Model Based Estimation for Multi-Modal User Interface Component Selection

L. Coetzee, I. Viviers, E. Barnard

Meraka Institute
(African Advanced Institute for Information & Communication Technology),
CSIR, P.O. Box 395, Pretoria, 0001, South Africa
lcoetze1@csir.co.za iviviers@csir.co.za ebarnard@csir.co.za

Abstract

Multi-modal human computer interfaces are becoming increasingly widespread, building on more capable and affordable devices along with advances in helper applications utilising advanced pattern recognition. These interfaces promise to improve human-computer interaction, not only for fully-abled people, but also for persons with disabilities. It is currently unclear how to map available multi-modal components to a specific user profile in a systematic fashion, especially when the abilities, perceptual preferences and literacy level of the user should be taken into account. This paper presents one approach to develop a cost-based model which can be used to derive appropriate mappings for specific user profiles. The model is explained through a number of small examples, where after the usage and benefits of the model are illustrated using a variety of different profiles. It is shown that the model is effective in identifying important multi-modal components for various user profiles.

1. Introduction

Advanced pattern recognition utilising a variety of different modalities such as speech recognition, gesture recognition and touch screens are changing the landscape of human-computer interaction (HCI). This change in landscape is to the benefit of all users, especially persons with disabilities, as the availability and incorporation of other modalities in a computer environment assist in breaking through accessibility constraints. Coetzee and Barnard [1] showed how advanced pattern recognition can break the access barrier and improve the quality of lives of persons with disabilities. Assistive technologies (such as a screen reader which voices out appropriate text from the computer) are often highly dependent on sophisticated pattern recognition as applied to a multi-modal environment.

The availability of various modalities to enhance interaction is not equally beneficial for all users, as a user's ability to actually utilise and interact through a modality (with output represented through a variety of content output formats and associated components and input as represented through various input mechanisms and devices) depends on the user's abilities (e.g. a user can see and hear) as well as a number of other factors including his perceptual preferences and literacy level. In addition, the lack of a specific ability often impacts on other aspects of a person's ability to optimally interface with a computer.

This raises the question of how a suitable HCI configuration for a user with various abilities and a specific perceptual preference, who utilises one or more assistive technologies, can be determined.

This paper presents one approach to determine which components in a multi-modal environment are important to a user, thus leading to an enhanced interaction experience.

The next section (Section 2) provides some background information regarding ability based modelling in a multi-modal computer based context. Section 3 introduces a variety of factors that need to be considered when attempting to define a configuration for a specific user profile. This is followed by a section containing information of the various technological representations of input and output modalities (Section 4). Section 5 presents a cost based estimation model that provides insight into the identification of the most important components for each user profile, while Section 6 illustrates the application of the cost model to real world examples. A conclusion is presented in Section 7.

2. Background

Substantial research has been conducted in the field of multi-modal interaction as associated with HCI.

Oviatt [2] investigates the use and benefits of multi-modal interfaces. Her aim is to provide users with a choice of switching to a better suited modality, depending on the specifics of their abilities, the task and the usage conditions. Oviatt presents the results of different studies which analysed the benefit of using multiple modalities for inputs (e.g. accented voice input combined with an alternate input – such as pen input). The results indicate that the use of multiple modalities lead to improved performance. Oviatt points out that further research is required in multi-modal interfaces that are capable to strategically adapt based on the user profile.

Kawai et.al. [3] present an architecture of a user interface toolkit that supports the flexibility required by persons with disabilities as well as fully-abled people. The toolkit is based on the premise of the user being able to select his/her preferred modalities.

Blattner and Glinert [4] highlight the fact that even though the strengths and weaknesses of each single modality for interaction are well understood, the general problem of integrated multi-modal systems are yet to be understood to the same level.

User modelling plays an important role within user-adaptive systems. Kobsa [5] presents a review on the development of numerous generic user modelling systems. One of the services of such systems can include the representation of assumptions about one or more types of user characteristics of individual users. Personalisation of systems benefits both users and providers of services and therefore user modelling tools will continue to play an important role in computer systems.

