MASTER THESIS
in
Universal Design of ICT

May 2017

Using non-speech sounds to increase web image accessibility for screen-reader users

Ratan Bahadur Thapa

Department of Computer Science

Faculty of Technology, Art and Design

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Preface

This thesis is competence for partial fulfillment of obtaining the Master in Universal Design of ICT at department of computer science, Høgskolen i Oslo og Akershus (HiOA) located in Oslo, Norway. The thesis work has been administrated and supervised by the college under the faculty of Technology, Art and Design.

I would like to thank those who made this journey possible for me. Foremost, I am extremely grateful to my mentor and supervisor **Mexhid Adem Ferati** for his endless encouragement, motivation, guidance, support, knowledge and friendship. I am indebted to him for introducing me in the field of auditory display and opening every possible door. Also, I am truly thankful to **Norun Christine Sanderson and** co-supervisor **George Anthony Giannoumis** for their valuable comments and suggestions in the process of completing this thesis.

I am obliged to many of my college colleagues for taking participation and providing their valuable time for the experiment. Finally, I take this opportunity to express my gratitude to my parents and friends for their unconditional and unfailing love.

Abstract

Screen-reader users access images on the Web using alternative text delivered via synthetic speech. However, research shows that this is tedious and unsatisfying for blind users, because text-to-speech applications lack expressiveness. This paper, poses an alternative approach using an experiment that compares audemes, a type of non-speech sounds, with alternative text delivered using synthetic speech. The findings show that audemes perform better across many areas. Specifically, audemes required lower mental and temporal demands and led to less effort and frustration and better task performance. Moreover, participants recognized audemes with higher accuracy and lower errors. Audemes were also perceived as more engaging compared to alternative text delivered using synthetic speech. Additionally, audemes were found to be richer in delivering information. This study suggests that non-speech sounds could substitute or complement alternative text when describing images on the Web.

keywords

Screen-reader; non-speech sounds; alternative text; Web image; accessibility.

Table of Contents

List of Tables

List of Figures

1. Introduction

Image has been ubiquitously exercised to deliver essential information in web. Now, Search engines are starting to appear allowing users to find multimedia content with the quantity growing even faster than text on the Web. Consequently, the wealth of textual information associated with multimedia on the web, especially images that can further be used for web indexing and searching is placing image in compensatory role to other source of information. But, perceiving image content and context has been a constant problem for visual impaired person who are primarily relying on the assistive technology to access web information. Content includes identity and properties of objects/events in the image while context carries surplus sensational information about the content such non-verbal, non- linguistic contents, emotions, environment knowledge and many mores. The most of available feasible assistive technologies are text to speech translators (TTS) such as Screen Reader(Bleicher & Bleicher, 1980), which basically translate textual information into synthetic speech.

WCAG (Chisholm & Jacobs, 1999; W. W. W. Consortium, 2008) suggest compulsory inclusion of textual description to represent pictorial content in web, known as Alternative Text. Though WCAG itself found insufficient to guarantee websites accessibility in user study with disable people(Rømen & Svanæs, 2012), alternative texts are known for an easy and proven method to make image accessible to the blind people(Evett & Brown, 2005; Slatin, 2001). An assistive technology basically reads textual description of image to the user. Alternative text should need to be meaningful and easily perceivable to propagate exact meaning as well as eradicate confusion among the user(Esposa Jr, 2008). But, choosing suitable and equivalent alt text for image which must convey same information as the image itself remains a challenge. In this context, WCAG 2.0 suggest alt text should be as short as possible to provide a substitute for an image and states “If the Alt text is greater than 100 characters (English) then it must be shortened or the user must confirm that it is the shortest Alt text possible”. Whereas other independent web standard and practice recommends different length restrictions on alternate text. This eventually undermines the quality of alternate text as well limit the amount of information conveyed to readers, and can affect the style of writing in a way that inhibits comprehension as well. Similarly, accessibility checkers also raise warnings if alternate text is too short or too long than their specifications. Moreover, tools that check and verify web sites' accessibility level currently ignore the quality of alternative text due to unavailability of solid standard and automatic method for judgment. Beside these, content of image might be easy to state and articulate in word when it comes to embed contextual information text suffers from its lack of expressiveness. Beside these, speech need to be expressive to support natural phenomenon of communication and deliver meaningful sensation about textual content. But, conventional Text-to-Speech (TTS) systems uses neutral synthetic speech which again lacks natural properties of human voice and fails to address human expression and emotion content summoned in written text. Thus, the lack of appropriate and meaningful way to represent web images remains an obstacle for blind users and others approaching the web with non-visual assistive medium.

In this scenario, non-speech sound could be an reasonable solution to represent image in web. It could deliver both content and contextual information. Auditory signal is already been in use computer to compensate general data that falls behind the focus point of our visual system, for example alerting user to things(Deatherage, 1972; Sanders & McCormick, 1987). Similar integration in multimodal computer systems(Blattner & Dannenberg, 1992) which help leveraging human-computer interaction while directing our visual attention to one task, such as surfing internet or editing a document, we could still monitor the state of other activities or tasks on our system. Extending these ideas encoded messages (Buxton, 1989) could be used to pass more complex information in sound such as inform peoples about their medication or about an upcoming events(M. McGee-Lennon, Wolters, McLachlan, Brewster, & Hall, 2011; M. R. McGee-Lennon, Wolters, & McBryan, 2007).

Speech as a complex integrated audio signal carries huge amount of details about message, speaker, speakers' intentions and reactions, emotion and others. The existing TTS systems process synthetic and natural speech effectively but their performance is not significant in compared to emotional speech. Similarly, It is difficult to model and incorporate natural human emotional phenomenon in conventional TTS. In real life conversation, non-verbal expression and emotion carries important aspect communication which play key role at interpretation and perception of message. Therefore, understanding non-linguistic emotional detail in a message conveyed through text and TTS translation of textual content is very difficult. The same textual content could be represented and conveyed more sensually and meaningfully by preserving and incorporating those emotional contents.

