.

The interpreter's role is to facilitate communication between the hearing and deaf participants. During the session, the interpreter might ask for the participants to clarify what is being said because he/she did not hear the content or under the meaning of the content being said.

Processing in public

While the interpreter is voicing for the deaf consumer in a face-to-face situation, it is assumed that the other participants are focused on the deaf speaker and the hearing people are only listening to the interpreter. In a video conference the remote interpreting may be the centre of attention because the remote interpreter is seen on the viewing screen. The interpreter must perform the complex cognitive task of interpreting while being in the unusual position of being the centre of attention. This potentially adds discomfort and cognitive load to the already considerable cognitive load from interpreting for a remote sign language interpreter.

Lag Time

During any interpreting assignment there is a lag, or delay, from the time the source message from the deaf participant is presented, to when the interpreter understands the message and presents it in the target language. This lag may be longer when videoconferencing technology is used due to the delay introduced by the technology. The ability of the deaf participant to participate on an equal level is restricted. Participants need to know to expect these delays and account for them meetings. Section 3.4.1.6 provides some suggested solutions to assist in managing the turn-taking difficulties imposed by the increased lag time. In addition, we suggest that a set of guidelines should be presented to all participants before the conference begins in order to educate people on the communication differences and the procedures put in place to accommodate these differences. The meeting chair or the access manager should be responsible for disseminating these guidelines.

Interpreters, deaf people and hearing people at several sites

Where there are several deaf people and interpreters for one video conference, the arrangement of cameras, viewing screens and seating becomes more complex. If several deaf people are participating in a multi-point call, they may have interpreters present with them at their site. In that case, the interpreter at that site voices and the information is re-signed by the interpreter at the other sites. The deaf people then have the choice of watching the deaf participant directly, or watching the "shadowed" version provided by the interpreter in person at their site. In a multi-point or continuous presence call, the frame speed of each square is reduced enough so as to increase the difficulty of the interpreting task. Having deaf people provide an interpreter at their site reduces some of this difficulty.

Interrupting the speaker (either deaf or hearing)

In person, there are opportunities for the interpreter and the deaf consumer to exchange brief signs that reinforce the accuracy of parts of the interpretation. For example, if a deaf person was talking about their family, an interpreter might check briefly, "Do you have FOUR brothers?" The deaf person could communicate the answer to the interpreter without necessarily stopping the natural flow of the communication. This checking of the accuracy of the initial interpretation might not be obvious to the audience listening to the voice interpretation. When interpreting remotely, it can be difficult for the interpreter to "check in" with the deaf consumer and interrupting becomes more complicated. The interpreter needs regular access to the consumers and the meeting chairperson so she can easily stop the proceedings in order to get clarification when needed. However, even if this occurs, the subtle accuracy "checking" becomes more conspicuous to the other participants and may cause disruption.

Preparation materials provided to the interpreter ahead of time can significantly offset this need for interrupting the speaker. As long as the interpreter has the proper background materials ahead of time, the need for interrupting the flow of the communication can be minimized.

Reduced Signing Space

An additional challenge to remote interpreters is they must adjust their natural signing space to accommodate the field of view of the camera. They cannot sign outside of a prescribed area that falls within the camera's visual field otherwise the deaf participant cannot see what is being said. Reducing the natural signing space can cause additional cognitive load and fatigue for the interpreter. More frequent breaks or team interpreting may be required to relieve this additional strain on the interpreter.

Auditory Referencing

Interpreters who are at the same location as their deaf and hearing consumers have the luxury of looking at exactly what is being referenced in the communication. Interpreters cannot sign the spatial relationship if they cannot see what is being referenced. For example, if a speaker says, "You put this over here," the interpreter will be at a loss as to how to describe what is being discussed. It is more effective for deaf and hearing participants to be explicit in what they are referencing such as, "Your name goes in the blue box in the right-hand corner of the first page."

Team Interpreting

Team interpreting is difficult to successfully carry out remotely without audio headsets that allow the interpreters to communicate together without disrupting the rest of the viewing audience. It is hard for the interpreters to confer on best linguistic choices (either for sign language or English) without individual headsets and microphones. If one of the team interpreters is at a remote setting, their audio is transmitted as soon as they offer linguistic suggestion. It is likely to be perceived as an auditory or visual distraction to the other participants.

Deaf Interpreter

The Certified Deaf Interpreter (CDI) is an individual who is deaf or hard of hearing. In addition to proficient communication skill and general interpreter training, the CDI has specialized training and/or experience in the use of gesture, mime, props, drawings and other tools to enhance communication. The CDI has knowledge and understanding of deafness, the Deaf community, and Deaf culture. The CDI possesses native or near-native fluency in American Sign Language (Registry of Deaf Interpreters, 2004).

There are some unique considerations for video remote interpretations that are applicable to Deaf interpreters. These include:

a. Negotiate with consumer(s) to create working conditions that will facilitate the most accurate and comfortable delivery of interpreting services
b. Inform consumers (hearing and deaf) of any problems the video conferencing session and make efforts to correct them. In this situation, the deaf interpreter may require more time to facilitate communication between the deaf person and the sign language interpreter. Depending on the communication needs of the consumer, the amount of time needed for effective communication may actually double.
c. Communicate with team member(s), particularly at the beginning of the video conference, to assess effectiveness of the interpreting.

d. At the completion of the assignment, it is critical that the deaf interpreter inform the consumers (hearing and deaf) about the clarity of the assignment.

4.3.1.6 Skills Needed for Remote Interpreters

There are very few documents describing specific interpreter skills required for successful remote interpreting. However, where there is mention of remote interpreting, there is agreement that remote interpreters must have considerable experience as face-to-face interpreters. Novices or new graduates of interpreter training programs may find remote interpreting extremely difficult. The types of skills that are learned through experience that are particularly transferable to remote interpreting are:

1. ***Closure skills*** – Interpreters must have the vocabulary, in English and in Sign Language, to appropriately complete sentences. An example of this might be a deaf person telling a story about going to an office to drop off their resume and having to leave it with a receptionist. The interpreter might miss the actual fingerspelling of the word "r-e-c-e-p-t-i-o-n-i-s-t" due to the time lag or pixelation caused by the video conferencing system but can still use their interpolation and closure skills to make a meaningful sentence.
2. ***Assertiveness*** – The remote interpreter must be willing to interrupt proceedings that are preventing participation by the deaf person for whatever reason, particularly if the chairperson is unable to manage the turn-taking and time lag issues. However, this is an added task for the interpreter and is not ideal. The interpreter may want to have a discussion with the chair person regarding some of the issues and solutions to turn-taking and accommodating the time delay to ensure equal access by the deaf person. Better yet, the interpreter can carry a set of guidelines such as the ones included in this document to provide to inexperienced meeting chairs.
3. ***Memory*** – The interpreter must track new, incoming information as well as the gist of what has just been interpreted. This is required because there are often ambiguous references made to previous utterance (such as to a previous PowerPoint slide that is no longer on screen) or location indicators such as "I put that document over here". In addition, remote interpreting tends to have more communication mis-understandings than face-to-face interpreting so the interpreter must have an idea of where the communication breakdown may have occurred, and where to go back to amend or correct the interpretation.
4. ***Lexical choices*** –The interpreter must have several lexical choices for the same concept, as some signs will be more conducive to a 2-dimensional medium than others. Experienced interpreters will have gained sufficient alternative expressions to accommodate and made lexical adjustments to accommodate this situation.
5. ***Varied Experiences*** – Interpreters with various kinds of community interpreting experiences will have the ability to interpret for a wide variety of consumers (both deaf and hearing). This is particularly important for remote interpreting because the likelihood of interpreting for people outside of the interpreters'

regular community is high. For example, an interpreter from Canada may be interpreting remotely for someone in the southern US.

Other factors to consider

1. The optimal background colour is a solid blue or black.
2. The interpreter's attire needs to be a solid blue or black colour as well.
3. Excessive movement is sometimes difficult for the camera to capture. It is best if the interpreter stays in one place.
4. A highly skilled interpreter is preferable, but he should be able to adjust his signing rate to account for less optimal camera frame rates.
5. The interpreter should arrive at least 15 minutes prior to the meeting. If it is an interpreter's first exposure to video conferencing (after training), it is recommended that the interpreter arrive at least 30 minutes early. The deaf person should also arrive 15 minutes prior to the meeting.

4.4 Illustrative cases and personal accounts highlighting issues

This section provides a short description of several cases and personal accounts illustrating the common issues using video conferencing for people with disabilities. Emphasis on experiences of deaf consumers is made due to the complications of language translation and the importance of high quality video.

The first case is of a family with a hearing father and deaf mother gathered to discuss parenting strategies with the hearing staff of a child protection agency in order to determine whether the child could stay in the home. The interpreter was connected remotely through a video conferencing system and all other participants were face-to-face. There were no remote camera controls so that the interpreter could not control the camera in the meeting room.

The mother was very upset and crying. As a result, she could not pay attention to the viewing screen, and the interpreter had difficulties gaining the attention of the other meeting members to assert the mother's needs. Also, the staff was seated with their backs to the viewing screen and facing the father, so that they could not see the interpreter's attempts to gain their attention. The interpreter was trying to accurately and intelligently represent the mother's communication but because the mother was so upset there was considerable delay in processing the mother's communication. This only added to the delay imposed by the technology. In addition, the hearing people continued to give new information and talk while the mother was trying to communicate and the interpreter was trying to interpret that communication and intervene on behalf of the mother. While having to process language as well as assert the deaf person's needs are normal responsibilities of the interpreter, the seating arrangement, the lack of support for and recognition of the mother in the physical setting, the lack of turn-taking structures and the lack of acknowledgement of the interpreter's importance resulted in a failed meeting. Many of these difficulties arose because the interpreter was remote and had difficulty gaining attention and recognition.

While this was a difficult situation even for a face-to-face meeting, the situation was exacerbated by the difficulties resulting from the remote video interpretation issues. To overcome some of these difficulties, there are a number of possible solutions. A formal attention-getting mechanism such as a flashing light with audio would have likely aided in allowing the interpreter and deaf mother to have more presence in the meeting. The hearing staff needed to be seated in such a way as to see the interpreter's eye gaze and turn-taking indicators. In addition, the staff did not have much experience with video conferencing technology, and did inadvertent things such as placing coffee cups down near the microphone causing audio spikes that were very disconcerting for the interpreter. Camera controls or properly setup camera view angles that suited the needs of the interpreter were required. An access manager would have greatly assisted in helping facilitate appropriate setup and use of the technology, and remove some of the barriers to successful communication.

The next three cases are first-person accounts from three individuals directly involved in providing video conferencing services at CHS; two are remote video interpreters, and one is a technical support person at CHS.

4.4.1 Personal Account: My First Impressions of Videoconferencing

By Video Remote Interpreter 1

Question**: What was your first impression of video conferencing? As a consumer?**

Answer: *"The first time I ever interpreted remotely, the deaf consumer copied everything I was signing as if he was watching a teleprompter. How many times does your television ever talk to you? It's a difficult concept for people to get their heads around. Those consumers whom I had met before seemed to pick up the concept faster. If the deaf consumer has never met the interpreter they may not know that the person signing to them on the television screen is actually expecting a response; that the process is interactive."*

Question: **What were your first experiences as a remote video interpreter?**

Answer: *"Sign Language Interpreters work hard at improving their English and Sign Language skills and be recognized as competent professionals. To impose technology that makes the interpretation feel more difficult and stilted feels like a step backward. Will people know that I'm not "new"? Will they see me as an interpreter with experience? As a "seasoned interpreter," why should I put myself in a position to feel inadequate when they are many easier assignments readily available?"*

4.4.2 Personal Account: My First Impressions of Videoconferencing

By Video Remote Interpreter 2

Question**: What was your first impression of video conferencing? As a consumer?**

Answer: *"The first experience I ever had using videoconferencing was two years ago for a job interview. I knew two of the people at the far site, but I was convinced they were angry with me. The camera had zoomed out in order to fit the three far-site people in the picture at the same time. I couldn't see their facial expressions very well and I felt as if my* *personality was really coming across like a lead balloon. Although it was a very clear connection (384 kbps and 30 fps) there was a slight delay in the amount of time it took for them to understand my questions and respond. The experience made me think that I might not want a job doing a lot of videoconferencing if this is what is was all about. I figured it wasn't much different (or better) than a teleconference call. I had a person with me that ran the remote control, but he wanted to leave for another commitment. I was terrified at the prospect of being left alone with the equipment (particularly in a job interview where you want to appear competent at what you're doing)."*

Question**: What were your first experiences of remote video interpreting? Difficulties experienced?**

Answer: *"My first experience interpreting by video conference was for a deaf consumer that I have known both personally and professionally for many years. I thought this would make it easier, but in fact it "upped the ante" and made me more fearful of making a mistake. I thought if I made a mistake, people would question my interpreting skills and think, "Boy, we thought she'd do a great job with Susan seeing that they've known each other for years!" It provided an extra pressure on top of the already difficult interpreting task. The situation was further complicated by well-intentioned individuals (who knew sign language) who noticed me struggling and yelled out either background context that I was missing, or yelled out the fingerspelling that I had missed, in order to help me. I had to get the deaf person to fingerspell very slowly and to change the orientation of some of the signs she was producing. Although the deaf person was quite willing to accommodate my needs, I felt like I was giving her an extra task to do. She was already concentrating on her thoughts, and then had to remember my needs. Interpreters are used to blending into the background so that the deaf person and hearing person communicate directly with each other. Having to interrupt either the hearing person or the deaf person more often made me feel more conspicuous in the interpreting event than I was accustomed to being."*

4.4.3 Personal Account: Web Applications Specialist

Question**: What was your first impression of video conferencing? As a consumer?**

Answer*: "IT'S ABOUT TIME! Back in 1991 I was very interested in videoconferencing and how we can use this to communicate visually. The equipment back then was using six telephone lines to communicate and the setting up of this was complex. We now have come to modern day videoconferencing using IP or IDSN. The quality of the visual information has gone up considerably, but there is still room for improvement.*

It a new way of communicating and requires new set of rules. The complexity of providing meetings with various communication needs is a challenge. I find myself in a group of deaf individuals who sign only is fine and the only problem is getting people's attention at the other location. Often we forget the person at the other end. There needs to be a way electronically or through human interactions to improve "attention getting" cues in this medium."

Question: What were your first experiences of providing video conferencing services at CHS?

Answer*: "Given the implementation of IP videoconferencing at The Canadian Hearing Society, it has provided us new ways of providing services. We have many deaf, deafened and hard of hearing people in the north who does not have access to a sign language interpreter. By providing the interpreter through videoconferencing, it's breaking down the barrier and at the same time saving costs. I know of individuals in Timmins who are in need of interpreters and we can provide this service from Toronto through videoconferencing. Literacy classes, sign language classes and counseling are other kinds of events that are taking place via videoconferencing. With the videoconferencing, we are breaking down the geographical barriers and being able to provide information in a visual medium."*

These three accounts provide some insight (need for training, pre-conference planning, technology improvements, ways to improve and acknowledge presence of remote person and but that the geographical a significant barriers that can be mediated by vc).