Even though the utilisation of multiple modalities to break down the access barrier has been addressed by several researchers, specific models that allow for the choice of a suitable configuration of modalities per user profile has not been pub-

lished. It is unclear which combinations of modalities are best suited for a specific user profile and how important the availability of a given modality is for such a user. The research presented in this paper seeks to address these important issues.

The following section presents factors required in building a user model which can be used to define a cost based model.

3. User Profile

Most computer applications are built around the concept of the "average user", in order to meet the needs of the largest part of the population, without requiring adaptation of the interface. This however, does not allow for differences between users which lead to exclusion and barriers in interfacing with the computer.

Individuals differ in many dimensions. For instance, persons with disabilities may have requirements quite different from those of the average user and use non-standard assistive technologies (e.g. a screen reader that voices out appropriate text from a computer user interface) to overcome the interaction barrier. This section presents a number of factors that need to be considered for each individual user when understanding that the *average approach* is not always sufficient to ensure acceptable interaction with a computer system.

3.1. Abilities and modalities

Each individual has different abilities which impact on how that individual interfaces with a computer through the modalities provided. A modality can be described as the sense through which a human can receive output from a computer (defined as an output modality), and the way that a computer can receive input from a human – defined as an input modality. Note that both input and output modalities may require specialised sensors or devices, and possibly also helper applications such as an automatic speech recogniser (e.g. for the entering of command and control commands on the computer). This software and hardware combination is commonly referred to as assistive technologies.

It is useful to think of an individual in terms of his abilities (i.e. what he can do) rather than his disabilities (that which he cannot do). For example: in the case of a person with a physical disability, the availability and use of an assistive technology (such as an eye tracker) will allow the user to interact with the computer system. With an "average" interface, this disability would have prevented him from navigating with a traditional mouse pointer. However, through the assistive technology, the user still has the ability to move the pointer. In essence, different modalities are used, but interaction still occurs.

Table 1 presents a list of abilities associated with generating output to a user. It should be noted from Table 1 that specific assumptions can be made with regard to a user's ability. One such assumption is that for a user to have the ability to understand Braille, he must be able to feel.

Table 2 presents a list of abilities linked with entering input into the computer.

From Tables 1 and 2 it is clear that individuals can have vastly different profiles based on their respective abilities alone. However, these abilities are only one part of the bigger picture associated with an individual. Section 3.2 presents user preferences, which also need to be taken into account when attempting to model a user.

Ability associated with output	**Assumptions**
Can See	None
Can Hear	None
Can Read	User can see
Can Read (simplified text)	User can see
Can Understand South African Sign Language	User can see
Can Feel	None
Can Understand Braille	User can feel
Can Lip Read	User can see

Table 1: *List of abilities linked with output.*

Abilities associated with input	**Method**
Can Talk	Regular voice combined with automatic speech recognition
Can Click	Input switch or dwell mode on pointer
Can Move pointer	Possible through standard mouse or assistive technology such as eye tracker
Can Utilise Keyboard	Possible through standard keyboard or assistive application such as on-screen keyboard
Can Make Physical Movement	Possible through sensors (e.g. gloves, switches and video cameras)

Table 2: *List of abilities linked with input.*

3.2. Perceptual preferences

In addition to the tangible abilities mentioned above, each individual's unique makeup is further defined by a number of other factors. These factors impact on what the user's preferences are in terms of internalising presented content. One such factor is the individual's *perceptual preference* which reflects his natural style. Perceptual preference indicates the preferred means by which individuals extract and internalise information through the use of their five senses. The five senses namely sight, hearing, touch, smell and taste can be translated into different perceptual pathways (or modalities).

The perceptual learning styles model developed by Russell French, Daryl Gilley, and Ed Cherry [6, 7, 8] in the late 1970s and early 1980s defines seven perceptual pathways namely print, aural, interactive, visual, haptic, kinesthetic and olfactory.