Non-speech sound at encapsulating emotion and expression ,Auditory emoticons (Froehlich & Hammer, 2004)which are an auditory briefing of smileys such as laughter, chuckling, or crying were found well perceived in user-centered study with an email reader application. Tailored auditory emotions in email reading resulted into reduced perceived workload without increasing user annoyance. In similar study with sighted and blind people (Wersenyi, 2009; Wersényi, 2010a) at representing emotional content of communication ,auditory emoticons were found significantly perceived and able to convey a surplus source of amusement. In compared to auditory icons(Gaver, 1986), they use human nonverbal voice with emotional load instead of a broader scope of environmental sounds. In recent, Spemoticons (Németh, Olaszy, & Csapó, 2011)an acoustic events synthesized based on meaningless vocalized expressions that do not occur in real life obtained by modifying the intensity, pitch, and temporal structure (by inserting breaks of various lengths) of TTS synthesized phrases provides real-time generation of audio which maps emotions to the sounds. But, interpretation of spemoticons could be culturally dependent.

Similarly, acute temporal resolution is another strengths of the auditory system(P. D. Kramer, 1994). In some cases, even reactions to audio stimuli have been found to be faster than reactions to visual stimuli (Bly, 1982). For example, emergence alert signals in detection of anticipated stimuli and danger(Papastavrou & Lehto, 1996), auditory enhancements for visual signals in a clinical environment (Chan & Ng, 2009) which minimizes the reaction time for safety measures. He found that perception of sound is different to visual perception where sound can offer a different intuitive view of the information it presents. This could allow us to understand and memorize information in different ways than that might be impossible by other means. Additionally, it is found that blind people have high level of acoustic ability (W Niemeyer & I

Starlinger, 1981) than sighted people(Sánchez, Lumbreras, & Cernuzzi, 2001). They could process auditory language stimuli faster than sighted people(Röder, Rösler, & Neville, 2000). For example, in a study with blind and visually impaired students shows that their accuracy at identifying the location of sound source is superior to that of sight student(Doucet et al., 2005).

Moreover, presenting information in synthetic speech is very slow as it suffers from same drawback as text in text-based computer system which is a serial medium. Text may takes many words even to describes simple information due to lack of its expressive capability(Barker & Manji, 1989). User needs to hear synthetic speech from start to end and many words might have to be comprehended before a message can be understood. Messages can be composed in shorter form and therefore heard more rapidly and sensationally with non-speech sound. User could hear an encoded non-speech message which could be able to convey content and contextual information more precisely and shortly. User will get true sensational source of information where they have opportunity to use their own comprehension and recalls meaning rather than directly listening pre-comprehending meaning described in words. Moreover, an experiment(Smither, 1993) to investigate memory load of natural sound against synthetic speech on young and old adults showed that synthetic speech puts a heavier load on short term memory than natural sound. Similarly, it is found that recognition accuracy of speech (Slowiaczek & Nusbaum, 1985) decrease significantly with the increased presentation rate. It requires much of practice even for highly skilled blind user to reach at higher recognition rates. But, none of these issues were found in non-speech sounds such as Auditory Icons(Gaver, 1986), Earcons (Blattner, Sumikawa, & Greenberg, 1989) and Audemes (M. A. Ferati, 2012a). Additionally, it can be visualized to help deaf people(Matthews, Fong, & Mankoff, 2005) to maintain an awareness of sound present in environment .

Auditory icons are a representation of natural sound that people listen to the world in their everyday lives, could be used to convey information. It is like sound effect produce or trigger by an item; for instance, variety of impact, bouncing, scraping and machine sounds. We could parameterize these categories of events and objects in order to reflect their proper dimensions as well. It provides a natural way to represent dimensional data as well as conceptual object in a computer system. For example, sound would be large according to file size, we could hear new surface if file is dragged over new surface, sounds becomes quicker if process starts executing more quickly and others. This parameterization of icon can serve rich source of information about object size, age, computational operation and their speed as well. Auditory icons are suitable for notifying user interaction, alerting running process and models, helpful for navigation and collaboration(Gaver, 1997). However, its true application is solely based on direct representation of associated concept. It is very difficult to accurately classify or even impossible to create absolute auditory icon for a word or concept.

Whereas, earcons are non-verbal audio messages that are used in HCI to provide information to user about computer object, operation or interaction. It uses abstract synthetic tones to create auditory message. This non-verbal music has useful application in user interaction and operation; they serve a non-visual representation of objects, functions and events in computer. There are two types of earcons; Compound and Hierarchical. Compound earcons are simply produced by integrating basic ones that severs an easy and effective method of earcons generation. But, building hierarchical ones required depth knowledge of earcons and inheritance properties of different parent component in the hierarchical tree. Despite, earcons uses a more musical approach than auditory icons but they do not need same natural association as auditory icons do. So, they are easier to use and create. But, it might be frustrating and difficult for users to learn and remember arbitrary earcon-based menu. There is no intuitive link between the sound and what it represents. The link must be learned by the listener. Moreover, earcons contains complex and highly specific musical grammar that needs to learn before using them in the system. Thus, remembering and decoding musical pattern may also be quite difficult task for non-musician. The same problem was found in Mathtalk system (Edwards, 1997)which uses musical earcons to indicate structural delimiters and provide an abstract overview of entire equation. They found that cognitive effort required to decode each pattern detract user from processing of the mathematical content.

However, audemes are a new category of non-speech sounds. They can convey theme based content(M. Ferati, Pfaff, Mannheimer, & Bolchini, 2012). Audemes are semantically more flexible than other non-speech sound. It could generate meaning for an object/event/context depending on the identity or meaning of the sound concatenated. Though meanings are not completely open and arbitrary, it starts wide and then narrow downs with additional sound cues which eventually merged into a single meaning. In terms of semiotics analysis, audemes are more likely to auditory icons. It is easy to interpret meaning. When an audemes is herd, object is clearly recognizable to the users. For example, when an audemes "keys jangling" is played, the object keys easily identifiable to users. Once object is known then it is easier to map meaning of the sounds. However, in contrary to auditory icons, meanings of audemes frequently changes with the addition of extra audio elements to audemes which extends its semantic grammar. For, example concatenation of "key jangling" sound with "car engine sparking audio" can represent "car being started for rides". This makes audemes a multi-vocal (ability to generate different meaning) non-speech sound in compared to others uni-vocal (having single meanings).