4.5 User Environment Scenarios:

In an effort to better understand and elucidate the issues inherent in different uses of video communication technologies, the following section will examine three key user group environments: health, education, and business, and common tasks used in these environments. For each user group environment, the challenges and issues connected with its unique requirements will be explored, and possible remedies will also be presented.

The scenarios discussed mainly involve one-to-one and small groups. Even though there are many different scenarios and combinations, our recommendations will be based on the most popular scenarios that are commonly presented on daily basis when video remote interpreting is used.

4.5.1 Health care Environments

Health care environment tend to involve limited sets of personnel with limited tasks to be achieved. However, it is very important that people have equal access to the health care environment and are able to successfully communicate. Telemedicine is becoming a much more common approach for providing access to medical services, particularly specialist services, for people in remote communities. It is important to consider the needs of people with disabilities.

4.5.1.1 Tasks

Common tasks that video conference participants may want to carry out in this setting are:

> Consultation – question and answer session regarding particular health needs or issues.
> Examination – inspection of physical, or mental state of patient.

Prescription – discussion of a particular course of treatment, intervention or action.

4.5.1.2 Physical scenarios

A deaf patient would like to consult with medical personnel (e.g., a doctor, nurse, psychiatrists, or medical technician) through an interpreter via video conferencing. The following list includes possible scenarios of remote interpreter/deaf consumer combination.

- The interpreter is in a remote location and patient and doctor, hospital staff or health-care providers at the local location.
- The interpreter is with the deaf person and the doctor, hospital staff or health-care providers are remote.
- Interpreter can be at one location, deaf person at another and doctor, hospital staff or health-care providers at yet another location.

4.5.1.3 Issues unique to health care settings

Physical Positioning

The National Council on Interpreting in Health Care (NCIHC) has produced guidelines for the role of verbal language interpreters working face-to-face in health care settings. In these settings, they recommend that interpreters be positioned near the patient's head with their backs to the patient giving the patient as much privacy as possible. However, sign language interpreters must always have eye contact with the patient for communication to be understood so they must stand in such a place as to see the patient's facial expressions. NCIHC also recommends that interpreters stand between the patient and health care provider so as to appear as unbiased participants.

A sign language interpreter working remotely cannot be guaranteed that busy health care providers in emergency settings will understand the interpreters' positioning preferences or be able to accommodate them at a bedside with limited room to maneuver the required medical equipment.

Sign language production

The ill deaf person lying in bed will not have the same clear sign language production as that of the interpreter, nor will they necessarily maintain a small enough signing space to fit within the camera's field of view. This can be particularly problematic when the interpreter is remote from the deaf person.

One recommendation is to use technical solutions such as zooming out the camera lens to capture their whole signing space and environment. However, the interpreter risks losing critical facial expressions that are achieved by having the camera closer to the deaf person's face. There can be a constant conflict between wanted to see a larger view of the room, and the need to go close to the deaf person to see nuances in the facial expression.

Socio/Political

When remote video interpretation is the method of service delivery, deaf participants may mistrust the interpreter because they have no working relationship with the interpreter and no common cultural context. They may withhold some of their communication. This may also work to the benefit of the deaf person. In a sensitive health care setting, deaf people may say more because they do not know the interpreter personally and chances are will not run into that same interpreter out in the community.

Socio/political factors emerge as some of the most important factors for successful video conferencing and remote interpreting. Failure to account for these factors can result in a failed conference and service to the deaf person. These factors include psychological aspects such as an interpreter's ability to cope with the stress of a remote interpreting assignment, or a deaf person's ability to cope with the stress of having medical needs and the confidence to express these needs via a video conferencing system and/or remote interpreter; having to rely on a screen to derive the visual support information necessary for carrying out the interpreting task; motivation; processing information from multiple sources; social isolation; and operating multiple controls while effectively communicating. For example, in personal statements it has been reported by interpreters and deaf people that the lack of proximity between a deaf person and her interpreter can create a feeling of alienation that may results in lack of motivation, a decrease in interpreting quality, and a decrease in the deaf person's involvement and participation in the video conference.

The deaf person may be at significant disadvantage because of these socio/political factors. In addition, they may feel intimidated by a lack of medical knowledge and and by not having a physical presence at the consultation, meeting or evaluation. This may result in an unbalanced power relationship between medical personnel and the deaf patient that can, in turn, cause withdrawl by the deaf participant, the need for the interpreter to take on a strong advocacy role, and errors made by the medical team by unconfirmed or incorrect assumptions. Awareness, experience and perhaps a patient advocate who is physically present with the medical team may be ways in which some of these difficulties can be mediated.

4.5.1.4 Guidelines and Protocols specific to Health Scenarios

- Follow general guidelines listed in Section 3.0.
- It is recommended that a system with "picture-in-picture" capabilities be used and that deaf consumers and interpreters be instructed to monitor themselves in the "Picture in Picture" screen to ensure that their signing stays within the camera's view. Individuals may need to adjust their signing space to accommodate small camera field of views.
- The deaf consumer at the remote end should have access to the camera controls so that they can ensure that the signs being transmitted by the interpreter. One important consideration is that the deaf person must also have training and a comfort level with the technology in order to manage camera controls, sign adequately and communicate his/her needs in an environment where the deaf person may be feeling sick or enjoy little power in the relationship.
- Video conferencing calls should be recorded for future reference or for legal needs.
- There can be a general need for a mobile video conferencing facility if people do not have the mobility to move to a "special video conferencing room". The recommendations in this document would apply to this type of portable system as well.

- A pre-conference meeting (using the video conferencing system) is recommended to alleviate some of the cultural and technical difficulties that might interfere with the success of the actual conference, and to provide new users with some practice.
- The technology should be tested and adjusted prior to the meeting to ensure that connection numbers, cameras, microphones and displays are working correctly. This also avoids time being spent on audio and video checks during the meeting.
- The interpreter should always remain in direct eye contact with the person who is deaf regardless of the position of equipment or other people.
- Be cognizant of the psychological stress imposed by medical situations and increased potential for the deaf participant to be alienated due to a reduced ability to communicate while ill and over a video conferencing system. It may be too easy for the medical practitioners to inadvertently ignore the deaf person's needs.

4.5.2 Education Environments

Distance learning has become more popular and gained wider acceptance as an effective instructional tool (Gowan & Downs, 1994; Benford et. al., 1998).
Students of all types have begun to take advantage of this services institution are offering. Video mediated communication is a valuable educational resource because it provides access to live instructors or teaching assistants; it can be more motivating than students working on their own, and can help students improve communication skills (Knight, 1998).

When a deaf student is registered at a post-secondary institution, whether it is through distance learning or face-to-face classes, she is often provided with an interpreter. As interpreting services can be difficult to obtain particularly for students at institutions in more remote locations (e.g., University of the Cariboo in northern British Columbia), the possibility of using remote interpretation is very real. Interpretation would be provided by interpreters in major centres that can often be quite a distance from the student.

4.5.2.1 Typical tasks

Common tasks that students (remote or local) may want to carry out in this setting are:

- Listening/pay attention to instructor for extended periods of time (e.g., one hour).
- Interacting with instructor with question/answer or discussion styles of interaction.
- Interacting with peers in formal group meetings, or in side conversations.
- Interrupting instructor to ask question.
- Formal presentation of academic material or point of view for entire class.
- Viewing class material (e.g., PowerPoint presentations, overheads, videos, music, blackboard/whiteboard markings, and images, websites, etc.)

4.5.2.2 Physical scenarios

There are a number of possible physical scenarios in which video conferencing is used to serve deaf students. These apply whether the video conferencing is used in a distanced education or a face-to-face teaching situation.

In the first scenario, the deaf student and her interpreter are in the same location. The interpreter would be signing the material being delivered by the instructor and voicing questions from the deaf student. This scenario is the most straight-forward as the interpreter can remain outside of the field of view of the camera and the viewing screen in the classroom only shows the deaf student. The image in the student's location shows the instructor and/or the classroom depending on the controls provided with the video conferencing system.
A second scenario is where the interpreter and instructor are in the same location and the deaf student is in the remote location. The interpreter is video remote interpreting the material being taught by the instructor while having a physical presence in the classroom. While the deaf student will always appear on the viewing screen, the interpreter potentially will have a stronger presence in the classroom particularly when voicing the deaf student's communication. Maintaining focus on the deaf student is a difficult task for the interpreter particularly when there are many other people in the classroom. The viewing screen is often located at the front of the classroom for the simple reason of access to electrical power. The deaf student is not an integrated part of the class but rather stands out at the front and may believe that everyone is always "watching" him. In addition, the eye gaze position of the deaf student due to the camera location may seem as though the deaf student is otherwise occupied or not paying attention. This may cause the student to take a passive stance and not participate in discussions or queries.

PEBBLES (Weiss & Fels, 2001) is one example of a video conferencing system that allows a remote student to be placed anywhere in a classroom. It is a video conferencing robot designed to allow students in hospital to attend school. Remote students have control over what they can view in their classroom and what others can see of them. Studies with PEBBLES (Fels, 2003) indicate that students using PEBBLES become integrated into the classroom quickly and that they do not receive more attention than other students after an initial novelty period. Placing the video conferencing unit at a normal "seat" in the classroom may remove some of the unwanted focus on the deaf student at the front of the class.

The third scenario would occur when the deaf student is physically attending a course in a classroom and the interpreter is remote. The video conferencing system would be setup in the classroom and transmit a live image and voice of the interpreter (the interpreter would sign what the instructor was saying and voice comments, questions, discussion points signed by the student).

Similar to the second scenario, the video conferencing system would usually be placed at the front of the classroom and potentially become the focus of attention particularly when the interpreter is voicing the deaf student's comments. The interpreter would have very little influence over where people are looking as the interpreter's eye gaze would seem to be towards all people in the classroom. Using a personal video conferencing system in the classroom that can be located close to the deaf student may potentially alleviate some of these problems including focusing the attention on the deaf student when he wants the floor.

4.5.2.3 Issues unique to education settings

Turn-taking

In classroom situations, turn-taking tends to be more formalized where students are expected to raise their hands to indicate they want the floor. Integrating a turn-taking indicator such as a hand or an audible light may be an appropriate method for the deaf student to also indicate she wants the floor. However, using only a light or flashing light has been shown to be ineffective in classroom settings (Weiss et al, 2001).

Shared applications

Where PowerPoint presentations or other applications must be shared, dual monitors must be used to ensure that the interpreter can always be seen by the deaf student. If captioning is available for audio material such as videos and films that have sound or music, it should be activated to ease some of the burden on the interpreter. If no captions are available for video or music material, the education delivery unit should consider producing them. This is important for live access to this material during use for the class but it is also important to allow students access to the material after class when an interpreter is not available.

Other important factors

Finally, constant attention to a video screen is fatiguing for deaf students as they rely exclusively on their visual system to acquire communication messages. Hearing students can take small visual breaks by using their auditory system during class. More frequent formal breaks, at least once per hour, are required to allow the deaf student to rest.

4.5.2.4 Guidelines and Recommendations specific to education

- Follow general guidelines for listed in Section 3.0
- It is recommended that a system with "picture-in-picture" capabilities be used and that deaf consumers and interpreters be instructed to monitor themselves in the "Picture in Picture" screen to ensure that their signing stays within the camera's view. Individuals may need to adjust their signing space to accommodate small camera field of views.
- The deaf consumer at the remote end should have access to the camera controls so that they can ensure that the signs being transmitted by the interpreter. One important consideration is that the deaf person must also have training and a comfort level with the technology in order to manage camera controls, sign and communicate adequately.
- As the instructor usually is responsible for managing class activities, pre-conference training is highly recommended. The instructor must be made aware of the unique communication and turn-taking needs of the deaf person/interpreter pair, and given strategies for ensuring that these needs are met.
- The technology should be tested and adjusted prior to the first class to ensure that connection numbers, cameras, microphones and displays are working correctly. This

also avoids time being spent on audio and video checks during the class and possible missed information from the instructor.

- The interpreter should always remain in direct eye contact with the person who is deaf regardless of the position of the instructor, other people or equipment in the classroom.
- Employ a formal turn-taking mechanism such as a remote hand that allows the deaf person to overtly gain the attention of the instructor or class peers.
- All class materials should be available to the deaf person and the instructor before the class begins.
- Any video or music material used in class should be captioned, and the captions turned on during viewing of the material in class.

4.5.3 Business Environments

One of the most common applications of video conferencing with deaf participants is a business meeting. These can take several forms including one-on-one, small groups of less than fifteen individuals, and large meeting with many people (e.g., annual general meeting). In each setting, there are unique issues that must be addressed so that the deaf person can enjoy equal participation and the interpreter is optimally positioned to facilitate this participation.

4.5.3.1 Tasks

Common tasks that video conference participants may want to carry out in this setting are:

- Participate in the discussion, contributing comments and listening to what others contribute. This can be accomplished using formal, agenda-driven rules for turn-taking and meeting progress (e.g., Robert's rules of order) or much more informal processes where people raise hands to request the floor or use non-verbal cues to take a turn when the opportunity arises. Large group meetings tend to be much more formal with set agendas and speakers, formal question/answer or discussion periods and formal turn-taking structures.
- Proposing or seconding motions.
- Presenting information or materials where participant has formal and exclusive control of the floor.
- Asking/answering questions/gaining attention
- Chairing meeting

4.5.3.2 Scenarios

The following three scenarios of deaf/interpreter pairs are possible

One to One Meetings

- The deaf person and the interpreter are in the remote location and the hearing person is at the local location or vice-versa.
- The deaf person and the hearing person are in the local location while the interpreter is at the remote location.

Small Group Meeting

- The group of deaf people and interpreter are at the local location while the hearing person/people are at the remote location.
- The deaf person is at the remote location and the hearing people and interpreter is at the local location.
- Deaf and hearing people at local location and the interpreter is at remote location.

Large meetings

- Group of deaf and hearing people are at the local location while the interpreter is at the remote location.
- Group of deaf and hearing people as well as the interpreter is at the local location while at the remote location there would be a small group of hearing and/or deaf people.

4.5.3.3 Issues unique to meeting settings

Turn-taking

Turn-taking issues are likely the more important issue to consider in meeting with deaf people. Equal participation in video conferenced meetings often involves equal access and particular sensitivity to turn-taking issues. Many of the issues discussed in Section 3.1.4.6 apply in business meeting scenarios.