Print refers to seeing printed or written words; *aural* refers to listening, while *visual* refers to seeing visual depictions. *Interactive* refers to verbal interaction, while *haptic* refers to the sense of touch or grasp. *Kinaesthetic* refers to the whole body movement, and *olfactory* refers to the sense of smell and taste. This research suggests that information should be presented in different ways to engage individuals with different preferences.

The research presented in this paper focused on four perceptual preferences (visual, aural, read/write and kinaesthetic) as presented in the VARK model as developed by Fleming [9]. These preferences (and their associated impacts) are typically associated with learning. However, the authors argue that these preferences also indicate general perceptual preferences in in-

terfacing with a computer.

Table 3 presents perceptual preferences according to the VARK model.

Perceptual Preference	Description
Visual (V)	Individual prefers pictures, graphs and diagrams.
Aural (A)	Individual prefers spoken words
Read/Write (R)	Individual prefers reading and writing texts
Kinaesthetic (K)	Individual prefers to move his/her body and manipulate things with his/her own hands

Table 3: *Perceptual preferences.*

It is clear from the above that individuals differ in their preferred preferences which impacts on the way they would prefer to interact with a computer.

3.3. Literacy level

In addition to the above-mentioned dimensions, individuals' literacy levels vary greatly. Also, not all computer users interface with a computer in their language of choice (such as their first language). An individual might be literate in a specific language but not in another. Literacy levels are also influenced by domain knowledge. For practical purposes, various categories of literacy could therefore be defined. For a given language, these categories include:

- **Illiterate**: A person is completely illiterate and cannot read or write.
- **Cultural**: A person does not understand the idioms, icons, expressions and role models associated with a language.
- **Grammatical**: A person tends to use grammar incorrectly.
- **Second language**: Literate in mother tongue, but generally less fluent in the language of the interface.
- **Deaf**: Literate in Sign Language, but not necessarily literate in a spoken language.

Literacy levels can also be influenced by disability, for example a Deaf person has difficulty to naturally acquire the reading and writing skills associated with an oral language to the same high level attained by persons with normal hearing. This can have the consequence that this individual is more comfortable interacting with a computer using a simplified version of text [10].

The combination of a user's preference, a specific set of abilities with the addition of literacy creates a complicated picture of a user. This picture is further complicated when the various technological components are introduced.

The next section describes a set of components associated with input and output on a computer based system.

4. Technologies

Various technology components or devices (associated with various modalities) have been developed to assist with human computer interfacing. These technology elements enable users with specific abilities (or lack of abilities) and specific perceptual preferences to interact in an appropriate way.

To make use of the available output modalities, content elements should be in relevant formats. For example, for an individual with an audio preference, content elements such as music, sounds and Text-To-Speech as audible output can be important, while these might not be important to an individual with a visual preference. Helper applications can be used to transform content from one format to another e.g. Text-To-Speech synthesis which transforms text into an audible format. Table 4 presents a list of applicable content formats used for output.

Content formats used by output modalities	Description
Image	An image or representation of an object or event
Video	A recorded video file, the visual component
Animation	Simulation of motion by presenting a series of pictures, the visual component
Sign Language	Text or audio presented by Sign Language Interpreter
Symbols	Small picture that represents something else by association
Icons	A small image or abstract representation of an object or event
Text-To-Speech synthesis	Text synthesised as audible output
Audio	Audible sound component from video files
Music	Audible music sounds
Sound	Audible sounds
Earcons	Audible abstract sounds
Text	Printed words
Simple Text	Printed text converted to a simplified version
Captions	Printed text captions
Braille	Text output onto a Braille display
Texture	Display pixels converted to texture maps
Tactile	Events represented through force feedback
Vibrations	Vibration alerts
Sound Vibrations	Sound frequencies converted to vibrations
Heat	Heat or the absence of heat (cold) signals or alerts

Table 4: *List of output content formats.*

Similarly, Table 5 presents a list of identified input devices and possible helper applications.

When considering the conjunction of the various dimensions as described in Section 3 and the variables introduced in Tables 4 and 5 it is clear that it would be difficult to map a user profile to a sensible configuration of output formats and input mechanisms. What is needed is a model to aid in the configuration determination. The next section introduces such a model with the aim to simplify the configuration process.