Additionally, design of audemes is entirely depends on empirical knowledge of designer which often results into creation of sounds derived from random selection or the personal preference of the sound designer. It could be sound from natural occurrences as well as abstract and musical tone or combination of them. Study found that audemes are significant at improving and increasing non-visual message encoding and recognition (Steven Mannheimer, Ferati, Huckleberry, & Palakal, 2009) in study with blind users. Audemes used as navigation and thematic landmarks in touch-screen were found much more variable and flexible, and still easy to learn, memorable and navigable to the visually impaired teenagers(M. Ferati, Mannheimer, & Bolchini, 2009). Similarly, audemes were found much helpful in reducing memory erosion as even after five months, content was remembered better along with audemes than without it(M. Ferati, Mannheimer, & Bolchini, 2011). They were also found potential in scientific application including gaming and productivity as well as an educational tool for better memory retention(Meyer et al., 2014) as well.

Moreover, meaning of non-speech sound is generally identified through reference to the cause of sound. But, meaning of audemes first starts with identifying cause of the sound, and then following reference. For example, “screaming voice of a horse” will be first linked to identify the type and nature animal. Then, this link is referenced to find the meaning of an audemes and that could be “horse riding”, “horse polo”, “horse racing”, “galloping horse in wilderness” or many others instance or event related to the subject. Therefore, this ability of generating meaning through the cause of sound and then following referential modes makes it more suitable for encoding content and contextual information than others.

1.2 Motivation of Research

There are many reasons why non-speech sound could offer advantages over Alt text to represent images on web:

1. Subtlety and non-intrusive enhancement quality

   Short nonverbal sound has been a successful common enrich interaction method for people with physical disability such as steering a wheelchair (Fehr, Langbein, & Skaar, 2000), User Interface (Poláček, Sporka, & Slavik, 2012) and the Vocal Joystick (Harada, Landay, Malkin, Li, & Bilmes, 2006, 2008).Gestures and speech recognition system have drawbacks of time delay for recognition algorithm and probability of being miss recognized. In that case, non-speech could be alternative approach which can delivery same information very shortly and precisely(Cowling & Sitte, 2002).They used a low latency input to precisely trigger events by shortly blowing into a microphone and found method allows completing task and performs better than expected against standard interaction device.

   Moreover, its non-intrusive quality helps enhancing the perceiving quality of systems because it allows increased refinement of information (G. Kramer, 1994). For example, homeland securities which are mostly depend on visual clues to detect an abnormal event but it does not cover all things. Non-speech audio can provide information in many cases where video systems fail to detect occurrences, such as screaming, glass breaking, knocking on a door, talking, footsteps sounds in nights. Thus, the addition of audio increases the quality of information for more security reasons. Similarly, its ability to represent usual temperature, pressure, size, cost, and rate makes more suitable for integration to main stream technology. We could also map and visualize more subjective and affective variable such as value, goodness, beauty, risk and many mores(Walker & Kramer, 1996, 2005) onto sound dimensions.

2. The audios are very powerful and have great potentials in information presentation(Bly, 1982).

   Music has complex structures and sub structures which makes its a potential medium to transmit complex information to the user(Alty, 1995). Thus, message can be codded and delivered in a different way that are more memorable and informationally rich. As Alty stated:

   “Music is all-pervasive in life and form a larger part of people’s daily lives. It is very memorable and durable. Most people are reasonably familiar with the language of music in their own culture. Once learned, tunes are difficult to forget.”

3. Pleasant and more natural

   Serial nature of speech make Information presentation and delivery slower than non-speech sound. User need to hear speech from beginning to end and comprehend many words in order to assimilate and understand full message. With non-speech sounds the messages could be more natural and melody therefore could be delivered and heard more pleasantly and rapidly.

4. Easy to habituate

   Due to its non-intrusive enhancement quality, non-speech sounds are easier to habituate (Fuller, Adams, & Buxton, 1989).For example an air conditioner sound. We only notice the sound produce during device turn off and on, operational sound becomes unnoticeable. Thus, non-sounds could be designed to facilitate habituation if required which is nearly impossible to achieve with speech sound. This helps to reduce and overcome annoyance and stress factor associated with non-speech sound application.

5. Unattended Speech Effect

   Background speech, even at low intensities, is found to be much more disruptive than non-speech sound when recalling information(Banbury & Berry, 1998). Whereas, non-speech sounds are much more helpful in meditations.

6. Non-speech sounds are more suitable for presenting continuous information

   Computer-generated sound patterns of two-dimensional line graphs (Mansur, Blattner, & Joy, 1985) were used to deliver line graph information in holistic manner to the blinds ,and were found very effective. Similarly, mapping of data into auditory signal is found to be successful in monitoring task, such as informative auditory displays used in supervising patient's wellbeing(Sanderson, Liu, & Jenkins, 2009), detecting trend in data streams (Walker & Kramer, 2004)and sonification-auditory icons hybrid techniques in a satellite-ground control environment(Albers, 1994). It is also found to be useful in other various scientific areas, such as seismology (Dombois, 2001), stock trading (Janata & Childs, 2004)and genetics (Won & Hey, 2005). More recent works include sonification(Hermann, Hunt, & Neuhoff, 2011; Kramer, 1993)of large amounts of geo-referenced data (S. H. Park, Kim, Lee, & Yeo, 2010) and emotional hand gesture data by accelerometer(Fabiani, Dubus, & Bresin, 2010). Likewise, application of sonification to touch screen in the process of map exploration (Delogu et al., 2010)and image perception (Yoshida, Kitani, Koike, Belongie, & Schlei, 2011) for blinds.