It is very important that turn-taking mechanisms be explicitly addressed and formalised in meetings with deaf participants to ensure that people are not isolated from discussions and can contribute equally. This is the responsibility of the meeting chair and remote interpreter (and/or the access manager) even when technological solutions are provided (e.g., a turn-taking indicator is used).

Visual materials

Business meetings like education scenarios often involve the use of visual presentation materials. As discussed in section 5.2.3.2 and 3.1.4.8, there are particular difficulties in using visual materials and speaking about those materials simultaneously. Large group meetings are often planned well in advance of the meeting. Providing visual materials to participants ahead of the meeting is the usual practice. One-on-one meetings often do not involve the use of other visual materials.

Small group meetings are often most affected by use of ad hoc or undistributed visual materials. In order to best manage the needs of deaf users for access to the visual materials and the interpreter through a single view screen, the following suggestions are made:

1) Use a fax, internet connection or file transfer capabilities on the video conferencing system to supply the deaf participant and the interpreter with visual materials during the meeting. The meeting may need to pause while this is carried out but can it can then resume at a faster pace once all people have access to the visual materials.
2) For a pre-planned meeting, the meeting organizer should request visual materials be made available prior to the meeting or suggest that they not be used.

3) The presenter should repeat/read each point verbally before discussing it as well as explicitly announce when a page/slide should be turned to the next one. This will be useful for blind and deaf participants.

Socio/political

The social/political issues that arise with for the health care environment are also potential issues in business meetings. Section 5.1.3.3 under Socio/political health provides details of those issues.

Technology alternatives

- Desktop computer videoconferencing equipment can be used for one to one videoconferencing.
- Small groups require boardroom style set up utilizing videoconferencing equipment that allows camera movement via the remote control.
- Small groups require boardroom style set up utilizing videoconferencing equipment that allows camera movement via the remote control.

4.5.3.4 Guidelines and Recommendations specific to meetings

- Deaf consumers and interpreters need to watch themselves in the "My Video" or "Picture in Picture" screen to be sure that they are signing within the camera's view. Deaf people might have to sign in a smaller space than they normally do, so they are not signing off screen.
- The deaf consumer at the other end should have the control to zoom in/out, up/down to be able to get the signs being transmitted by the interpreter.
- For large and small group meetings, a meeting chair should be assigned. The chair is responsible for ensuring that participants with disabilities and interpreters are explicitly included and meeting procedures address their specific access needs. Planning this prior to the meeting ensures that there are no meeting delays because these issues have not been addressed.
- It is ideal to have a separate access manager from the meeting chair who is responsible for the operation of the technology and for ensuring that access needs are met.
- Video conferencing calls should be recorded and archived for future references and as video/audio notes. Confidentiality issues should be taken into account.
- Visual materials should be made available to deaf participants and interpreters at least one day prior to the meeting.
- Turn-taking procedures should follow formal turn-taking rules. The meeting chair must be responsible for the management of turn-taking.
- Training sessions for video conferencing should include training and suggestions to address access issues.

4.6 Live Captioning: An Overview

Real-time or live captioning is a technique that is not currently available in video conferencing technologies and that could provide many benefits to users with disabilities.

Real-time captioning involves the translation of dialog or spoken information into text as it is occurring similar using court stenographic procedures (court stenographer and technology). A verbatim text transcript is provided to conference participants as participant dialog is being produced. Not only does this allow participation by hard of hearing participants and some deaf participants who use spoken language, but also it allows a complete text archive of the meeting to be produced while the meeting is in progress. The text stream could be displayed along with the camera images on the view screen or it could be displayed over parallel technologies such as through a website.
Services already exist that provide real-time captioning through remote connections. The real-time captionist does not need to physically attend the meeting or class but can use a remote audio/video connection to "listen" to the meeting, and produce and display the real-time captions for meeting participants.

Some of the difficulties with real-time captioning relate to accuracy levels (or error rates during to typing inaccuracies, poor audio fidelity, and mis-hearing words), interpreter's ability to physically and cognitively cope with speed of conversation, interpreter fatigue, and ways to easily provide emotive or speaker identification information. Many of these difficulties remain unresolved and additional research and development is required to address them.

4.7 Recommendations and Guidelines – Summary

- *Recommended minimum bandwidth is 384kbps for a typical business quality videoconference.*
- Good quality videoconference requires 30 frames per second video transmission.
- Data rate higher than or equal to 384 kbps will support a video screen update of 30 frames per second, equal to VCR playback quality television.
- *Data rate lower than 384 kbps will support video screen update of 15 frames per second or less, which is usable, but can seem jerky under rapid motion.*
- Video transmission frame rate greater than 15 frames per second is critical.
- Audio is less important (it is still important for the interpreter), than video, so need for camera controls (remote and local) is very critical.
- Personal microphone such as a lavaliere or lapel microphone for the interpreter rather than a high-end room microphone can be used.
- If interpreter is with the hearing participants, a room microphone and a personal microphone may be required.
- Intelligent audio signal processing (in the form of intelligent hardware or software mixer) must be available to switch between the various audio sources (e.g., interpreter voicing and meeting participants).
- Microphone should be placed next to interpreter voicing for deaf person.
- Remote interpreters should only work for three hours per day, since video interpreting is more fatiguing than face-to-face interpreting.
- Maintaining eye contact is crucial for deaf people in video conferencing.
- Sitting back from the viewing screen (approximately 244 cm or 8' from a 32" monitor) simulates the most natural eye contact.

- Interpreter should position herself on camera by sitting far enough back or by zooming out to appear to have eye contact with the deaf person .
- The interpreter should always remain in direct eye contact with the person who is deaf regardless of the position of equipment or other people.
- Interpreters should avoid eye contact with a hearing person as a non-verbal reminder that he should direct his remarks to the deaf person.
- If interpreter is remote to all parties, he must view screen to watch the person who is deaf.
- To eliminate "visual clutter" in the camera's field of view, minimize visible items on the table such as flowers, water bottles, computer screens,
- Table 1 suggests appropriate seating arrangements for all participants.
- An office divider in a neutral colour makes an appropriate backdrop.
- Walls in the video conferencing room should be painted a solid neutral colour (blue and green are also acceptable).
- The environment must have minimal "visual noise" such as windows, wall coverings, logos, murals or any thing or person that can distract users.
- The optimal background colour is a solid blue or black.
- The interpreter's attire needs to be a solid blue or black colour as well and should contrast with the skin colour.
- Excessive movement is sometimes difficult for the camera to capture. It is best if the interpreter stays in one place.
- Tables with a matte finish are recommended to reduce additional glare**.**
- Depending on the size of the sessions, tables should be portable (eg: wheels on legs, or detachable legs) to allow optimal set up.
- Curtains are recommended to allow the least amount of visual distraction.
- Chairs should not be "squeaky" when rocked back and forth. Casters on casters are most appropriate for tiled floors to prevent scraping sounds when the chairs are moved. The ideal floor covering is carpet with under padding to minimize the chair moving sounds.
- Adequate room lighting ensures productive communication. Optimal lighting should illuminate the user's face and hands directly and minimizes shadows on face and hands.
- The interpreter must be hyper-vigilant and constantly assess the importance of non-speech and unseen audio events
- If a presentation is incorporated in video conferencing one viewing screen should be for the interpreter and a second one for the other images.
- Chairperson must ensure that all participants are included equally.
- When interpreters and deaf people are using PowerPoint in video conferencing, there must be a pause between each slide and then interpreter can resume watching the signing. Communication will not be accessible if the hearing person continues to talk over the PowerPoint.
- If deaf participants announce themselves before signing; it will make their participation in the meeting more equal to their hearing counterparts.

- Information about participants, meeting purpose, agenda for discussion, intended outcomes of assignment, and any other materials should be provided at least one day ahead of the scheduled video conference.
- Scheduled introductory time between remote interpreter and deaf consumer ahead of the meeting (at least 10 minutes) is needed to allow these individuals to establish rapport, and acknowledge and accommodate for regional differences.
- Arrival of interpreter should be at least 15 minutes prior to the meeting, but if it's a first time it should be ½ hour earlier. The client (deaf person) should arrive 15 minutes prior to the meeting as well.
- A set of guidelines should be presented to all participants before the conference to educate them on the communication differences and the procedures of video remote interpreting.
- The interpreter requires regular access to consumers and meeting chairperson to easily stop the proceedings to get clarification if needed.
- Remote interpreters should be experienced face-to-face interpreters.
- Interpreters must have the vocabulary, in English and in Sign Language, to appropriately complete sentences.
- Remote interpreter must be willing to interrupt proceedings that are preventing participation by deaf person for any reason, particularly if the chairperson is unable to manage the turn-taking and time lag issues.
- Interpreters must track incoming information as well as the gist of what has just been interpreted.
- The interpreter must have several lexical choices for the same concept, as some signs will be more conducive to a 2-dimensional medium.
- Interpreters with diverse community interpreting experiences will have the ability to interpret for a wide variety of consumers (both deaf and hearing).
- The interpreter should be able to accommodate signing slower.
- To process language as well as assert the deaf person's needs are normal responsibilities of the interpreter
- System with "picture-in-picture" capabilities should be used and deaf consumers and interpreters be instructed to monitor themselves on the "picture-in-picture" screen to ensure their signing is within camera's view.
- The deaf consumer at the remote end should have access to the camera controls to better view the signs being transmitted by the interpreter.
- Consider a general need for a mobile video conferencing facility if people do not have the mobility to move to a "special video conferencing room".
- A pre-conference meeting (using video conferencing system) can ease cultural and technical difficulties that might interfere with actual conference, and provide new users with practice.
- The technology should be tested and adjusted prior to meeting to ensure connection numbers, cameras, microphones and displays are in order. This also avoids time spent on audio and video checks during meeting.
- Medical situations can impose Psychological stress and increase potential for deaf participant to be alienated due to a reduced ability to communicate while ill and

over a video conferencing system. The medical practitioners may inadvertently ignore the deaf person's needs.

- Interpreters should stand between the patient and health care provider to appear as unbiased participants.
- Pre-conference training is highly recommended for a classroom instructor and he must be made aware of the unique communication and turn-taking needs of the deaf person/interpreter pair, and given strategies for ensuring that these needs are met.
- Employ a formal turn-taking mechanism such as a remote hand that allows deaf person to overtly gain attention of the instructor or class peers.
- All class materials should be available to the deaf person and the instructor before the class begins.
- Any video or music material used in class should be captioned, and the captions turned on during viewing of the material in class.
- Desktop computer videoconferencing equipment can be used for one to one videoconferencing.
- Small groups require boardroom style set up utilizing videoconferencing equipment that allows camera movement via the remote control.
- The deaf consumer at the other end should have the ability to zoom in/out, up/down to get the signs being transmitted by the interpreter.
- Video conferencing calls can be recorded for any future legal issues as well as being able to go through notes in case he/she feels that they have missed something.
- Instructors are required to allow deaf student to rest at least once a hour.
- A traveling "Road Show" that would show the videoconferencing equipment to deaf individuals and explain its impact on service delivery. Deaf people in remote areas will not know that it is possible for the television to communicate with them. More education and exposure to this technology will encourage cultural acceptance.
- Agencies that do a lot of remote interpreting should invest in "studio quality" resources: the proper videoconferencing furniture, backdrop colours and lighting.
- A separate certification process for remote video interpreters should be investigated so that deaf and hearing consumers will have faith in the professionals providing service to them in this medium.
- Headsets, laptops should be considered natural parts of the remote interpreting process and should have resources designated to that purpose only.

4.8 Glossary of terms

Frames per second (fps): number of frames that pass by per second. NTSC (north American) Broadcast quality video is 29.95 fps

H.323 The ITU standard for videoconferencing over packet switched networks

Acknowledgements

Canadian Network for Inclusive Cultural Exchange (CNICE)

Creating Accessible Online Cultural Content Discussion Document Series

CNICE Partners

Adaptive Technology Resource Centre
Banff New Media Institute
Canadian Abilities Foundation
Canadian Centre on Disability Studies
Canadian Cultural Society of the Deaf
Canadian Hearing Society
Canadian Learning Television
Centre for Learning Technology
Centre for Research in Disability Studies
Digital Frog International
marblemedia
National Broadcast Reading Service

Editors and Contributors:

Michael Boyce, Sara Diamond, Deborah Fels, Elle Gadsby, Rob Harvie, Wendy Porch, Lake Porter, Jan Richards, Vera Roberts, Marc Rosen, Nancy Sicchia, Charles Silverman, Jutta Treviranus

This project was made possible through the funding by the New Media Research Networks Fund of the Canadian Heritage Canadian Culture Online Programme.

Canadian Heritage Patrimoine canadien

5 Representations of Visual Geo-Spatial Information

5.1 Introduction

There is a strong visual bias to web content. This document aims to provide some solutions to barriers in accessing online geo-spatial content, particularly to those who are blind or have low vision. It discusses sensory modality translation techniques, considerations and guidelines, resources, and the design of recommended digital media formats capable of conveying spatial information. Some of the content involves relatively new technologies, such as tactile interfaces. It is hoped that government on-line content designers can benefit from this document, and in turn, a larger spectrum of their clientele.

This discussion is primarily about tactile and audio cues in electronic map and web3D environments: how to enhance detail, give context, and ease navigation and way-finding. It bootstraps off of previous work done at the Adaptive Technology Resource Centre of the University of Toronto. The CANARIE Health Education Network project *Adding Feeling, Audio and Equal Access to Distance Education* had the technical objectives of facilitating the creation of accessible environments, and included a model courseware module designed to assist in teaching geography to students who are blind. Major research influences on this document were:

- Carleton University's Drs. Taylor and Siekierska *et al* work on audio-tactile mapping on the web with voice annotations and haptic feedback, to be found on the Natural Resources Canada site http://tactile.nrcan.gc.ca.
- Drs. Bulatov and Gardner at Oregon State's Science Access Project, who gave several new paradigms for tactile information display.
- Dr. Calle Sjöström at Certec, Lund Institute of Technology, Sweden, on non-visual haptic interaction design.
- Dr. Norm Vinson of the National Research Council's Institute for Information Technology, who has kindly provided his research on navigation in virtual environments.

This document starts with the more familiar, computer-based 2D geo-spatial context, usually in the form maps and explores transformations from visual media to audio modes, progressing to the sense of touch. The on-line graphical format Scalable Vector Graphics (SVG) is a primary focus, with resources and application techniques provided. The Web-based 3D technologies of Virtual Reality Modeling Language (VRML) and its offspring X3D are in turn explored, with more resources and application techniques on the issues.

Discussion of sensory modality complementarity and supplementary guidelines on designing for navigation and way-finding in virtual environments wrap up the document.

5.2 2D Content

Online visuals no longer entail simple static imagery. The complexities of information display call for more intuitive ways of accessing meaning through alternate or concurrent channels. Most online geo-spatial content takes the form of maps, which are not only essential but extremely valuable tools. Digital maps on the Web are often provided through increasingly flexible Geographic Information Systems (GIS) and allow people from different sectors and disciplines to make and use maps to analyze data in sophisticated ways.