Input Device	Description
Microphone	Requires automatic speech recogniser to create character string
Joystick	Sends pointer events
Eye Tracker	Requires helper application to send pointer events
Camera	Requires helper application to create pointer events
Mouse	Sends pointer events
Head Pointer	Requires helper application to send pointer events
Touch Screen	Sends pointer events
Keyboard	Sends character string
Stylus	Sends pointer events
Switches	Sends pointer events

Table 5: *List of input devices.*

5. Mapping model

In the preceding sections we have introduced the various components (consisting of the user's abilities, the user's style and literacy as well as possible input and output components) which influence the possible configurations for a user. It is clear from the large number of variables that it is not straightforward to determine which configuration of possible input and output components are most suited for a specific user profile, especially when availability constraints are taken into consideration. What is needed is an approach to model the variables which would result in adaptable configurations for each user. This sections contains such a mathematical analysis and model which allows for the prediction of configurations.

5.1. Cost Model

Let $\vec{p_i}$ be a vector of real values scaled between 0 and 1 of length n, where each element in the vector represents the user's abilities according to Tables 1 and 2.

Similarly, let the diagonal matrix S_j of size $n \times n$ contain real values scaled between 0 and 1 to represent the perceptual preferences according to Table 3. The four basic representations of S_j correspond to each of the perceptual preferences, and these can be weighted and combined for individuals with mixed preferences.

Combining $\vec{p_i}$ and S_j as presented in Equation 1 provides a vector $\vec{w_k}$ which represents an adjusted user profile as based on his perceptual preferences.

$$\vec{w_k} = S_j \times \vec{p_i}^T \tag{1}$$

Let the matrix D of size $n \times m$ (where n is the number of modelled user abilities and m the number of modelled available input and output components) represent a matrix of "dominant" user abilities as required for a specific modality. (The concept of "domination" is explained below.)

Using D and $\vec{w_k}$ as is presented in Equation 2 provides us with a cost estimation of suggested components to be used per adjusted user profile.

$$\vec{c_l} = D \times \vec{w_k}^T \tag{2}$$

Larger values in $\vec{c_l}$ thus indicates which are the more important components for a specific user profile.

5.2. Application of Cost Model

Equation 2 provides a cost vector indicating important components for a specific user profile. The following simplified example illustrates the concept.

Let $\vec{p_i}$ represent the abilities *Can See*, *Can Hear* and *Can Read*. A fully-abled user can thus be represented as in Equation 3 while a user that can only hear is represented as in Equation 4.

$$\vec{p}_{Fully-able} = [0.3, 0.3, 0.3] \tag{3}$$

$$\vec{p}_{Can\ only\ hear} = [0.0, 1.0, 0.0] \tag{4}$$

The perceptual preferences for a visually biased user can be represented as in Equation 5, while an aural bias can be represented as in Equation 6.

$$S_{Visual} = \begin{bmatrix} & Can\ See & Can\ Hear & Can\ Read \\ Can\ See & 0.6 & 0.0 & 0.0 \\ Can\ Hear & 0.0 & 0.2 & 0.0 \\ Can\ Read & 0.0 & 0.0 & 0.2 \end{bmatrix} \tag{5}$$

$$S_{Aural} = \begin{bmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{bmatrix} \tag{6}$$

The adjusted user profile for a fully-able user as calculated through Equation 1 for a visual sense preference will be:

$$\begin{aligned} \vec{w}_{Fully-able\ with\ visual\ preference} &= \begin{bmatrix} 0.6 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{bmatrix} \\ &\times [0.3, 0.3, 0.3]^T \\ &= [0.18, 0.06, 0.06] \end{aligned} \tag{7}$$

while a fully-abled profile with an aural bias will be:

$$\vec{w}_{Fully-able\ with\ aural\ preference} = [0.06, 0.18, 0.06]. \tag{8}$$

A hearing-only profile adjusted according to the visual and aural bias results in:

$$\vec{w}_{Hearing\ only\ with\ visual\ preference} = [0.0, 0.2, 0.0] \tag{9}$$

and

$$\vec{w}_{Hearing\ only\ with\ aural\ preference} = [0.0, 0.6, 0.0] \tag{10}$$

A possible dominant matrix D representing abilities against components (for components *Text – output*, *Audio – output* and *Image – output* and abilities *Can See*, *Can Hear* and *Can Read*) is:

$$D = \begin{bmatrix} & Can\ See & Can\ Hear & Can\ Read \\ Text & 0 & 0 & 100 \\ Audio & 0 & 100 & 0 \\ Image & 100 & 0 & 0 \end{bmatrix} \tag{11}$$

From Equation 11 we see that a weight of 100 has been assigned to the *Can Read* ability for the *Text* output component, while the other abilities have been assigned a weighting of zero. Similarly, a weighting of 100 is assigned for the *Can Hear* ability for the *Audio* component and a weighting of 100 is assigned

for the *Can See* ability for the *Image* output component. The dominance aspect of the matrix D is illustrated in the first row. Even though a user would require the ability to see to utilise text, only the ability *Can Read* is activated (as it is implicit that a user must be able to see, to be able to read).

Utilising Equation 2 by applying (7), (8), (9) and (10) results in cost vector $\vec{c}_l$ for the different profile examples. The first cell in $\vec{c}_l$ represents the importance of the *Text* to the user, the second cell the importance of *Audio* and the last the importance of *Image*.

$$\vec{c}_{Fully-able\ visual\ preference\ cost} = [6.0, 6.0, 18.0] \quad (12)$$

$$\vec{c}_{Fully-able\ aural\ preference\ cost} = [6.0, 18.0, 6.0] \quad (13)$$

$$\vec{c}_{Hearing\ only\ with\ visual\ preference\ cost} = [0.0, 20.0, 0.0] \quad (14)$$

$$\vec{c}_{Hearing\ only\ with\ aural\ preference\ cost} = [0.0, 60.0, 0.0] \quad (15)$$

The cost vectors as presented in this section allow for the identification of the appropriate HCI components for a specific user profile. It must be noted that the purpose of the cost vectors for the individual profiles is not to compare them across users, but to indicate the appropriate component selection for a specific profile. Section 6 presents some results of a more complete modelling of users in an environment with more available components.

6. Mapping Results

The application of the cost model as presented in Section 5.1 to a variety of different profiles (including all four perceptual preferences) provides an interesting perspective on appropriate components. Section 6.1 presents mapping results associated with various output profiles, while Section 6.2 presents results for input profiles.

6.1. Output

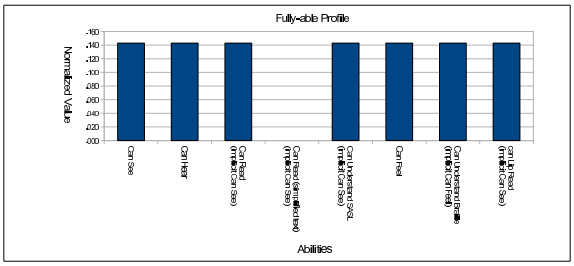

Figure 1: *Fully-abled abilities.*

Figure 1 presents a baseline profile for a person with all the abilities associated with output as presented in Section 3.1. Figure 2 presents the results of the application of the cost model.

In Figure 2 the importance of a specific output format per perceptual preference is clearly visible. A user with a visual preference would prefer content presented in a visual format (e.g. icons, symbols, video) even though the user has the ability

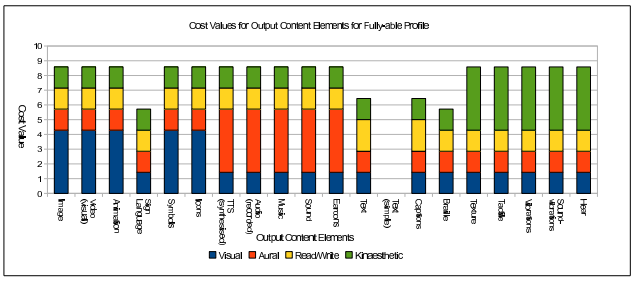

Figure 2: *Fully-abled cost model representation.*

to consume information as presented in any modality. Similarly, a user with an aural preference would prefer content presented as audio, music, sounds and earcons. Similar observations can be made for Read/Write and Kinaesthetic preferences.