1.3 Research Aims

The overall aim of the research is to investigate and elaborate application of non-speech sound to represent image in the web. Study explores semantic ability of non-speech sound and their context of use and ease at implementation issues. Consequently, audemes were tested to encode and increase accessibility of web image to the screen reader users in a comparative study with Alt text. Audemes were expected to be effective and efficient at presenting web image to the screen readers' user and those who approaching web with non-visual mediums. Overall aims of the study are:

1. To investigate whether audemes are feasible to represent web image.
2. To evaluate effectiveness and usability of audemes to represent web image.

2. **Problem statement**

For most people, accessing or surfing the web is an activity with informational benefits but, this might be a completely different story for disable one. Disable people has special needs which need to address in web design and policy. It is essential that the Web design should be universally accessible in order to provide equal access and equal opportunity to people with diverse abilities and difficulties. One way to ensure web accessibility rights of disable people is by exercising human or civil legislation. For examples, Australian blind people court won case against Sydney Organizing Committee of the Olympic under Disability Discrimination Act 1992 in 2000, Brazil published guidelines for accessibility on 18 January 2005 for public reviewing, European Parliament stated that all websites managed by public sector have to be accessible to everyone in February 2014, and many more. For example, Norway has a legal obligation under the Act June 20, 2008 No 42 relating to a prohibition against discrimination based on disability, known as the Anti-Discrimination Accessibility Act. The Act went into force in 2009, and the Ministry of Government Administration, Reform and Church Affairs published the Regulations for universal design of information and communication technology (ICT) solutions in 2013. The regulations require compliance with Web Content Accessibility Guidelines 2.0 (WCAG 2.0) / NS / ISO / IEC 40500: 2012, level A and AA with some exceptions.

However, W3C web accessibility standards exist over a decade, implementation of accessible websites, software, and web technologies are still dragging. Lack of appropriate knowledge and expertise among developers, fragmentation of web accessibility approaches and legal binding are some key reasons. Alt text in webpage describes pictorial content in text. Consequently, TTS read out those textual equivalent of visual content , and becomes accessible to blinds and visually impaired users. In contrary to image scope and usage to cover and illustrate wide range of information and context in the web, makes it extremely difficult to explain within few words. Similarly, long text description also makes web sluggish and inefficient thus increases cognitive loads on users. Moreover, characteristic of unarticulated synthetic speech remains another additive complexity to Alt text.

2.1 Hypothesis

Access to web image is improved when enriched with content enriched non-speech sounds, audemes to the screen reader user.

H.1. The overall subjective workload is reduced under the audemes enrichment of images compared to alternative text.

H.2. There is a decrease in the frustration level while accessing web images with audemes compared to alternative text.

H.3. There is an increase in usability of webpage when images are enriched with audemes compared to alternative text.

2.2 Research Questions

RQ1: How do alternative text and audemes compare in terms of perceived workload when comprehending images on the Web?

RQ2: Will users recognize images more accurately under audemes enrichment compared to alt text?

RQ3: Does accessing images under audemes enrichment delivers better website usability compared to alt text

3. Literature review

3.1 Overview of non-speech sounds

In early days, non-speech sounds in computer are used to inform programmers (Krokstad, Svean, & Ramstad, 1994)about computer's bus. Depending on the patterns and rhythms of the sound trigger by program counter they could identify types of alert, and fixed those bugs. In general, user has clue about computer operation such as copy and paste, saving and installation of programs and many more depending on noise produced during hard disk operations. Consequently, It is suggested that non-speech audio could aid in comprehension and analysis of programs behavior (DiGiano & Baecker, 1992) in sound enhancing programming environment. Later on, Brewster et al.(Stephen A Brewster, Wright, & Edwards, 1995a) suggest and described use of musical parameters such as timbre, register pitch and duration for composing suitable earcons to deliver information through sound. Earcons to aid people in scanning-input method(S. A. Brewster, V.-P. Raty, & A. Kortekangas, 1996) which is found naturally perceived rhythms and favorable to disable participants. Similar use of musical notes in 3D spatial auditory information ,described as TouchMelody (Ramloll & Brewster, 2002), found to be increased usefulness, information content and reduce tactile clutter in tactile diagrams to the blind users. They found system is easy to understand and explore by hands with dynamic non-speech sounds. Consequently, new categories of non-speech sound based on music and environment cue were emerged which further broaden research and application scopes such as feature extraction(Bolea, Grau, & Sanfeliu, 2003) ,sensory substitution for tactile perception in robotic human interactions(Csapo & Baranyi, 2010), therapeutic assistance (Ferrari, Robins, & Dautenhahn, 2009), Gammatone filters in biology(Valero & Alias, 2012) and so on. In this section, paper briefs existing non-speech auditory representations in HCI and their specification.

3.1.1 Earcons

Earcons (Blattner et al., 1989) were introduced as non-speech audio messages(Blattner & Greenberg, 1992) in attempt to overcome cognitive load imposed by heavy use of graphical icons in computer(Ramsey & Atwood, 1980). They are useful in auditory enhancement of computer user interface. They can represent computer objects, operations and interactions such as graphical icons, menus, files and their compiling and executing operations such as editing, saving, deleting ectaras. They become auditory representation of graphical counterparts that are used in HCI. They help to overcome computer user interface design and operation in universal design and accessibility perspectives. Earcon (Stephen A Brewster, Wright, & Edwards, 1993) were first tested as brief musical abstract sound made up by altering timbre, register, and tempo systematically. The meaning of sound is arbitrary and has not any association with objects signature or environmental sounds. They have abstract relationship between sounds and its meanings that user need to learn based on environmental experience. Application of earcons demands a learning curve to user that they need to confront how earcons are linked with objects and events. Similarly, earcons were tested for structured integration in HCI (Stephen Anthony Brewster, 1994)by sonically enhancing scrollbars, buttons and windows which resulted into improving usability, reducing workload and overcoming recovery from error (Stephen A Brewster, Wright, Dix, & Edwards, 1995)without increasing annoyance. Integration of structured earcons yield effective interface widgets. Parallel earcons(Stephen A Brewster, Wright, & Edwards, 1995b) were also tested in search for usable design and increase sounds presentation rates. Presenting earcons in parallel to reduce time duration and increase interactive pace with HCI shows no difference in recall and recognition of parallel earcons in compared to serial compound earcons. In fact, parallel earcons were found effective at increasing audio message presentation rate without impacting recognition rate. Further, earcons were studied for presenting navigational information in menu hierarchy (Stephen A Brewster, 1998),and are found a robust and extensible method of representing hierarchical message with increasing recall and memory advantages. Other studies include use of earcons to overcome small screen space in handheld devices(S. Brewster, Leplâtre, & Crease, 1998) and maximizing screen space on mobile computing(Stephen A Brewster & Cryer, 1999).