5.2.1 Translations: Visual to Sound

In cartographic documents symbols are defined as "the graphic elements shown on a map designed to represent geographic features or communicate a message, at a given scale" (*The Atlas of Canada*, Natural Resources, atlas.gc.ca/site/english/learning_resources/carto). Map symbols can be pictorial or abstract in nature. Similarly aural symbols can either be metaphorical or abstract in nature. In a cartographic environment there is also the possibility of providing the user with sound actually recorded at the site being represented. While this possibility remains untested, possible uses for this type of representation are worthwhile. In this section text will also be discussed as a possibility for representing elements of a cartographic document.

Aural Text (Voices)

Textual descriptions of objects are the most intuitive way to describe elements of a map. For example, having a voice say "school" easily and informatively does the task of identifying a school symbol. Also, aural text can easily provide additional information about the school without a steep learning curve from the user (provided the user speaks the language being spoken in the aural text).

There are two ways aural text can be integrated into a user interface. Recording a person's voice saying the text is perhaps the most obvious way to include aural text. The sound files generated in the recording can then be attached to the appropriate objects to be played according to the parameters of the interface (activating the sound on mouse over, on focus, on keystroke).

Consideration: Currently Adobe's SVG viewer will play *wav* and *au* files. The SVG standard has yet to implement a standardized way of attaching sounds to objects in an SVG environment. The delay in this implementation was caused by a lack of open source sound format. Currently the only open source sound format is *ogg*. Future SVG browsers will probably be capable of playing *ogg* files.

The second way to include aural text as part of an interface is to implement a text to speech engine, which will allow the computer to speak any electronic text out loud in a synthesized voice. When a text to speech engine is implemented all that need be done is attach the electronic text version of what is to be said to the object it is to represent. Aural interfaces, such as screen readers, implement such text to speech engines.

Consideration: A significant advantage of this method of including aural text is the saving of file size. Electronic text takes up very little space and leaving the creation of the audio to the interface is very efficient. However, the trick to successfully making electronic text available to a text to speech engine is knowing where to put it. In most well developed technologies an aural interface such as a screen reader will be able to flexibly access any electronic text contained in a document. In the case of SVG, a technology that is still early in its development, a screen reader may not be able to find electronic text unless it is put in a very specific place in the file. So far we have been able to show that a screen reader can access electronic text contained in the "title" tag, but not the "label" or "text" tags.

When a designer is creating aural text for an interface there are several things that should be kept in mind. The text should be as clear and concise as possible. Just as map symbols visually communicate information in clear and concise ways the aural text that is attached to a symbol must be equivalently clear at delivering the information in an aural format. If recorded sound is being used the quality of sound must be such that the voice is easily comprehended.

If more than one voice is being used to disseminate the information there are some additional factors the designer must consider. When a person listens to disembodied voices the natural thing for the persons mind to do is to establish context. The easiest way to do this is by establishing what relationship the voices have. If two voices are present the listener will try to establish what roll each voice is playing (teacher/student, boss/employee, buyer/seller,) and which voice is dominant. As the number of voices increases the number of possible relationships between the voices will increase exponentially. As it is not desirable to have the user preoccupied with trying to determine which voice is speaking in what context, it is best to keep the number of voices being used to a minimum. While it is important to minimize the number of voices being used, multiple voices can assist the user in quickly comprehending information if different voices are used to delineate types of information. For instance, one voice may be used to give directional information, while another is used to voice basic symbol titles, and perhaps yet another voice gives geographic information.

Consideration: Use two different voices in the description of an object, one voice to give the type of data, another voice to deliver the specific data itself. For example:

Voice 1: School
Voice 2: University of Toronto, St George Campus

<u>Abstract sounds</u>

Sounds of an abstract nature comprise a category that encompasses all the beeps, whistles and buzzes computer applications make to convey information. The sounds may also be musical in nature. It is very easy to make this sort of sound short with respect to time and small with respect to storage space. Sounds are very good for giving cause and effect related feedback to users and to indicate the state of an element. The other value of abstract sounds is that it is easy to sonify data about the object being represented. For instance, a sound representing a topographical line might be higher or lower depending on the height the line represents. This gives the user information about what map symbol they are viewing and the approximate value it represents. This sound variation is an excellent way to communicate a relatively large amount of information in a very short amount of time. There is a long history of aural interfaces being used to sonify data as in the case of hospital heart monitors or Geiger counters.

Consideration: Of all the attributes a sound can have (pitch, duration, volume) pitch or change in pitch is the most effective at communicating the state or change in state of an object (Walker and Ehrenstein, 2000).

While abstract sounds can be used to identify objects it becomes quickly difficult for a user to memorize a large list of objects represented with abstract sounds. Abstract sounds may also have musical qualities (melody, harmony, rhythm) that can be used to enhance its ability to contain information. This possibility is useful when expressing more complex data sets. A set of abstract sounds with musical attributes could be used to give a user a summary of the type and scope of the elements contained in a geographic area (Joseph and Lodha, 2002).

Consideration: Rhythm is the most memorable element that a group of sounds can have. A list of abstract sounds that employ rhythm as one of their defining characteristics will be easier for a user to remember.

Aural metaphors

Aural metaphors are sounds that are reminiscent of sounds in the real world. It would seem that these types of sounds would be a natural fit for representing symbols on a map, as a map is a representation of the real world. These types of sounds work well for representing objects as the sound itself assists the user in identifying or recalling what the associated object is. This basis in the real-world can greatly reduce the learning curve a user goes through as they learn to read the map, and reduce reliance on the legend.

Consideration: Aural metaphors work very well for representing real world objects, such as types of buildings (school, bank, museum), areas (park or quarry), or infrastructure (roads, highways, transit or rail). There are some exceptions to this general rule of thumb. There are some real world objects that do not make sound or are not intuitively associate with a sound, such as a mountain range or even a generic building. Maps also represent abstraction of the real world such as political boundaries or topographical measurements. Aural metaphors don’t naturally represent these mapping elements.

Most objects in the real world that easily have sounds associated with them do not make a single type sound. For instance, a park can easily be associated with the sound of children playing, dogs barking, playground equipment squeaking, or people talking. A designer must approach most objects that will be represented with an aural metaphor with the intention of using synecdoche, where the part will represent the whole. For example, a park may be represented by the sound of a child’s laugh.

It is difficult to control the length of aural metaphors as they must endure long enough allow the metaphor to be recognized, and many real world sounds have an absolute length. Lengthy representations of symbols mean that it will slow the user down while looking for the information they want, and file sizes will be larger. In many cases this length may be unavoidable, in which case the effect on the ability of the user to efficiently read the map must be considered.

Consideration: If a real world sound is to be edited for length there are some things to keep in mind. The human ear identifies a sound based on its attack (the beginning of a sound) and its decay (the end of a sound). Any changes to the attack or delay of a sound will affect the listener's ability to identify the sound. The body of a sound (the middle part) is the most generic to the human ear. The body tells the listener how loud the sound is and how long it lasts, but its quality is not used to identify the sound. If a sound is to be changed in length the changes must be made to the body of the sound if it is to remain identifiable to the human ear.

<u>Location based sounds</u>
There is a fourth possibility for representing elements on a map with sound that could be referred to as location based sounds. These sounds would be used in the representation of actual places and would consist of a recording taken from the place being represented. For instance a location based sound for a specific park would involve recording a short (and hopefully representative) amount of time of the sounds from the park. Such a representation could give much information about the element being represented to the listener, who upon listening to the sound could potentially determine whether there was a playground in the park, whether there was a body of water in the park, or whether dogs were commonly found in the park. To the best of the writer's knowledge, this sort of representation has never been implemented in a geo-spatial context, and the practical usefulness of such representations is not known.

5.2.1.1 Spatialized Audio

Maps express geo-spatial data in a two dimensional, visual, symbolic format. When considering an aural alternative to visual geo-spatial maps it seems natural to take advantage of our ability to hear in three dimensions.

Aural displays (generally, acoustic speakers) are capable of creating the illusion that a sound is coming from a specific location. A stereo auditory display uses two speakers to generate the illusion of a sound coming from a specific spot, which can be located anywhere between the two speakers. This illusion, called the *stereo image*, is created by playing the sound louder on one speaker, creating the impression that the sound is located between the two speakers, though closer to the louder. The successful creation of a stereo image depends on the correct placement of the speakers, preferably at ear level and at the correct distance apart. This placement and method can provide the illusion of sound coming from a very specific location. A sound designer attaches panning or EQ effects to a sound to allow it to be displayed in three dimensions by a stereo system.

Panning is a feature of an electronic sound that has been authored for spatialized display systems, such as stereo or surround speakers. The panning effect is measured in percentages of left or right; centered referring to the state of being 0% left and 0% right. If panning is set to 100% right, the stereo image will give the illusion of a sound source coming from the right speaker. Likewise, panning 100% left gives the illusion of sound coming from the position of the left speaker. This is not to say that when panning is set to 100% right, there is *no* sound coming from the left speaker. If a person hears a sound that comes from their front-right, their left ear will still pick up some of that sound. To create a true illusion of sound coming from the far right, some sound would still have to come from the left speaker. The further an object is, the more sound a person will hear in the ear furthest from the source. The closer an object, the less sound a person will hear in the ear furthest from the source, until the source is so close that the sound is only heard in one ear. (Thus, the effect of something whispering in one ear.) We call this attribute of panning *divergence*, which can be manipulated by a sound designer to further create the illusion that a sound is coming from a specific location. If divergence is set to 100%, the sound source will seem to be extremely close to the listener. If divergence is set to 0%, sound will always be equal in both speakers, creating the effect of a distant sound coming from beyond the horizon.

Consideration: Remember that panning and divergence have relative values, displayed differently on different sets of stereo speakers. It is impossible to express absolute proximity or position using panning. The position of the sound source will stay relative to the position of the two speakers, with the absolute position varying from one set of speakers to another.

EQ (equalization) is the determination dominant frequencies in a sound. The effect of EQ on the spatial effect is related to how distinctly a sound stands out when played in the context of other sounds. If the frequencies that are distinguished in a sound are unique in the soundscape, the human ear will easily hear it as being in the foreground and distinct. A sound that shares its dominant frequencies with other sounds in a soundscape will "blend," causing the sound to recede into the aural background.

If a designer creates a group of sounds playing at the same time in order to depict a complete set of geo-spatial data, distinctness is vital. Ensuring that each sound has its own space in the spectrum of frequencies will help users distinguish different sounds from one another. The process of consciously doing this is called "carving EQ holes."

Consideration: When carving EQ holes, determine whether the sound that is to be predominant (or made "present") is in the low (20Hz-100Hz), mid-low (100Hz-500Hz), middle (500Hz-2kHz), mid-high (2kHz-8kHz) or high (8kHz-20kHz) frequency range. Next determine what the central frequency of the sound is, and boost it. Take all other sounds in the group sharing the same frequency range with the sound in question and reduce that same frequency.

Consideration: Sibilance is the hissing sound people make when they say the letter "s." When a recorded sound has sibilance, there is a hissing accompanying the actual sound. If the sound designer notices that a sound has strong sibilance, they can lower the EQ of the sound at 6-8kHz to eliminate the hiss.

5.2.1.2 The Aural Legend

On a traditional map the reader will find a legend detailing the meaning of symbols found on the map and explaining the mapping method used (such as scale or measurement systems.) The reader will typically use the strategy of glancing back and forth between the legend and the area being viewed. The reader of a map who uses an aural interface will need access to the same information. Due to the limited scanning ability of the aural sense this strategy would not work well for the latter, especially if the number of items represented in the legend were extensive.

Consideration: It becomes much more important to the user of an aural interface that objects represented on a map have their characteristics intrinsically associated with its representation.

One of the greatest differences between visual and aural interfaces with respect to geo-spatial maps is the issue of scale. A necessity of visual geo-spatial maps, scale becomes irrelevant to the user of an aural interface.

Consideration: All measurements of distance can be given in real terms with out any encumbrance on the usability of the map.

5.2.1.3 Information Scalability

In the study "Multichannel Auditory Search: Toward Understanding Control Processes in Polychotic Auditory Listening" (Lee, 2001), Mark Lee observes that our aural attentional systems experience overload faster than our visual attentional systems when presented with data sets of increasing size. That is to say, when a person is presented with a large amount of concurrent aural data, his/her ability to synthesize the data is quickly compromised. Lee found that our ability to pick information out of multiple data sets (that are spatially differentiated) is far more limited than comparative tasks using the visual sense.

Consideration: When displaying geo-spatial mapping data with sound, avoid overwhelming users with too much concurrent information.

A single map can give information about topography, political boundaries, infrastructure (roads, rail, shipping), population (cities, towns, regions), or structures for an area potentially covering the entire earth. Geo-spatial maps are potentially among the most complex data sets possible, simultaneously giving great quantities of many different types of information. The limited ability of the auditory sense to scan complex sets of data must be considered when an aural interface is designed, especially when the information to be disseminated is know to be a complex set.

Consideration: In the case of geo-spatial mapping data it becomes very important for the different types of data to be contained in different layers that can be hidden and revealed by the user. Providing this ability to hide and reveal information enables the user to create a simple data stream that can be more easily scanned with the aural sense while containing the most relevant information.

However the user may require many types of information delivered at once in a related context. This would require many layers to be revealed to the user plaited in the output of the aural interface. To compensate for the complexity of this output the user should have another way to scale the delivered information.

Consideration: Another way geo-spatial mapping information can be scaled is by restricting the size of the area being displayed. Just as an ideal aural interface for geo-spatial information should have different types of information separated into layers the total area of a geo-spatial map should be divided so as to allow the display to be restricted to a single area. For instance, if a user is able to restrict the displayed area to a single intersection, many different types of information can be displayed (streets, buildings, topography) without overwhelming the scanning ability of the aural sense.

5.2.1.4 Translation Algorithms

Translation algorithm suggests aural equivalents to the visual elements of a geo-spatial map. Keep in mind that across modalities, there is simply no one-to-one full translation equivalent. In choosing approximate meaning equivalents for visual mapping data we have three types of sound to choose from: aural text, real world sound (relating to object type), and abstract sound effect.

To make an aural map useful a method of navigation that is independent of hand eye coordination is very important. For a small and simple map section, such as would be used for the following example, we believe that selecting the arrow keys to navigate from object to object will allow the users to orient themselves to the geo-spatial content. When a user moves cursor focus onto a map object there is an initial sonification consisting of an audio icon (sound) and short descriptive text. The user may then use the tab key to prompt for even more audio information about the object, such as further description or directional information.

Naturally in a larger or more complicated geo-spatial map, such simple navigation would not suffice. More elaborate methods of filtering and selecting and navigating through information is suggested in the geo-political translation algorithm to follow.