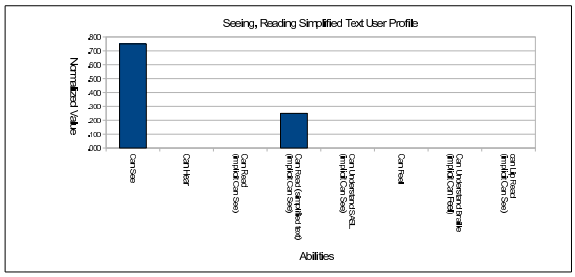

Figure 3: *Can See and Can Read Simplified Text abilities.*

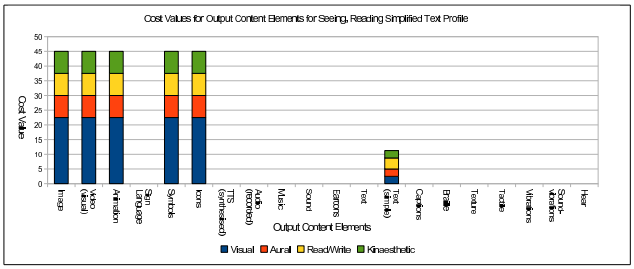

Figure 4: *Can See and Can Read Simplified Text cost model representation.*

The power and the benefit of the cost model is illustrated in Figures 3 and 4. Figure 3 presents a user profile where the user has the ability to see and only read simplified text. The model results are shown in Figure 4. Simplified text as output is important to an aural preference profile, while the visual components increase in relative importance for a visual preference.

6.2. Input

Figure 5 presents an input profile for a user who has no clear preference for any of the available inputs. Figure 6 presents the calculated cost model for this profile. Figure 6 shows the importance of having an automatic speech recognition engine for an aural perceptual. Similarly, Figure 6 shows that a Kinaesthetic profile would prefer to use motion and sensors as input.

Figure 7 shows a profile weighing the *Can Talk* ability more compared to the other presented abilities. For this profile voice

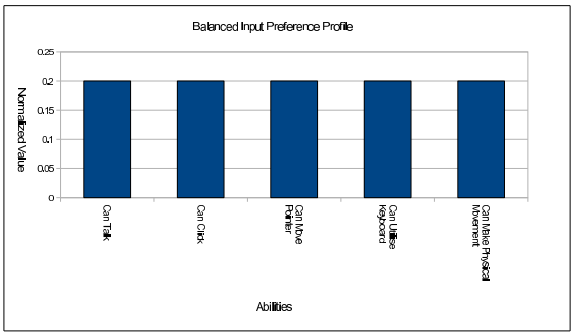

Figure 5: *Balanced input ability preference.*

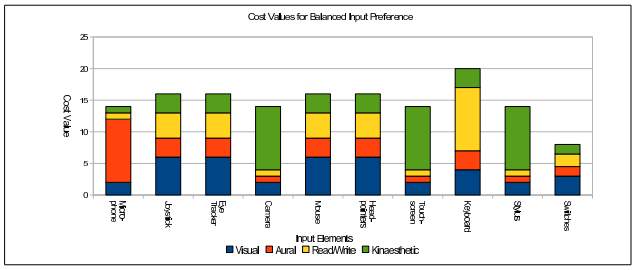

Figure 6: *Balanced input cost model representation.*

input is most important for an Aural preference, while a keyboard is the most important for a Read/Write preference.