3.1.2 Auditory Icons

Auditory icons were introduced and explored in a new way of understanding everyday sounds, the way of recognizing events that happen every-day in environment by listening to their sounds or sounds they make. The basic idea was integration of those every-day sounds in computer interface, which may be applicable to present information to computer users in analogous to visual icons. Auditory icons have semantic connections to object or events they represent. Their natural and metaphoric association to things they represent makes they easy to learn and interpret. User could visualize and map object associated with auditory icons even with their experience gained from every-day life such as simulated sound of recycle bin and printing icon. Auditory icons were first investigated in terms of audible source attributes(Gaver, 1993), and introduced(Gaver, 1986) to represent dimensional data and conceptual objects in computer interface. Further, auditory icons were tested in SonicFinder(Gaver, 1989), an auditory interface developed for Apple Computer, Inc. Interface was enhanced with auditory icons and standard graphical feedback. The SonicFinder has same feature as of original interface but sampled with recording of everyday sound-producing events to simulated its operation such as move, copy, delete files. This extended visual desktop icons into auditory representation. Result shows that using auditory icons in interface increased users level of direct engagement and improved flexibility and usability. This helped in extending mapping of auditory icons beyond their implied literal meaning. However, usability of auditory icons associated with environment sound was found affected by requirement of higher cognitive and perceptual resource in recognition(Leech, Gygi, Aydelott, & Dick, 2009) and resistive or vulnerable to distortion of meaning in filtering(Ballas, 1993). Others studies shows that logical and expected auditory were more recognizable than unexpected (Leech et al., 2009)and approximated heightened version of auditory were superior to realistic one(Fernström & Brazil, 2004).

Application of auditory icons in warnings were first studied for collision avoidance (Graham, Hirst, & Carter, 1995)over visual and haptic to reduce reaction times and eye free use. Paper argues in favor of auditory icons applicable in conveying everyday events in life general. Later, learnability and retention of auditory icons in auditory warnings (Leung, Smith, Parker, & Martin, 1997)was discussed in comparative study between auditory icons, speech and abstract sound for better memory invocation and learnability in emergency warning with 18 participants in 9 session over 3 weeks to represent a set of eight different warnings occurs in different events of military aircraft cockpits, auditory icons were found easy to learn and remember like speech than abstract sounds. Participants learnt and remembered auditory icons more easily than those of far difficult abstract sounds for events warning exercise. Research suggested that close natural association between auditory icons and event it represents makes is more memorable and recognizable than other types of abstract sounds, and those of having significance like speech indicates a potential use in informationally enriched auditory warnings. Other study includes designing auditory sign for In-vehicle information systems(IVIS) to support traffic awareness and reduce operational cognitive load(Fagerlönn & Alm, 2010) found that arbitrary sounds used for auditory cue have difficult learning curves, degraded user response performance to warning, negative impact in driving satisfaction than those of sounds having natural association in the driving context. Finding indicates auditory icons could be a reasonable auditory solution in developing wide range of on-broad intelligent transport system for assisting drivers in various traffic situations .Auditory signal as a warning in truck(Fagerlönn, 2011) were found highly acceptable but depended on subjective selections of auditory icons for warnings and different driving situations.

Application in increasing accessibility includes auditory icons enhanced GUIB (Textual And Graphical User Interface for Blind users) for MS and X window applications(Mynatt & Weber, 1994), educational software application(Jacko, 1996),and many mores. For example, Auditory icons combined with spatial sounds to create spatial auditory icons for navigating distance and direction in BATS (Blind Audio Tactile Mapping System) to help blind children exploring spatial information(Parente & Bishop, 2003) in computer such as maps. System design was found insightful for further design and improvement to increase accessibility to spatial information. Auditory icons used in representing obstacle information such as traffic poles, garbage bins, tree in sidewalk and many mores in AudioGuide (an outdoor navigation aid system) (Xu, Fang, Dong, & Zhou, 2010)for visually impaired along with arbitrary abstract non-speech sounds to represent dynamic status and progress of auditory events such as direction discrepancy and the distance to the destination were found to improve safety and reliability of aid system to visually impaired users.

3.1.3 Spearcons

Spearcons are different sound briefing than earcons and auditory icons with a more effective acoustic association. Spearcons are produced by automatic translation of textual content through TTS and then speeding up the speech without changing pitch to the point that is no longer recognizable as speech. They are better organizable in terms acoustic association and features. For example, spearcons for menu such as save, save as, save as with file extension which all of them are unique with different length but acoustically similar at the very beginning of audio. User can hear and feel different length of menu navigation as scanning through list. Though spearcons are not comprehensible by words, they are non-arbitrary which demands learning curve but less than earcons(Dingler, Lindsay, & Walker, 2008) and auditory icons(Palladino & Walker, 2007). It supports dynamic algorithmic creation as well as more flexible menu structures. However, they are not effective at navigating menu hierarchy as eracon they provide more direct mappings. Spearcons(Walker, Nance, & Lindsay, 2006b) are first introduced as a speed-up sound in improving navigation performance in menu navigation(Walker & Lindsay, 2006). Consequently, spearcons statistically found well effective for auditory navigation of graphical menu and significantly increased performance as well. Other studies show spearcons alone results superior in useful user interface in compared to hybrid auditory-earcon(Dingler et al., 2008). The outcomes highlighted important aspect of spearcons in auditory interface design. Further, in experiment visually impaired participants (Wersenyi, 2008)it was found that appropriated adjustment in compression ratio and other parameters such as pitch and speed helps to improve accessibility of spearcons. Recent study in compression and usability of spearcons to increase application shows that though spearcons resembles same as spoke work in acoustic sense it starts lose identification drastically below compression of 40% percentage of original TTS audio(Sun & Jeon, 2015). The results suggest that in overall in-combination with accuracy, user reaction time and usability, spearcons supports bottom limits of 40 % compression of original sounds. Moreover, in experiment of modality and encoding strategy effects on a verification task with accelerated speech (spearcons), visual text and tones with 51 users (Nees &