Table 5 and Table 6 represent aural equivalents for a mapped intersection consisting of two streets, four buildings/structures, and a Park. The two charts, representing the same intersection, test the representation of the mapped elements fusing aural metaphors in the first chart, and abstract sounds in the second chart.

Urban Intersection Scenario

The following translation matrix conveys the sounds in an aural interface for an urban intersection map.

Object	Audio icon	Descriptive Text	Directional Text
Street 1	Traffic	James Street	North/south
Street 2	Traffic	Helen Avenue	East/west
Intersection	Toronto Intersection sound	James and Helen	Map coordinates x, y
Building 1 (school)	Chalk on board	School; Saint George Elementary	North west corner; James and Helen
Building 2 (Bus Shelter)	Bus doors opening bus doors	James street bus stop; bus # 54	South east corner; James and Helen
Building 3 (Bank)	Coin	First national bank	North west corner; James and Helen
Building 4 (Library)	Book pages	Brook View public library	South west corner; James and Helen
Area 1 (Park)	Children playing/cheering	Begrundi Memorial park	North of Helen on James; west side

Table 5: Real world or aural metaphor sounds for an intersection map.

Object	Audio icon	Descriptive Text	Directional Text
Street 1	Abstract 1	James Street	North/south
Street 2	Abstract 2	Helen Avenue	East/west
Intersection	Abstract 3	James and Helen	Map coordinates x, y
Building 1 (school)	Abstract 4	School; Saint George Elementary	North west corner; James and Helen
Building 2 (Bus Shelter)	Abstract 5	James street bus stop; bus # 54	South east corner; James and Helen
Building 3 (Bank)	Abstract 6	First national bank	North west corner; James and Helen
Building 4 (Library)	Abstract 7	Brook View public library	South west corner; James and Helen
Area 1 (Park)	Abstract 8	Begrundi Memorial park	North of Helen on James; west side

Table 6: Abstract sounds for an intersection map.

Geo-Political Map Scenario

The following translation matrix conveys the sounds in an aural interface for a geo-political map.

Geographic Element	User Action	Non-speech Audio	Speech
Latitude and Longitude Lines	Proximity		
	Over	Elastic, Boing	
	Mouse Click		"X Longitude"
Provincial Boundaries	Proximity	Rapid clicking	
	Over	Sucking in and then out sound	"Provincial Boundary between x and y" (X being province just left)
	Mouse Click		
Provinces	Proximity		
	Over		
	Mouse Double Click	Three note ascending scale with clarinet, continuous	Shape of Ontario
Regions, Maritimes, Central Canada, Prairies, Arctic, West Coast	Proximity		
	Over		
	F1 key	Three note ascending scale with oboe, continuous	X Region of Canada
Cities	Proximity	Busy City Street Sound, cars honking, people talking, traffic	
	Over	Increase volume upon entering zone, interrupted by speech	Name of City spoken
	Mouse Click		City and Population Size spoken
Provincial Capital	Proximity	Busy City Street Sound, cars honking, people talking, traffic	
	Over	Increase volume upon entering zone, interrupted by	Name of City spoken

Geographic Element	**User Action**	**Non-speech Audio**	**Speech**
		speech	
	Mouse Click		City, Capital of X, Population Y
Canadian Capital Ottawa	Proximity	National Anthem	
	Over	Louder interrupted by speech	Speak "Ottawa, Canada's Capital”
	Mouse Click		City and Size spoken
Lakes	Proximity	Small wave lapping with loon sound	
	Over	Louder interrupted by speech	Lake Y
	Mouse Click		Lake Y, size in Square miles
Rivers	Proximity	Whitewater river sound	
	Over	Louder interrupted by speech	Y River
	Mouse Click		Y River, x miles long, flowing into y
Waterways	Proximity	Harbour sound, boats chugging, tooting	
	Over	Louder interrupted by speech	St. Lawrence Seaway
	Mouse Click		X waterway, route from x to y
Ocean	Proximity	Waves, Seagulls	
	Over	Louder, Interrupted by speech	X Ocean
	Mouse Click		
Towns	Proximity	Playground sounds	
	Over	Louder interrupted by Speech	Name of Town
	Mouse Click		Name of Town, Population X
Major roadways	Proximity	Highway, Cars zooming past	
	Over	Louder interrupted by speech	Highway x
	Mouse Click		
Railways	Proximity	Train sounds	

Geographic Element	User Action	Non-speech Audio	Speech
	Over	Louder interrupted by speech	X Rail
	Mouse Click		X Rail between Y and Z
International Boundaries	Proximity	American Anthem	
	Over	Louder interrupted by speech	"International Border with the United States"
	Mouse Click		
	Mouse Double Click		

Table 7: Geopolitical map translation matrix.

5.2.2 Translations: Visual to Touch

For general discussions on types of Haptic Devices see 1.5.3.1 Haptic Devices and 1.6.2.3 Visuals to Haptics. This document focuses on tactile effect authoring software for the consumer level haptic device technology (called TouchSense) licensed by Immersion Corporation.

5.2.2.1 Full Force and Tactile Feedback Distinctions

Immersion Corporation defines the difference between their force and tactile feedback devices as follows (www.immersion.com/developer/technology/devices):

> The term **full force feedback** is used to refer to devices that apply sideways forces to your hand to resist your motion or give the impression of turbulence, recoil, impact, G-forces, or countless other phenomena. If the device can physically push on you, it's probably a full force feedback device. Many force feedback gaming devices fall into this category.
>
> The term **tactile feedback** is used to describe devices that play high-fidelity tactile sensations, but generally won't move or inhibit the movement of either the device or the hand holding it. A tactile feedback device can play a wide variety of distinguishable taps, textures, and vibration effects to communicate with the user and greatly enrich the computing or gaming experience. A number of pointing and gaming devices fall into this category.

5.2.2.2 Haptic Effects Provided by Immersion Corporation

Following are pre-set, editable effects offered from the Immersion Studio library. While there is a large selection for force feedback devices, there is a small selection for tactile feedback devices, due to their limited nature.

Effects for *Immersion* Full-Force Feedback Devices:

- ***Angle Wall Effect*** – feel like a wall at an arbitrary angle
- ***Axis Wall Effect*** – feel like vertical or horizontal wall
- ***Damper Effect*** – feel like dragging a stick through water
- ***Ellipse Effect*** – used to attract the cursor to the inside of the ellipse, keep the cursor outside of the ellipse, or attract the cursor to the border surround the ellipse
- ***Enclosure Effect*** – used to attract the cursor to the inside of the rectangle, keep the cursor outside of the rectangle, or attract the cursor to the border surround the rectangle
- ***Friction Effect*** – feel like pushing a mass across a rough surface
- ***Grid Effect*** – creates a 2-dimensional array of snap points or snap lines
- ***Inertia Effect*** – feel like pushing a mass on wheels
- ***Slope Effect*** – feel like rolling a ball up and/or down a hill
- ***Spring Effect*** – feel like compressing a spring
- ***Texture Effect*** – causes mouse to feel as if it were traveling over a series of bumps
- ***Periodic Effect*** – feel like a simple back and forth motion or a high frequency vibration
- ***Pulse Effect*** – identical to Vector Force but its default duration is much shorter to create a "pop" sensation
- ***Ramp Effect*** – creates a force that either steadily increases or steadily decreases over a discrete time
- ***Vector Force Effect*** – constant force over a discrete time

Effects for *Immersion* Tactile-Feedback Devices

- ***Ellipse Effect*** – a small "pulse" will be emitted when cursor is over the border of the ellipse
- ***Enclosure Effect*** – a small "pulse" will be emitted when cursor is over the border of the rectangle
- ***Grid Effect*** – creates a 2-dimensional array of lines which emits a "pulse" when cursor is over it.
- ***Texture Effect*** – causes mouse to feel as if it were traveling over a series of bumps by pulses as the mouse moves.
- ***Periodic Effect*** – high or low frequency vibrations
- ***Pulse Effect*** – create a "pop" sensation

5.2.2.3 Tactile Effects Applied to Geospatial Content

Each effect has parameters (such as duration, gain, magnitude, phase, waveform) that can be finessed for specific results. User action of proximity refers to cursor being outside of but in a predefined number of pixels from the element.

Following is a translation algorithm for representing typical map elements with haptic effects using a **full-force feedback** mouse or trackball.

Geographic Element	User Action	Possible Haptic Effects	Description of Effect
Latitude and Longitude Lines	Over	Spring	Elastic
Provincial Boundaries	Proximity	Slope	Thin trench on both sides of curtain
	Over	Spring	Elastic curtain
Provinces	Mouse Double Click	Ramp	Relief of province rises/ Alternatively sinks
Regions, Maritimes, Central Canada, Prairies, Arctic, West Coast	F1 key	Ramp	Relief of Region Rises/Falls (toggled)
Cities	Proximity	Ellipse	Slight gravity well around sphere
	Over	Slope	1/2 sphere
Provincial Capital	Proximity	Enclosure	Slight Gravity Well around Pyramid
	Over	Slope	Pyramid
Canadian Capital Ottawa	Proximity	Ellipse	Slight gravity well around stars
	Over	Slope	1/2 sphere
Lakes	Proximity	Slope	Depression around lake
	Over	Periodic	Pulsing waves of resistance
Rivers	Over	Slope	Trench
Waterways	Proximity	Slope	Trench along sides
	Over	Periodic	Washboard in direction of flow
Ocean	Proximity	Slope	Depression Around Coast
	Over	Periodic	Pronounced Waves
Towns	Over	Slope	Small slightly raised square
Major roadways	Over	Slope	Raised ribbon with indentation in middle
Railways	Over	Texture	Raised ribbon with horizontally running indentations
International Boundaries	Over	Slope	Very high Edge

Table 8: Translation algorithm for representing typical map elements with haptic effects.

Unlike Full-Force feedback devices, which can create the effect of resistance when the pointing device is over an object, **tactile feedback** pointing devices are limited to variations of pulses and vibrations. As a result, the number of geographic elements that can possibly be represented by tactile effects is very limited, and should be accompanied by other modalities of representation (see 5.4 Modality Complementarity).

Consideration: Pulses experienced through a tactile feedback device may not be very informative since a single "bump" may indicate that either a user has passed over *or* is currently on an object. Consider, for example, an image of a thin line such that when the cursor is over the image, a single pulse is emitted. However, when a user is moving the cursor about and passes over the image, the user will also feel a pulse. If the goal for the user was to move the cursor along the line, it will be very difficult for them to distinguish between being currently over the image or having just passed over it.

Periodic effects can best be utilized in geo-spatial applications as follows:

Vibrations with a constant frequency or amplitude can be used as indications for a single geographic element, such as roads or relief. Using the example of roads, varying the frequency can indicate the type of road when the cursor is over the object. A high frequency might be used for highways, low frequency for side streets.

Vibrations with varying frequencies or amplitudes can be used as indications for specific objects, such as towns, buildings or points of interest. For example, a building can be indicated by a "buzzing" effect where vibrations cycle from large to small amplitudes when the cursor is over the object.

5.2.3 Format: Scalable Vector Graphics (SVG)

Scalable Vector Graphics (SVG) is an XML compliant graphics format being developed by the World Wide Web Consortium (W3C, www.w3.org) to be the dominant method of rendering graphics on the Web. An open standard, it is being adapted by conventional web browsers. Full SVG specifications can be found at http://www.w3.org/TR/SVG. While this technology continues to undergo development, it already shows great promise in terms of creating accessible, interactive graphical environments.

Access to information can be greatly enhanced through interactive features made possible by the SVG format. By allowing a user to hide different elements, information can be simplified and made easier to read in a SVG document. A strength of the SVG format is the ability of the creator to layer information. By grouping similar elements in a layer within the context of a larger document allows the user to more easily locate the element, or group of elements that contain the information they are searching for.

Geo-spatial content is typically rendered in a graphical medium with an iconic language that is very difficult to translate into another medium. SVG gives the map creator the ability to associate electronic text with graphical elements and layers of graphical elements. This electronic text, unlike the iconic language of geo-spatial maps, is easily transformed into another mode, such as audio. An audio interface such as a screen reader can render the electronic text in the form of a spoken voice giving the aural user access to a graphical element. Of course, the creator of a map in the SVG format will have to include descriptive text for the aural interface to read. SVG authoring programs in the future should contain prompts allowing content creators to easily include text information about the elements and layers they are creating.

Earlier iterations of the SVG standard had some limitations that affected the accessibility of SVG documents. Java script handlers controlling keyboard control, vital to the success of an aural interface, could not be used due to a lack of keyboard support in the Document Object Model (DOM) before the inception of level 3. This lack keyboard support prevented testing aural interfaces built with SVG for the duration of this project. With the inception of DOM level 3 validation on January 27, 2004 and the continued development of SVG 1.2 (still in working draft as of March 18, 2004) this problem has been solved. It is not known how many of the SVG browsers have implemented support for the working version of SVG 1.2.

In testing SVG documents with the JAWS screen reader it was shown that JAWS could read "title" tags, but we not able to get "desc," "label," or "text" tags to be read. The scripting group at Freedom scientific (maker of the JAWS screen reader) has not yet addressed JAWS' ability to work with this relatively new standard.

5.2.3.1 Resources

Accessibility features of SVG are discussed in detail at the W3C's Web Accessibility Initiative (WAI) site www.w3.org/TR/SVG-access.

The Mapping for the Visually Impaired portal (MVI, tactile.nrcan.gc.ca) at Natural Resources Canada is a forerunner in attempting to make maps and geo-spatial data accessible on-line. Associated MVI members presented a seminal paper at the summer 2003 SVG OPEN conference, "SVG Maps for People with Visual Impairment" (www.svgopen.org/2003/papers/svgmappingforpeoplewithvisualimpairments), describing techniques for creation of audio-tactile maps in SVG and Flash formats. Some of this work bootstrapped off of trailblazing by Dr. John Gardner (himself blind), Vladimir Bulatov and team at Oregon State University's Science Access Project (dots.physics.orst.edu). A paper describing their development and application of the Accessible SVG Viewer, incorporating haptic feedback and object description via a speech engine, was presented at the 2001 CSUN conference, "Smart Figures, SVG, And Accessible Web Graphics"(www.csun.edu/cod/conf/2001/proceedings/0103gardner.htm).