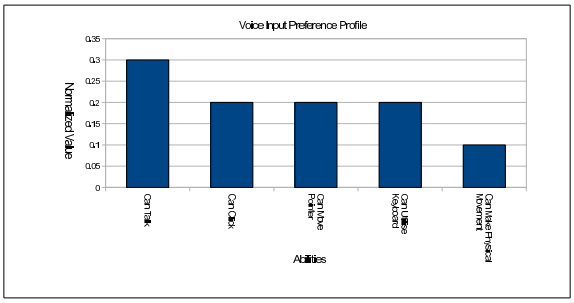

Figure 7: *Voice input ability preference.*

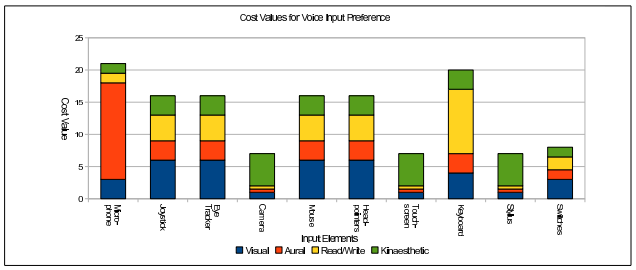

Figure 8: *Voice input cost model representation.*

This section has provided some examples of utilising the cost model to determine which elements are more important based on the preferences and abilities of selected users. The examples clearly show how a multi-modal environment should be configured per user profile. The power of the model is that it allows such configuration to happen in an automated fashion - based on knowledge of user preferences and abilities as well as the available modalities for a particular task. An application can use this model to select a preferred presentation style automatically.

7. Conclusions

Multi-modal human computer interaction which utilises pattern recognition approaches are gaining in popularity and importance. Such interfaces entail a number of advantages, especially for persons with disabilities, as they promote inclusion and the removal of barriers. The identification of appropriate modalities based on a user profile is not a trivial matter, especially when cognisance is taken of the many possible factors associated with a user (including the fact that a user might have one or more perceptual preferences).

This paper has presented an approach to model the user, taking into account his abilities, literacy level, and perceptual preferences. The presented cost model leads to interesting insights into what would be most appropriate for a specific user and provides us with the ability to automatically configure and utilise a multi-modal environment based on user characteristics.

8. References

[1] L. Coetzee and E. Barnard, "Pattern recognition in service of people with disabilities," in *Proceedings of the Fifteenth Annual Symposium of the Pattern Recognition Association of South Africa*, vol. 15, November 2004, pp. 75–80.

[2] S. Oviatt, "Designing robust multimodal systems for universal access," in *Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing: providing for the elderly*. Alcácer do Sal, Portugal: ACM, 2001, pp. 71–74. [Online]. Available: http://portal.acm.org/citation.cfm?id=564526.564546

[3] S. Kawai, H. Aida, and T. Saito, "Designing interface toolkit with dynamic selectable modality," in *Proceedings of the second annual ACM conference on Assistive technologies*. Vancouver, British Columbia, Canada: ACM, 1996, pp. 72–79. [Online]. Available: http://portal.acm.org/citation.cfm?id=228360

[4] M. Blattner and E. Glinert, "Multimodal integration," *Multimedia, IEEE*, vol. 3, no. 4, pp. 14–24, 1996.

[5] A. Kobsa, "Generic User Modeling Systems," *User Modeling and User-Adapted Interaction*, no. 11, pp. 46–63, 2001.

[6] R. L. French, "Teaching strategies and learning," 1975, department of Curriculum and Instruction, University of Tennessee, Knoxville, TN.

[7] D. V. Gilley, "Personal learning styles: exploring the individual's sensory input processes," Ph.D. dissertation, University of Tennessee, Knoxville,TN, USA, 1975.

[8] C. E. Cherry, "The measurement of adult learning styles: Perceptual modality," Ph.D. dissertation, University of Tennesse, Knoxville, TN, USA, 1981.

[9] N. D. Fleming and C. Mills, "VARK a guide to learning styles," http://www.vark-learn.com/English/index.asp (Last accessed 28 January 2008).

[10] D. Aarons and M. Glaser, "A Deaf Adult Literacy Collective," *Stellenbosch Papers in Linguistics*, no. 34, pp. 1–18, 2002.