Best, 2013)where participants were instructed to memorize the prescribed encoded stimuli either as words (verbal word) or tone (auditory cue) in verification task. Spearcons were found to have same stimuli as verbal working memory code in visual text but in compared to tone result shows more specific-stimuli dependence effect in verification of task. This indicates that spearcons (auditory cue) seems to have same auditory stimuli perception effects as text in memory invocation verification task.

Application includes spearcons to represent distance and forward direction in pedestrian navigation in compared to earcons and short pulses in eye-free environment(Jeon et al., 2015) were found efficient in conveying measurements data accurately and helped participants to complete task. Participants found favored spearcons in compared to others more satisfactory auditory guidance in pedestrian navigation. Similar, comparative study in English based spearcons and Chinese based spearcons in representing distance and forward direction in eye free pedestrian navigation engravement with 10 native Chinese participants (Hussain et al., 2016)shows subjective language dependent preference and satisfaction . Test with TTS in compared to both English based spearcons and Chinese based spearcons, they preferred more satisfaction rating to Chinese based auditory feedback in navigation overall. Spearcons with speech in hybrid auditory feedback in mobility assistance for the visually impaired(Hussain, Chen, Mirza, Chen, & Hassan, 2015) was found more effective than non-speech and speech alone. Mostly used speech were replaced with spearcons and integrated along with speech in subjective design. In in-door study with Visually impaired participants shows greater improvement in auditory feedback. Participants were found exciting about the feedback design and rated model as more efficient and less annoying than speech only. This indicates hybrid integration of spearcons yield easy, fast and more meaningful design of feedback navigation system. Spearcons to reduce visual distraction from in-vehicle human–machine Interfaces(Larsson & Niemand, 2015) while using on board navigation and media player system (the main reasons of driving attention distraction) were found increased participants driving performance by reducing off-road visual attentions and distractions. In highway driving simulation test with 14 participants, spearcons enhanced in-vehicle user interface were found efficient at reducing off-road eye glance demand in using on-board music and navigation system while driving in dual task scenario in compared to earcon enhanced auditory display and without any auditory cue at all.

3.1.4 Spindex

Spindex are speech index enhancement of spearcons(Jeon & Walker, 2009). They are generally produced by accelerating initial letter of menu items so user could get direct sense of navigation. It offers great advantages when user has to navigate longer menu items and items are alphabetically ordered. Navigating longer mobile contact list with TTS alone and TTS plus spindex over visual menu displayed or not, it is found that spindex are preferred by the participants, and they reduced search time and improved navigation in long lists(Jeon & Walker, 2009). Similarly, spindex auditory enhancement in music player where user has to navigate long songs list as a secondary task in dual task situations(for example while driving or conversation with others), spindex were preferred in less subjective workload and higher usability(Jeon, Davison, Nees, Wilson, & Walker, 2009). Later, three design alternative of spindex tested with sighted and blind users (Jeon & Walker, 2011). They are “Attenuated spindex” where first occurrence of each spindex is louder than all the rest; “decreased spindex” where succession of the same spindex becomes gradually softer as the user browse the list; and “minimal spindex” where only the first occurrence of a spindex can be heard. The result were not clearly conclusive in terms of preference between participants groups but apparently “ attenuated” spindex were found mostly preferred.

3.1.5 Spemoticons

There are basically two types of non-speech auditory cues which are used in HCI, one is use of real life environmental sounds such as auditory icons and other is artificially generated synthetic sounds such as earcons. Others are speed up audio of TTS synthetic output known as spearcons, and spindex as auditory index. Auditory emoticons as non-verbal by product of human emotional expression. In contrary, spemoticons are auditory emotional state in compared to emoticon characters generated by TTS. Spemoticons are unrecognizable and unintelligible emotional by product of TTS like spearcons in linguistic sense. They were first introduced as acoustically synthesized unintelligible vocal expression(Németh et al., 2011).They are generally produced by altering temporal structures, intensity and pitch of TTS synthesized phrases. However, mapping of emotion into spemoticons might be culturally dependent. The generation of spemoticons involves generation of basic audio through TTS first then control alteration of acoustic and other parameters of audio until desired emotional product result. This facilitates real time generation of spemoticons which could directly maps human emotion into auditory representation

3.1.6 Auditory Emoticons

Auditory emoticons are vocal based, non-speech and auditory by product of human emotional expression representation of smileys such as winking, smiling, chucking, crying, kiss ectaras. Auditory emoticons were first studied with email reader (Froehlich & Hammer, 2004)to express non- verbal elements occurred in written text. Study showed that tailored auditory emoticons in email reader efficiently support valid human emotional expression in user interface. Integration of appropriate auditory emoticons to convey non-linguistic expression resulted into increased performance and user satisfaction. Tailored auditory emoticons showed that they improved user perception rate regarding understanding structure and meanings of audio content in email by reducing subjective workload. Sound helped in organizing structural textual audio into categories of highlighted and normal text while secondary sounds guided user attentions and reduced perceptual efforts. Overall outcomes of the study indicated that user preferred more expressive email (enabled with auditory emoticons) over text only. Diversity of user choice over music and gender in sound expression was found important in perception of auditory icons. In extensive evaluation of auditory emoticons with auditory icons, earcons and spearcons for a better HCI with blind and sighted user(Wersényi, 2010c), auditory icons using environmental, familiar sounds used to represent additional emotional content were found well perceived.