Not relevant to SVG so much as to Cartography and Map Symbols, Natural Resources Canada's online Atlas of Canada (atlas.gc.ca/site/english/learning_resources/carto) has a highly readable section (intended for students and teachers) entitled *The Fundamentals of Cartography*, which includes subsections *Map Content and Design for the Web* and *Cartographic Symbology.*

5.2.3.2 Methodology

Our recommendation for associating haptic effects (using Immersion TouchSense) with web-based visual information using SVG has been previously described in *5.2.3 Visuals to Haptics*. The procedure following differs in that it adds detail for adding sound, and provides code samples. User requirements are:

1. An SVG viewer plug-in associated with their browser, such as the Adobe SVG Viewer, found at www.adobe.com/svg/viewer/install
2. Immersion Web Plug-in, found at www.immersion.com/plugins

The general procedure[3] for adding accessible graphics to HTML (in this context, a map, with audio and tactility (using Immersion-licensed TouchSense devices)) is as follows:

1. *Construct the SVG*

   Any text editor can be used. Third-party software, such as Adobe Illustrator 10 or 11 (www.adobe.com/products/illustrator), can greatly speed up development.

5.2.3.3 Code Sample 1: An SVG that draws a fill red circle.

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.0//EN"
   "http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd" >
<svg xmlns="http://www.w3.org/2000/svg"
        xmlns:xlink="http://www.w3.org/1999/xlink"
        xmlns:a="http://ns.adobe.com/AdobeSVGViewerExtensions/3.0/" >
        <circle cx="25" cy="25" r="20" fill="red"/>
</svg>
```

2. *Organize Objects in Layers for Interactivity*

   Group all elements in appropriate layers and label both elements and layers. For example, in a map setting, put symbols for train station, bus station, and airport into one group, primary, secondary and tertiary roads into a second group, mountains and bodies of water forming two more groups. All interactive elements need to be grouped, which creates a group tag in the SVG document; in Illustrator, use the layers palette. Each element identifiable by screen readers needs a label, which will in turn be spoken.

3. *Assign Haptic and Sound Effects*

   Assign JavaScript functions to layers. Associate each haptic effect with a sound effect, which makes map features more evident to people with visual impairment, by placing before the opening of the group tag <g> an audio anchor <a:audio> with the attribute xlink:href set to the file name of the sound to be played. The event handler must also be set in the audio anchor with the "begin" attribute. Each

[3] Steps two and three are adapted from SVG Maps for People with Visual Impairment (www.svgopen.org/2003/papers/svgmappingforpeoplewithvisualimpairments)

category of map elements gets assigned a separate haptic effect with JavaScript, e.g. one for roads, one for built-up areas, one for water and occasionally one for borders. JavaScript functions to detect user input and trigger the events; in Illustrator, use the interactivity palette. Otherwise, for those versed in JavaScript, code can be added to the SVG with a simple text editor.

5.2.3.4 Code Sample 2: Using Sample 1 code, group objects together, add sound and assign JavaScript functions as needed (shown in bold).

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.0//EN"
"http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd" >
<svg xmlns="http://www.w3.org/2000/svg"
          xmlns:xlink="http://www.w3.org/1999/xlink"
          xmlns:a="http://ns.adobe.com/AdobeSVGViewerExtensions/3.0/" >
     <a:audio xlink:href="button.wav" begin="mybutton.mouseover"/>
   <g id="mybutton" onmouseover='parent.startHaptic()'>
          <circle cx="25" cy="25" r="20" fill="red"/>
   </g>
</svg>
```

4. Incorporate SVG and haptic effects into HTML

Use an EMBED tag to place the SVG graphic in a webpage. An object reference to the Immersion Web Plug-in must be made, along with JavaScript referring to the tactile effects.

5.2.3.5 Code Sample 3: The minimum HTML code required to implement haptics with SVG.

("Sports.ifr" is a resource file containing haptic effect "Punch" created with Immersion Studios.The previous sample provides "circleWithJavaScript.svg".)

```
<HTML>

<BODY onload="IMM_openIFR();">


<!-- Object reference to the Immersion Web Plugin must be included in the HTML.-->
<OBJECT
     classid="CLSID:5CA42785-ABC3-11d2-9F81-00104B2225C5"
    codeBase="http://www.immersion.com/plugins/ImmWeb.cab#version=3,2,3,0"
    id="ImmWeb" height="0" width="0">
</OBJECT>

<SCRIPT LANGUAGE="JavaScript">
var imm_eref;

function IMM_openIFR( )
{
    imm_eref = document.ImmWeb.OpenForceResource("Sports.ifr");
}
```

```
function startHaptic( )
{
    document.ImmWeb.StartEffect("Punch", imm_eref);
}
</SCRIPT>
<EMBED NAME="circle" SRC="circleWithJavaScript.svg" WIDTH="700"
HEIGHT="400">
</BODY>

</HTML>
```

5. *Add both Immersion Web plug-in and user environment detection to HTML*

Use an OBJECT tag and JavaScript to detect presence of the Immersion plug-in associated with users' web browser. (Code supplied in Appendix. It is available from Immersion Developer site www.immersion.com/developer/technology/tutorials/index.php, log in required, within the tutorial "TouchSense in JavaScript - Web Tutorials: Detecting Immersion Web Plugin".)

5.2.3.6 GIS Applications and SVG for Government On-Line (GOL) Content

Maps are not only essential but powerful tools. Digital maps on the Web and in Geographic Information Systems (GIS) allow more and more people to make and use maps to analyze data in sophisticated ways. GIS applications are found in various government sectors, business, the natural and social sciences, urban planning and management, and many other fields. The two most prevalent GIS software applications are ESRI's ArcView and MapInfo, both of which do not have native SVG output. Fortunately, two third-parties have recognized the promise and utility of the SVG format, and have developed extensions to give ArcView and MapInfo better web deliverability with SVG export.

ESRI's ArcView (www.esri.com/software/arcview) is perhaps the most popular desktop GIS and mapping software, with more than 500,000 copies in use worldwide. ArcView provides data visualization, query, analysis, and integration capabilities along with the ability to create and edit geographic data.

MapView*SVG* (www.uismedia.de/mapview/eng) is a third party extension for ArcView that converts maps within ArcView to the SVG format for dissemination on the Web. All vector based objects and text objects are converted into SVG, which can be panned and are infinitely zoomable without losing cartographic quality. Various elements can be turned on and off in the display, queries made, and many other features including web and email links, scale bars and overview maps, coordinate read-outs and interactive measurement tools.

MapInfo (www.mapinfo.com) is another leading business mapping solution that allows users to create detailed maps, reveal patterns and trends in data that may otherwise be hard to discern, perform data analysis, manage geographically based assets, etc.

DBx Geomatic's SVGmapmaker (www.dbxgeomatics.com) is a program that exports MapInfo content to SVG in such a way that the result looks as much as possible like the original document in MapInfo. Several options are also available directly in MapInfo that take advantage of SVG's advanced graphic capabilities, such as filter effects, color gradients, opacity, hyperlinking, etc. DBx Geomatics is a Canadian company, with a strong clientele across federal (Public Works and Government Services Canada, Environment Canada, Canadian Coast Guard, Fisheries and Oceans Canada, Treasury Board of Canada Secretariat) through to municipal governments.

5.3 3D Content

3D content is usually an object, collection of objects, or a virtual environment. This document focuses on web3D, which entails an on-line, bandwidth efficient technology. Panoramic technologies are not discussed herein. Web3D is a navigable, interactive experience involving more than the spin and zoom nature of panoramas. This format allows for a large variety of time and user triggered events, with cause and effect through scripting and potential interface capabilities such as query and filter.

The term "Web3D" generally refers to a spectrum of on-line interactive 3D technologies, many of them proprietary. While we are concerned with the predominant, open-source standards of Virtual Reality Modeling Language (VRML) and X3D, readers are encouraged to extend accessibility and usability considerations herein to other formats.

5.3.1 Transformations: Visual to Sound

For individuals who are blind, environmental sounds can reveal much more perceptual and navigational information about a modern urban landscape than the sighted would fathom. The following scenario is used by Massof (2003, p.271) to describe the level of aural acuity involved:

> *After vision, hearing has the broadest band for acquiring information. Blind people rely almost entirely on hearing to perceive the environment beyond their reach. A highly skilled blind pedestrian can approach an intersection, listen to traffic, and on the basis of auditory information alone, judge the number and spatial layout of intersecting streets, the width of the street, the number of lanes of traffic in each direction, the presence of pedestrian islands or medians, whether or not the intersection is signalized, the nature of the signalizations, if there are turning vehicles, and the location of the street crossing destination.*

Massof concedes that "not all blind people demonstrate this skill" (2003, p.271); however, this example is still useful in illustrating how a 3D soundscape may be used to convey geo-spatial information to individuals who cannot view a map, or see their virtual environment surroundings.

Not only man made, but natural sound sources are important perceptual cues, though, as Potard (2003) states, they are rarely implemented or properly controlled in virtual 3D sound scenes. Sound sources such as a beachfront or waterfall usually exhibit a particular spatial extent which provides information on the physical dimensions, geometry and distance of the sound emitting object.

In way-finding tasks, auditory beacons can prove extremely useful in guiding a user to locations that are not visible due to buildings or walls. In observing the navigation paths of subjects, Grohn *et al* (2003) found that users used auditory cues to define approximate target locations, preferring visual cues only in the final approach.

Many of the consideration made in 5.2.1.1 Spatialized Audio can be applied to 3D contexts. However, more sophisticated 3D sound playback technologies can be found in almost all current computer sound hardware, which support not only two channel stereo, but surround speaker systems which are both widely available and affordable. Web3D technologies have usually supported spatialized audio playback, but newer browsers are pushing the sound envelope with streaming, multi-channel aural content.

Vision, in these implementations, can be combined with different types of sound, like speech, real-life sounds (auditory icons) or abstract musical sounds (earcons). Each of these types of sounds has advantages and disadvantages. The benefit of speech, for instance is that most of the time, the meaning of the message is relatively unambiguous. However, users need to listen to the whole message to understand the meaning.

5.3.1.1 Visual to Touch

Calle Sjöström and the Haptics Group at Certec (http://www.english.certec.lth.se), Lund Institute of Technology, Sweden, has created the doctoral dissertation, "Non-Visual Haptic Interaction Design: Guidelines and Applications" (www.english.certec.lth.se/doc/hapticinteraction). They present five guidelines for researchers, designers, testers, developers and users of non-visual haptic devices:

- *Elaborate a virtual object design of its own*
    - Avoid objects with small and scattered surfaces. Objects with large connected surfaces are easier to find and explore
    - Use rounded corners rather than sharp ones.
    - Virtual objects in virtual worlds can be given virtual properties. Utilize them.
    - Optimize your haptic interface widgets as well. Think about affordance.
    - Make sure that the models are haptically accurate and work without vision.
    - Be aware that orientation of the object matters.
    - Consider different representations to enhance different properties (negative relief emphasizes the line whereas positive relief emphasizes the contained surface).
- *Facilitate navigation and overview*
    - Provide well defined and easy-to-find reference points in the environment.
    - Avoid changing the reference system.
    - Make any added reference points easy to find and to get back to. They should

    also provide an efficient pointer to whatever they are referring to.
  - Utilize constraints and paths, but do so with care.
  - Virtual search tools can also be used.
- *Provide contextual information (from different starting points)*
  - Present the haptic model or environment in its natural context.
  - Provide information about the purpose of the program.
  - Provide information about possibilities and pitfalls in the environment
  - Use a short text message such as a caption to an image or model, provided as speech or Braille. This can make a significant difference.
  - Idea: Consider using an agent or virtual guide that introduces the user to the object and also gives additional information if requested.
- *Support the user in learning the interaction method and the specific environments and programs*
  - Be consistent; limit the number of rules to remember.
  - Give clear and timely feedback on the user's actions.
  - Facilitate imitating of other users and situations if possible.
  - Develop elaborated exercises to make the handling of the interaction tools and methods automatic in the user.
  - Idea: Consider using a virtual guide or remote users to help when a user comes to a new environment.

These guidelines represent the learning the research group had gained during the testing of blind subjects using haptics devices in tasks such as the recognition of both geometrical and real life objects (with texture) and navigation in a traffic environment. Their work showed strong hope for effectively using haptic technology to translate both 2D and 3D graphical information and make it comprehensible to the Blind. Not only did they demonstrate that a blind person can orient and navigate in a virtual haptic environment, but that these tasks can be more efficient with complementary information via sound or Braille. They ascertained that a blind person can use virtual world familiarization for real life scenarios.

For *virtual environments*, the researchers also provide five basic prerequisites for being able to work efficiently:

- To be able to explore, understand and manipulate the objects in the environment
- To navigate and to gain an overview
- To understand the context
- To use all modalities that are normally used
- To learn the interaction method and the specific environments/programs

5.3.1.2 Visual to multimodal

In Sjöström's work, an emphasis on the need to utilize multiple sensory modalities is made. His guidelines for including sound are as follows:

- Utilize all available modalities

- Combine haptics with sound labels, a Braille display and/or synthetic speech for text output.
- Try environmental sound to aid in getting an overview.
- Use audio (both sound labels and environmental sound) to provide a context.
- Provide feedback to the user via any available sense.

5.3.2 Format : Web3D - Virtual Reality Modeling Language (VRML)/X3D

Virtual Reality Modeling Language (VRML) is the ISO standard for interactive 3D web content. While there are many proprietary and open-source types of online, interactive, three-dimensional technologies, VRML has been the primary Web3D format. Various working groups have developed the next generation called X3D (X for eXtensible), which is both backward compatible with VRML and integrated into XML. Specifications of VRML and X3D can be found at www.web3d.org/x3d/spec.

VRML does provide spatialized audio capabilities, though text-to-speech (i.e. passing text to a speech synthesizer) isn't. The benefit of speech, in contrast to abstract sound, is that meaning of the signal is relatively unambiguous. However, users generally have to listen to the full duration to comprehend. The *Adding Feeling, Audio and Equal Access to Distance Education Project* undertaken by the Adaptive Technology Resource Centre (University of Toronto) was able to take advantage of the VRML PROTO node to enable such a capability in the form of *SpeechClip* nodes. Features included the ability to set or modify voice, volume, speech rate, and pitch, as well as some ability to control the "play back" of the speech: interrupt a speech, pause/resume a speech, and loop a speech. However, the ideal implementation of text-to-speech in VRML could not be fully implemented at the time due to JSAPI (Java Speech Application Programming Interface) and browser security issues.

As X3D is an XML based technology, the best hope for seamless speech processing in Web3D may be in the parallel developing markup language efforts of HumanML (www.humanmarkup.org) and the Humanoid Animation Working Group (H-ANIM, www.h-anim.org). As for tactile interface support, there is very little in VRML. The *Adding Feeling, Audio and Equal Access to Distance Education Project* rendered VRML scenes haptically, though it was specific to a force feedback device no longer commercially available. An obvious choice of haptic device technology to pursue VRML/X3D compatibility with is the most widespread and affordable—Immersion Corporation's TouchSense line (www.immersion.com/consumer_electronics/applications/mice_trackballs.php).

5.3.2.1 Methodologies

The following two procedures implement VRML with haptics effects readable by Immersion-licensed TouchSense devices.