3.1.7 Anthropomorphic Auditory Icons

Anthropomorphic auditory icons (Schmitz, Fehringer, & Akbal, 2015)are synthetically produced affect bursts(Scherer, 1994).They are short, discrete, nonverbal expressions used for representing non-linguistic, nonverbal emotional state and characteristics of anthropomorphic objects such as animated characters. It is produced through the process of filtering all human acoustic attributes from recorded bursts, Montreal Affective Voices (MAV)(Scherer, 1994) by keeping sufficient acoustical information to propagate intended affection and emotion with same efficiency as original affect bursts. Consequently, these intermediate sounds further combined with carrier sounds associated with characters to encapsulate different affectionate and emotional state and characters such as admiration, happy, suffering, worry, sad, pleasure, relief and others like coughs, groans, snoring and many mores. Though in an online user study with 20 users to explore recognition of emotions which are selected and categorized from Ekman Faces(Ekman, 1993) in andromorphic auditory icons were found inconclusive in contextual use, it is concluded that sounds perceived differently in real world so concrete domain application could provide more contextual use in perceiving acoustic feedback more accurately and precisely.

3.1.8 Auditory Scrollbars

Concept of auditory scrollbars emerged from the use of non-speech sound to elaborate and discover hidden information in user interface depending on events, status and modes(Brewster, Wright, & Edwards, 1994). Two types of sound were used for representing window scroll or thumb movement, and page scroll and position to represent events, status and mode of visual scrollbar in user interactions. In experiment with enhanced auditory scrollbar it is found that timing and error along with subjective perception of workload was significantly reduced. And, overall participants preferred enhanced scrollbar over visual scrollbar. The study shown that auditory enhance of graphical widgets improves the usability of interface. Consequently, auditory enhancement of user interface in small hand held device to overcome screen space and resolution along with noise and speech clutter produced by voice enabled application, study on buttons, scrollbars, menus and tool palettes show increased user preference, improved usability and reduced subjective workload(S. Brewster, 1998; S. A. Brewster, 1998). Study also found that time demand for completing task and time required to recover from the error was improved significantly.

Moreover, in an attempt of improving navigation of auditory menu for blinds and impaired people and develop design guidelines, auditory scrollbar was test with five different auditory enhancement which are pitch polarity, single-tone, double-tone, alphabetical grouping, and proportional grouping(Yalla & Walker, 2008). In user testing with sighted and blind participants, among others proportional grouping scrollbar was found best at increasing performance and preferred by the user. Study shows proportional grouping as an effective and acceptable auditory menu enhancement for auditory scrollbar design. The study demonstrated application of auditory enhancement of scrollbar in long list menus such as song list

3.1.9 Musicons

Musicons are snippets of small piece of familiar music which are produced by sampling well-known music or song piece(M. McGee-Lennon et al., 2011). The small sampling piece of famous music are found highly memorable and robust auditory reminders. Since people are personally associates with memories of music and songs, musicons are easy to remember and recall. They are specially suggested as an auditory remainder in home and work circumstances such as options for alarms and notifications. In terms design and characteristics, they have less meaning association than auditory icons and build on more familiarity than earcons. So, their learning curve might fall in-between auditory icons and earcons which remains to investigate.

User performance was found increased with musicons in compared to short spoken reminders. Participants attended a high level of accuracy (89%) which further increased (90% or better across sessions) among participants who understand meaningful links between music and the associated tasks(M. McGee-Lennon et al., 2011). Recent study (McLachlan, McGee-Lennon, & Brewster, 2012)suggested that musicons of length 5 seconds , produced by user own preference and selections have higher recognition and usability rates. Musicons were also found useful in music classification and search operation in design and development of large scale music repositories(Roy, Pappu, & Prabhakar, 2004). Musicons guided user to search music by certain characteristics and helped to arrive at subjective choice of music.

3.1.10 Morphocons

Morphocons (Parseihian & Katz, 2012)are short morphologically customized audios that have the same basic role as the earcons, also known as morphological earcons. They support hierarchical sounds grammar and recognizable without knowledge of sound context and musical meanings. Thus, highly learnable than earcons. Morphocons do not have naturally perceptible meaning like auditory icons which provides intuitive link between sounds and the associated objects/events. Morphocons are similar to the Temporal Semiotic Units(Frémiot, 1999; Schaeffer et al., 1967) which are morphological description of acoustical parameters such as envelope, rhythm, frequency, and sound length. They are generally produced by morphological sound grammar. Morphocons can be generated by mapping envelope, harmonic and rhythmic properties of the sound onto natural, musical and synthesized sounds which kept general properties of sound constant and specifics changes according to subjective preference of user. Morphocons were first tested in an audio-based navigation assistance system (called "NAVIG")for visually impaired which help them recognize obstacles and path(Katz et al., 2012).

3.1.11 Lyricon

Lyricon are audio icon based on musical features and lyrical keywords of songs(Machida & Itoh, 2011). Thus, lyricon contains both the musical pattern as well as lyrics of story. It enables selection of musics according to users' preference based on their impression of musical sounds and lyrics. It found effective at supporting user interface and adaptive display in musical player.