Method 1: HTML plus VRML plus a Java Applet using EAI

Approach: Traditional VRML with Java using the External Author Interface, interface which was designed to allow an external program (the applet) to access nodes in the VRML scene.

User requirements:

1. Web3D Viewer with complete VRML support and EAI (External Author Interface) support. (Cortona VRML Client from Parallel Graphics (www.parallelgraphics.com/products/cortona) works well).
2. Immersion Web Plug-in, found at www.immersion.com/plugins
3. Required additional java classes*:*
   vrml.external.*
   vrml.external.field.*
   vrml.external.exception.*

*Cortona places these classes in corteai.zip ("old" and "new" External Authoring Interface (EAI) for Internet Explorer), plywood.jar (SAI and old EAI for Netscape). The default location of these are C:\Program Files\Common Files\ParallelGraphics\Cortona\corteai.zip and C:\Program Files\Netscape\Netscape\Plugins\plywood.jar

General procedure:

1. Create HTML with the necessary JavaScript code needed to add touch sensations to web pages.

5.3.2.2 Code Sample 1: Initial HTML with JavaScript to enable immersion effects

```
<HTML>

<BODY onload="IMM_openIFR();">

<!-- An object reference to the Immersion Web Plugin must be included in the HTML.-->
<OBJECT
    classid="CLSID:5CA42785-ABC3-11d2-9F81-00104B2225C5"
    codeBase="http://www.immersion.com/plugins/ImmWeb.cab#version=3,2,3,0"
    id="ImmWeb"
    height="0" width="0">
</OBJECT>

<SCRIPT LANGUAGE="JavaScript">
var imm_eref;

function IMM_openIFR()
{
    imm_eref = document.ImmWeb.OpenForceResource("Sports.ifr");
}

function startHaptic()
{
    document.ImmWeb.StartEffect("Punch", imm_eref);
}

</SCRIPT>
```

```
</BODY>

</HTML>
```

2. Create VRML content with events to be detected (in our case it was collision)

5.3.2.3 Code Sample 2: VRML generating a red sphere that will detect collisions

```
#VRML V2.0 utf8

DEF COLLIDER Collision{
children [

Transform {
  children Shape {
    appearance Appearance {
      material Material {  diffuseColor 1.0 0 0  }
    }
    geometry Sphere { }
  }
}
]
collide TRUE
}
```

3. Create a small applet to detect events from VRML using EAI. Detected events will call the JavaScript functions in the html to start effects.

5.3.2.4 Code Sample 3: Java applet used with VRML from Step 2 to detect collisions.

```
import java.awt.*;
import java.lang.*;
import java.applet.Applet;

import netscape.JavaScript.*;
import netscape.JavaScript.JSObject;

// Needed for External Authoring Interface (EAI)
import vrml.external.*;
import vrml.external.field.*;
import vrml.external.exception.*;
import vrml.external.Node;
import vrml.external.Browser;

public class EAI extends Applet implements EventOutObserver{

  Browser browser = null;
  Object eventKey;
  JSObject win;
  boolean once = true;
  int counter = 0;

  public void init(){
```

```
 // Get a reference to the Navigator Window:
 try {
   win = JSObject.getWindow(this);
 }
 catch (Exception e){
     // Don't throw exception information away, print it.
     e.printStackTrace();
 }

 // Get the VRML Browser:
 browser = Browser.getBrowser(this);

 // Get a handle on the TouchSensor node (in this case it is the COLLIDER node):
 Node ts = browser.getNode("COLLIDER");

 // Get collission time from the event collideTime
 EventOutSFTime bump = (EventOutSFTime)ts.getEventOut("collideTime");
 bump.advise(this,null);
}

// What do when when an event occurred
public void callback(EventOut event, double timestamp, Object key){

 // Retrieve the state of isActive (see above the advise function)
 EventOutSFBool state = (EventOutSFBool) key;

 // Muffle the number of calls made to the callback method
 // (and consequently the haptic device as well) by only allowing
 // a call to JavaScript once every 20th time.
 if (counter % 20 == 0)   // calls js every 20th collision
   callJS();
   counter ++ ;
 }

 // Call JavaScript in HTML
 public void callJS()          {
   win.call("startHaptic", new String[1]);
 }
}
```

4. Embed the VRML and Applet into the HTML.

5.3.2.5 Code Sample 4: Embedding VRML and applet into the HTML from Step 1

(“Sports.ifr” is a resource file containing haptic effect “Punch” created with Immersion Studios.)

```
<HTML>

<BODY onload="IMM_openIFR();">


<!-- An object reference to the Immersion Web Plugin must be included in the HTML.-->
<OBJECT
    classid="CLSID:5CA42785-ABC3-11d2-9F81-00104B2225C5"
```

```
    codeBase="http://www.immersion.com/plugins/ImmWeb.cab#version=3,2,3,0"
    id="ImmWeb"
    height="0" width="0">
</OBJECT>

<SCRIPT LANGUAGE="JavaScript">
var imm_eref;

function IMM_openIFR()
{
    imm_eref = document.ImmWeb.OpenForceResource("Sports.ifr");
}

function startHaptic()
{
    document.ImmWeb.StartEffect("Punch", imm_eref);
}

</SCRIPT>

    <EMBED NAME="vrml" SRC="sphereWithCollision.wrl" WIDTH="700"
HEIGHT="400">
    <applet name="EAI" code="EAI.class" MAYSCRIPT align="baseline" width="500"
height="100">
            <PARAM NAME="name" VALUE="EAI">
            <PARAM NAME="MAYSCRIPT" VALUE="">
    </applet>
</BODY>

</HTML>.
```

Additional Note: Only works with JVM from Microsoft, which no longer will be supported after September 30, 2004. In future, Cortona's VRML client will support the latest versions of Java and Java Virtual Machine from Sun. In the meantime, this version of Java can be used with a solution offered by J-Integra from Intrinsyc (j-integra.intrinsyc.com). See details at www.parallelgraphics.com/developer/kb/article/162

Method 2: HTML plus VRML plus a Java Applet using Xj3D

Approach: This method uses Xj3D, a toolkit written in Java for VRML97 and X3D content. It was a project initiated by Sun Microsystems in 1999 and handed over to the Web3D Consortium. The toolkit allows users to incorporate VRML97 and X3D content into their Java applications, such as using it to load VRML/X3D content.

However, it is important to note that Xj3D in still in development. Although the toolkit in its present state can be utilized effectively for some applications, it may be inadequate for visual-to-haptic applications. The current status of Xj3D (www.xj3d.org/status.html) indicates that many of the nodes/features, such as collision, have been implemented or partially implemented and tested locally, but have yet to be proven compliant with all Web3D published conformance tests.

Theoretically, when Xj3D is fully compliant, it may be used effectively in visual-to-haptic applications. This would eliminate the need for additional plug-ins, such as Parallel Graphics' Cortona. Furthermore, since the toolkit was written with Java 1.4 it is fully compatible with Sun Microsystems' JVM. The Xj3D toolkit can be obtained from www.xj3d.org/download.html.

User requirements:

1. Immersion Web Plug-in, found at www.immersion.com/plugins
2. Required additional java classes:
   Download the latest stable release of the toolkit from www.xj3d.org/download.html to implement an xj3d browser capable of viewing VRML.

General procedure summary*:*

1. Create HTML with the necessary JavaScript code needed to add touch sensations to web pages.

2. Create VRML content with events to be detected (in our case it was collision).

3. Create a small applet to load the VRML file to an Xj3D browser and to detect events from VRML using EAI. Detected events will call the JavaScript functions in the html to start effects.

4. Embed the Applet into the HTML.

5.3.2.6 VRML/X3D and GIS

GeoVRML
GeoVRML, an amendment to the Virtual Reality Modeling Language specification (and incorporated in VRML's offspring X3D), gives web3D a geographic coordinate specification. The GeoVRML Working Group of the Web3D Consortium has developed tools and recommended practices for the representation of geographical data using the web3D. The group's goal is to enable geo-referenced data, such as maps and 3D terrain models, to be viewed over the web with conventional browsers.

Resources
Geo-spatial example scenes in VRML, X3D and XHTL formats are available at the Web3D Consortium's site www.web3d.org/x3d/content/examples/GeoSpatial

The 3D Metanet Atlas Platform ("3MAP", www.ping.com.au) is an initiative to represent the Earth in it's entirety, giving users the ability to navigate (or zoom down) to the planet's surface for finer detail of any geographic region, or further still, rural setting—given existing data linked on the Web for that specific setting. Standards utilized for the platform include that of the Open GIS Consortium, Java and X3D/VRML. User navigation is limited (as unconstrained movement in 3D environments often disorients the user) to four modes:

- Orbital – examine or rotate the globe from space
- Satellite/map – "bird's eye view" looking down
- Flight – above land or cityscape with view angle control
- Walk – across land or through cityscape, in an enhanced "target lock facility"

An added user interface component allows the user to dynamically query content or filter out unwanted information.

5.3.3 Navigation and Way-finding Guidelines

Norm Vinson of the National Research Council's Institute for Information Technology had published the notable work "Design Guidelines for Landmarks to Support Navigation in Virtual Environments"[4] In it, he addresses user difficulty in gaining and maintaining bearings and contexts by promoting proper landmark provision.

The guidelines are not only applicable to aiding user navigation of geographic content. Vinson writes, "As computers become more powerful and 3D visualization more common, fly-throughs of networks (electronic or biological) and complex molecular structures, simulations, and data visualizations will also pose navigational problems." Following are only excerpts of core elements from the paper, where individual guidelines may not be clearly elaborated; refer to the original for clarity. Vinson makes constant use of the acronym VE, which refers to Virtual Environment.

Learning about an Environment

Newcomers to an environment rely heavily on landmarks as points of reference. As experience with the environment increases, navigators acquire route knowledge that allows them to navigate from one point in the environment to another. Route knowledge is acquired and expanded by associating navigational actions to landmarks.

Guideline 1: It is essential that the VE contain several landmarks.

Generally, additional experience with the environment increases the representational precision of route distances, and of the relative orientations and positions of landmarks. Additional experience may also transform the representation from route knowledge to *survey knowledge*. Survey knowledge is analogous to a map of the environment, except that it does not encode a typical map's top-down or bird's-eye-view perspective. Rather survey knowledge allows the navigator to adopt the most convenient perspective on the environment for a particular task. Survey knowledge acquired through navigational experience also incorporates route knowledge. In comparison to route knowledge, survey knowledge more precisely encodes the spatial proprieties of the environment and its objects.

Landmark Types and Functions

To include landmarks in a VE, one must know what constitutes a landmark. In his seminal work on urban planning and cognitive maps, Kevin Lynch found that people's cognitive

[4] Proceedings of CHI '99, Pittsburgh, PA., May, 1999. Available online through portal.acm.org by search.

maps generally contained five types of elements: *paths, edges, districts, nodes,* and *landmarks*.

Guideline 2: Include all five types of landmarks in your VE:

Types	*Examples*	*Functions*
Paths	Street, canal, transit line	Channel for navigator movement
Edges	Fence, river	Indicate district limits
Districts	Neighborhood	Reference point
Nodes	Town square, public bldg.	Focal point for travel
Landmarks[5]	Statue	Reference point into which one does not enter

Table 9: Five landmark types.

Landmark Composition

It is important to include objects intended to serve as landmarks in a VE. However, it is also important that those objects be designed so that navigators will choose them as landmarks. There are two issues regarding the way in which landmarks should be constructed. One issue relates to the landmark's physical features. The other issue relates to the ways in which landmarks should be distinctive.

Guideline 3: Make your (manmade) landmarks distinctive with features contributing to memorability of both object and location:

- Significant height
- Expensive building materials & good maintenance
- Complex shape
- Free standing (visible)
- Bright exterior
- Surrounded by landscaping
- Large, visible signs
- Unique exterior color, texture

Guideline 4: Use concrete objects, not abstract ones, for landmarks.

A study of VE landmarks also suggests that memorable landmarks increase navigability. Landmarks consisting of familiar 3D objects, like a model car and a fork, made the VE easier to navigate. In contrast, landmarks consisting of colorful abstract paintings were of no help. It was felt that the 3D objects were easier to remember than the abstract art and that this accounted for the difference in navigability.

In a natural environment, any large manmade object stands out. Accordingly, experts in orienteering (natural environment navigation) relied most on manmade objects as cues when navigating. They also used land contours and water features. However, they tried not

[5] Author's footnote: Note that Lynch refers to these items as "elements" and reserves a specific meaning for the term "landmark".

to rely on vegetation since it is a rapidly changing, and therefore unreliable, feature in natural environments.

Landmarks in Natural Environments

Manmade Items	Land Contours	Water Features
roads	hills	lakes
sheds	slopes	streams
fences	cliff faces	rivers

Table 10: Sample landmark items.

Guideline 5: Landmarks should be visible at all navigable scales.
Consider environment scales that differ from that of a city. For instance, on a larger scale, a cognitive map of a country could have cities themselves as landmarks. It is not unusual for a user to have the ability to view a VE at different scales by "zooming in" or "zooming out". In such cases, it is important for the designer to provide landmarks at all the scales in which navigation takes place.

It is important to remember that the distinctiveness of an object is a crucial factor in its serving as a landmark.

Guideline 6: A landmark must be easy to distinguish from nearby objects and other landmarks.

Guideline 7: The sides of a landmark must differ from each other.
Objects intended to serve as landmarks must be distinctive in several ways. First, they must be distinctive in regard to nearby objects. Second, a landmark must be easy to distinguish from other landmarks, especially nearby ones. Otherwise, a navigator could confuse one landmark with another, and, as a result, select the wrong navigational action (e.g. make a wrong turn). Third, the sides of each landmark should differ from one another. These differences can help navigators determine their orientation.

Guideline 8: Landmark distinctiveness can be increased by placing other objects nearby.

Guideline 9: Landmarks must carry a common element to distinguish them, as a group, from data objects.
Consider VEs whose features are constrained by the underlying data, such as the human circulatory system. Although some of the objects in these VEs can serve as landmarks, it is possible to further assist navigation by augmenting the VE with additional objects that only function as landmarks. However, navigators must easily recognize these objects as landmarks and realize that they are *only* landmarks.

Combining Paths and Landmarks: Landmark Placement

Guideline 10: Place landmarks on major paths and at path junctions.
Memorability of a building and its position was also affected by the building's location in

the environment. Memorability is enhanced when the building is located on a major path or at a path junction.

Building Positions Contributing to Memorability:

- Located on major path
- Visible from major road
- Direct access from street (esp. no plaza or porch)
- Located at important choice points in circulation pattern

Minimize distortions in cognitive maps, while being aware that the term "cognitive map" is misleading in that it suggests that mental representations of environments are very much like images. In reality, our cognitive maps contain many features that are not image-like. Cognitive maps contain categorical and hierarchical structures and many spatial distortions, some of which cannot even be represented in an image.

Using a Grid

Guideline 11: Arrange paths and edges (see Guideline 2) to form a grid.

Guideline 12: Align the landmarks' main axes with the path/edge grid's main axes.

Guideline 13: Align each landmark's main axes with those of the other landmarks.

To minimize the distortions, the designer must create a VE that induces a hierarchical representation whose districts form a grid. A consequence of the grid's spatial regularity is that the spatial relationships between districts are a good approximation of the spatial relationships between objects in those districts.

5.4 Modality Complementarity

Cross-modal interactions are characteristic of normal perception. Experiencing information in one sensory modality while having congruent, supporting information in another assists us in carrying out our most mundane tasks effectively. Stating such would not be hard to find general agreement on. What might surprise people are some findings by cognitive and neuroscientists about sensory "processing", to use a computational paradigm expression. While the sensory systems are thought of as being distinct, there are indications that we have our "wires crossed," though in a beneficial way. For example, PET scans have revealed to researchers that visual processing is not only involved in tactile perception, but is necessary for optimal performance in tactile orientation discrimination. The visual cortex is intimately involved in processing certain kinds of tactile information (Sathian, 2001).

Regardless, there are strong arguments against a seemingly synesthesic mixing or synergistic "whole greater than its parts" sensory processing phenomenon at work in the human mind. Translation of a full meaning equivalence from one sensory modality through to another is debatable. It is unlikely that the combination of hearing and touching can ever replace the sensation of seeing something. Loomis (2003) observes:

> If the sensory bandwidth of the substituting sense (or senses) is grossly inadequate, it is simply not possible to carry out the desired function. For example, the informational demands of driving a car are not likely to be met through a combination of audition and touch.

When attempting to optimize the interaction of user with interface, online media content designers have an increasing variety of modalities they can apply. Information is presentable through vision, touch, sound, or preferably, a combination of these sensory modalities. We are challenged with the task of choosing the most apt combination of information channels to communicate a specific message, or, the context for users to forage for what they consider relevant, in the manner they are best suited to. Sheer efficiency, of course, is not the only measure of success. Short of established guidelines for sensory modality transformation, we urge the designer of any kind of online content to employ reiterative, inclusive user testing.

5.5 Appendix: Code Sample for adding both Immersion Web plug-in and user environment detection to HTML.

(Available from Immersion Developer site www.immersion.com/developer/technology/tutorials/index.php, log in required, within the tutorial "TouchSense in JavaScript - Web Tutorials: Detecting Immersion Web Plugin".)

```
...
<OBJECT
classid="CLSID:5CA42785-ABC3-11d2-9F81-00104B2225C5"
codebase="#version=3,2,3,0"
id="ImmWeb"
height="0" width="0">
 <EMBED type="application/x-Immersion-Web-Plugin"
 name="ImmWeb" hidden="true" src="myfile.txt">
</OBJECT>

<SCRIPT language="JavaScript">
<!--hide this script from non-JavaScript-enabled browsers

var imm_onTouch = false;

var imm_usrAgt = navigator.userAgent.toLowerCase();
var imm_isIE = (navigator.appName == "Microsoft Internet Explorer");
var imm_isNN = (navigator.appName == "Netscape");
var imm_isIE4plus = imm_isIE && (parseInt(navigator.appVersion) >= 4);
var imm_isNN45plus = imm_isNN && (parseFloat(navigator.appVersion) >= 4.5);
var imm_isNN6plus = imm_isNN && (parseFloat(navigator.appVersion) >= 5);
var imm_isNN7plus = imm_isNN6plus && (navigator.vendor == "Netscape");
var imm_isWin = (imm_usrAgt.indexOf("win")!=-1);
var imm_supportsImm = imm_isWin;
var imm_NNMinPlugin = new Array(3, 2, 3, 0);
var imm_IEDontAskOff = (IMM_getCookie("DontAskIE")!= "ON");
var imm_NNDontAskOff = (IMM_getCookie("DontAskNN")!= "ON");

if(imm_supportsImm) {
 IMM_sufficientSystem();
}

function IMM_sufficientSystem() {
// If Netscape Navigator 4.x, Netscape 7.x or Mozilla 1.x and not Netscape 6
if(imm_isNN45plus && !imm_isNN6plus || imm_isNN7plus) {
 if(navigator.plugins["Immersion Web Netscape Plugin"] &&
   IMM_sufficientNNPlugin()) {
  imm_onTouch=true;
 }
 else {    // ImmWeb plugin is not installed
  if(imm_NNDontAskOff) {
   imm_installNow = confirm("You must install the Immersion Web plugin to experience
```

```
tactile sensations on this web page. Install now?");
    if(imm_installNow)  {
                                    if(imm_isNN7plus)
                                                alert("Sorry, Gecko based browsers are not currently supported.");
                           else
                                       self.location.href="download_NN_plugin.html";
    } else  {
      imm_askAgain = confirm("You have chosen not to install the Immersion Web plugin at this time. Would you like to be reminded again later?");
      if(!imm_askAgain)  {
        IMM_setCookie("DontAskNN", "ON", "month");
      }
      else  {
        IMM_setCookie("DontAskNN", "ON", "now");
      }
    }
  } else {
                        alert("You're cookie setting is inhibiting you from downloading the Immersion Web Plugin. If you have changed your mind and would like to install, please click on the appropriate link above.");
            }
 }
}

// If Internet Explorer 4 or greater
if(imm_isIE4plus)  {
 if(document.ImmWeb.ImmWebInstalled)  {
  imm_onTouch=true;
 }
 else  {    // ImmWeb plugin is not installed
  if(imm_IEDontAskOff)  {
    imm_installNow = confirm("You must install the Immersion Web plugin to experience tactile sensations on this web page. Install now?");
    if(imm_installNow)  {
      self.location.href="http://www.immersion.com/plugins/download_IE_plugin.html"
    }
    else  {
      imm_askAgain = confirm("You have chosen not to install the Immersion Web plugin at this time. Would you like to be reminded again later?");
      if(!imm_askAgain)  {
        IMM_setCookie("DontAskIE", "ON", "month");
      }
      else  {
        IMM_setCookie("DontAskIE", "ON", "now");
      }
    }
  }
 }
}
}   // end function IMM_sufficientSystem()
```

```
function IMM_sufficientNNPlugin()  {
            // Assumes Netscape plugin description = "Immersion Web Plugin a.b.c.d" where
a.b.c.d = version
            var needDownload = false;
            var installedVersion = navigator.plugins["Immersion Web Netscape
Plugin"].description.split(" ")[3].split(".");
            // Need Download if installed plugin version < minimum plugin version
            for(index=0; index<4; index++)  {
        if(installedVersion[index] > imm_NNMinPlugin[index])  {
                        break;
                  }
                  else if(installedVersion[index] < imm_NNMinPlugin[index])  {
                        needDownload = true;
                        break;
                  }
            }
            return (!needDownload);
}   // end function IMM_sufficientNNPlugin()


//****** Cookie utility functions ******//

function IMM_setCookie(name, value, expire) {
 var today = new Date();
 var expiry = new Date();

 switch(expire)  {
   case "day":
     expiry.setTime(today.getTime() + 1000*60*60*24*1);
     break;
   case "week":
     expiry.setTime(today.getTime() + 1000*60*60*24*7);
     break;
   case "month":
     expiry.setTime(today.getTime() + 1000*60*60*24*30);
     break;
   case "year":
     expiry.setTime(today.getTime() + 1000*60*60*24*365);
     break;
   default:
     expiry.setTime(today.getTime());
 }
 document.cookie = name + "=" + escape(value) + "; path=/; expires=" +
expiry.toGMTString()
}

function IMM_getCookie(Name) {
 var search = Name + "="
 if (document.cookie.length > 0) {                // if there are any cookies
   offset = document.cookie.indexOf(search)
```

```
  if (offset != -1) {                          // if cookie exists
    offset += search.length                         // set index of beginning of value
    end = document.cookie.indexOf(";", offset)      // set index of end of cookie value
    if (end == -1)
      end = document.cookie.length
    return unescape(document.cookie.substring(offset, end))
  }
 }
}

// stop hiding -->
</SCRIPT>

...
```

References

Americans with Disabilities Act, US Department of Justice. 2003 from http://www.usdoj.gov/crt/ada/adahom1.htm

Audio Description Associates Inc. (2004). Training. Retrieved March 9, 2004 from shttp://www.audiodescribe.com/training1.html.

Avery, M.B. (2001). The role of the health care interpreter. Retrieved Feb. 27, 2004 from http://www.ncihc.org/workingpapers.htm.

Babich, F., & Vitez, M., (2000). A novel wide-band audio transmission scheme over the Internet with a smooth quality degradation. ACM SIGCOMM Computer Communication Review. 30(1). 37 – 44.

Barstow, C., McKell, M., Rothberg, M., & Schmidt, C. (2002, June 27). *IMS Guidelines for Developing Accessible Learning Applications*. Retrieved January 12, 2004, from http://www.imsglobal.org/accessibility/accv1p0/

Blacking, J. (1973). *How musical is man?* Seattle: University of Washington Press.

Brossard, N. (1987). *Le désert mauve : roman*. Montréal: L'Hexagone.

Brown, S. E. (2002). *A Celebration fo Diversity: An Annotated Bibliography About Disability Culture* (No. Second Edition). Las Cruces: Institute on Disability Culture.

Bussemakers, M. P., & de Haan, A. (2002). *When it sounds like a duck and it looks like a dog . . . Auditory icons vs earcons in multimedia environments.* Paper presented at the Eighth International Conference on Auditory Display, ICAD, Kyoto, Japan.

Canadian Radio-television and Telecommunications Commission. (2004). Broadcasting Decision CRTC 2004-21.

Chisolm, W., Vanderheiden, G., & Jacobs, I. (1999, May 5). *Web Content Accessibility Guidelines 1.0*. Retrieved January 12, 2004, from http://www.w3.org/TR/WCAG10/)

Dardailler, D., Palmer, S. B., & McCathieNevile, C. (2002, October 3). *XML Accessibility Guidelines*. Retrieved March 1, 2004, from http://www.w3.org/TR/xag

Freed, G., Rothberg, M., & Wlodkowski, T. (2003, January). *Making Educational Software and Web Sites Accessible: Design Guidelines Including Math and Science Solutions*. Retrieved January 12, 2004, from http://ncam.wgbh.org/cdrom/guideline/

IBM. (2002). *IBM Software Accessibility Checklist*. Retrieved March 1, 2004, from http://www-3.ibm.com/able/guidelines/software/accesssoftware.html

Jacobs, I., & Brewer, J. (1999, August 4). *Accessibility Features of CSS*. Retrieved March 1, 2004, from http://www.w3.org/TR/CSS-access

Jacobs, I., Brewer, J., & Dardailler, D. (2000, November 8). *WAI Resource: HTML 4.0 Accessibility Improvements*. Retrieved March 1, 2004, from http://www.w3.org/WAI/References/HTML4-access

Jacobs, I., Gunderson, J., & Hansen, E. (2002, December 17). *User Agent Accessibility Guidelines 1.0*. Retrieved March 1, 2004, from http://www.w3.org/TR/2002/REC-UAAG10-20021217/

Jacobs, I., & Koivunen, M.-R. (1999, September 21). *Accessibility Features of SMIL*. Retrieved March 1, 2004, from http://www.w3.org/TR/SMIL-access/

Massof, R. (2003). *Auditory assistive devices for the blind.* Paper presented at the International Conference on Auditory Display, Boston.

McCathieNevile, C., & Koivunen, M.-R. (2000, August 7). *Accessibility Features of SVG*. Retrieved March 1, 2004, from http://www.w3.org/TR/SVG-access/

National Information Standards Organization. (2002, March 6). *Specifications for the Digital Talking Book*. Retrieved March 1, 2003, from http://www.niso.org/standards/resources/Z39-86-2002.html

NC State University, & The Center for Universal Design. (1997, April 1). *The Principles of Universal Design*. Retrieved January 28, 2004, from http://www.design.ncsu.edu:8120/cud/univ_design/principles/udprinciples.htm

Peters, S. (2000). Is there a disability culture? a syncretisation of three possible world views. *Disability & Society, 15*(4), 583-602.

Purcell, P. (2002, June). Extreme Interfaces. *i3 Magazine: The European Network for Intelligent Information Interfaces, 12,* 35-37.

Randel, D. M., & Apel, W. (1986). *The New Harvard dictionary of music*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Roig, X. C. (2004). Mom, I wanna be an artist. *Translation Journal, 8*(1).

Schwerdtfeger, R. S. (2000, August 24). *IBM Guidelines for Writing Accessible Applications Using 100% Pure Java (tm)*. Retrieved January 12, 2004, from http://www-306.ibm.com/able/guidelines/java/snsjavag.html

Sign Language Center.*About Deaf Culture*. Retrieved March 1, 2004, from http://www.signlanguagecenter.com/AboutDeafCulture.html

Stravinsky, I. (1936). *Chronicle of My Life*. London: Gollancz.

Taylor, M. M., Lederman, S. J., & Gibson, R. H. (1973). Tactual perception of texture. In E. Carterette & M. Friedman (Eds.), *Handbook of Perception* (Vol. III, pp. 251-272). NY: Academic Press.

The DAISY Consortium. (2001, February 28). *DAISY 2.02 Specification*. Retrieved March 1, 2004, from http://www.daisy.org/publications/specifications/daisy_202.html

Thornton, C., Kolb, A., Gemperle, F., & Hirsch, T. (2003). *Audiocentric interface design: a building blocks approach.* Paper presented at the International Conference on Auditory Display, Boston.

Titon, J. T. (1996). *Worlds of music : an introduction to the music of the world's peoples* (3rd ed.). New York, NY: Schirmer Books.

Treviranus, J., McCathieNevile, C., Jacobs, I., & Richards, J. (2000). *Authoring Tool Accessibility Guidelines 1.0*. Retrieved March 1, 2004, from http://www.w3.org/TR/ATAG10/

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8*(4), 211-221.

Walker, B. N., & Kramer, G. (1996). *Mappings and metaphors in auditory displays: an experimental assessment.* Paper presented at the Third International Conference on Auditory Display, ICAD, Palo Alto, CA.

Walker, B. N., & Lindsay, J. (2003). *Effect of beacon sounds on navigation performance in a virtual reality environment.* Paper presented at the Ninth International Conference on Auditory Display, ICAD, Boston, MA.

Weeks, J. (2002). Bringing ballet to the blind. *Dance Magazine*. 23-24.

Weiss T., Whiteley C., Treviranus, J., and Fels, D.I. (2001). PEBBLES: A Personal Technology for Meeting Educational, Social and Emotional Needs of Hospitalized. Children. Personal and Ubiquitous Computing. 5. 157-168.