3.1.12 Lyricons

Speech transformation of non-speech sound have number of significant application in electronic devices along with completely instrumental and natural non-speech sounds. Lyricons(Jeon, 2013) were developed in a consideration to revamp the significance of auditory cue of speech with non-speech by combing them into a new category of non-speech sound. Lyricons generally combines the direct meaning mapping accuracy of speech with musical aesthetics of earcon to generate high level of direct sound mapping than abstract and uncertain acoustic relevant earcons. Thus, resulted lyricons is expected to have more accessibility and usability features for inclusive user groups such as visually impaired and kids. In study with novel delivery approach of auditory cue, earcons were combined with two concurrent layers of musical speech (musical transformation of lyrics) to study functional meaning mapping to sound generated by the integration process (Lyricons). It is found that lyricons tend to have significantly strong meaning relevance between sound and the function it represents(Sun & Jeon, 2015). Result showed that lyricons has almost double accuracy rate (83.35%) in compared to earcons (46.53%) in average. Lyricons supports strong single referent relationship and overcome uncertain casual meaning mapping problems of earcons and others factor impacting identification and applicability of auditory cue in synthetic environment. They show reduction in time and error in hierarchal selection of sounds in compared to earcons as well. Moreover, in a comparative exploratory study of lyricons with spearcons, spindex and earcons in visual display(Jeon, 2015) suggests basic outline for lyricons musical structure construction such as pitch ,timbre, note and rhythm to increase meaning relevance and optimal usability. Consequently, broaden the applicability of sound in virtual environment.

3.1.13 Artificial subtle expressions (ASEs)

ASEs are intuitive notification audio for conveying artifacts' internal states to the users(Komatsu, Yamada, Kobayashi, Funakoshi, & Nakano, 2010). Two types of ASEs audios, flat artificial ASEs for high level of confidence and decreasing in pitch ASEs for lower level of confidence in suggestion were tested with robots to investigate impact of robotic suggestion on user acceptance. It is found that flat ASEs robotic suggestion were highly accepted by user while they rejected decreasing in pitch ASEs suggestion which shows ASEs are highly successful in conveying robot's internal state to users accurately and intuitively. This indicates ASEs are highly application in ATM machines, ticketing and reservation system, robotic assistance and support system in avionics and process plants.

3.1.11 Audemes

The review of existing non-speech sound used in HCI above shows that they have been used for a brief representation of objects or events in user interface. They support user notification, warnings and limited information coding such as human emotions and mood expression in communication, remainders and navigational cues in auditory display. This indicates that they are more suitable to carry and express univocal meanings rather than representing large informational content such as messages, educational contents. To address this need of a new category of theme based content rich non-speech sounds for designing acoustic user interface in auditory display, audemes as a short, non-verbal sound symbols were introduced as educational sound symbols(Steven Mannheimer et al., 2009) and learning medium (Steve Mannheimer, Ferati, Bolchini, & Palakal, 2009b)for visually impaired students. Consequently, audemes were found potential in acoustic design to assist blind and visually impaired users (M. Ferati et al., 2009)and enhance the usability of the product(M. Ferati et al., 2011). Audemes used for navigational and thematic landmarks in touch-screen application (AEDIN: Acoustic EDutainment INterface) produced richer and more engaging user experience to the users when browsing a large collection of data, and found highly usable to blind and visually impaired users(M. A. Ferati, 2012a, p. 67).Moreover, in investigating audemes with blind participant to leverage auditory enhancement of workplace(M. Ferati et al., 2012), sequential concatenation of auditory element were found to attribute accurate meaning recognition of audemes than overlapping sounds which further used for fostering guidelines for audemes generation. Audemes were found to have potential in scientific applications including gaming and productivity as well in education as a tool for better memory retention (Meyer et al., 2014).

3.2 Application of non-speech sound in increasing and improving accessibility and usability

3.2.1 Aid Application

Structured non-speech sounds can deliver visual information in more promising ways. One of convincing application is sonification (Kramer, 1993) which could provide continuous data much better than speech to blind users (Brown & Brewster, 2003; Mansur et al., 1985; Ramloll et al., 2000; Roth, Petrucci, Assimacopoulos, & Pun, 2000). Additional application includes light weight, cheap and easy to use Electronic Travel Aids (ETAs), often called as Orientation and Mobility, such as wearable instruments and cell phones that helps navigation and mobility using auditory display techniques(Dobrucki & Sinusas, 2010). For examples, Navbelt (Shoval, Borenstein, & Koren, 1998) and Tyflos System (Bourbakis, 2008) which help blind user to avoid collision with obstacles by providing vibration and sound feedback based on surrounding information(Andò, Baglio, La Malfa, & Marletta, 2010; Andò, Baglio, Marletta, & Pitrone, 2009). Others such as camera-based audio representation of visual content where horizontal and vertical dimension of camera vision are mapped with continuous frequency based non-speech sound (Auvray, Hanneton, & O'Regan, 2007).Similarly, wearable audio navigation system (Walker & Lindsay, 2006; Wilson, Walker, Lindsay, Cambias, & Dellaert, 2007) that assist persons with visual disability in communication and navigation surroundings.

More specific application of non-speech sounds in aid includes combination of earcons and auditory icons non-speech sound used as a supportive information to boost effectiveness and memorability of education contents along with speaking avatars in e-learning interface (Alseid & Rigas, 2011)were found increased user satisfaction and memorable of educational content. Instrumental sounds and sonic effect used to represent 3-D image in a context aware aid system for visually impaired and blind users (Gomez, Bologna, & Pun, 2012) found promising in providing awareness of environment context and alerting them for possible stumble and obstacles. Non-speech auditory representation of physical world, ColOr (Gomez Valencia, 2014) as an audio based sensory substitution device which could sense and map color and depth into instrumental music found usable in spatial navigation and awareness to blinds and visually impaired users.

Moreover, in compared to conventional touch keyboard ,enhanced auditory feedback techniques in Korean touch screen keyboards(EAF) in mobile phone (Y. Park, Heo, & Lee, 2015)found improved usability of the touch screen keyboards. EAF used phonetic auditory cue (feedback generated depending up on acoustic phonetic features of human voice) to replicate default sounds of mechanical keyboards present in most design of touch screen keyboard. Test with 30 volunteers where data were collected in correspond to participants all input data including corrected error, transcribed text, correction time demand, input before corrections, time duration for typing phrases, and keystrokes per phrase; result shows that auditory enhanced keyboard produce better auditory cue than conventional keyboard which increased user performance in all dimension of test. Similar, sonification used in representing electroencephalography(EEG) data in the form of piano note in order to cover the limited screen space for large amount of overloaded visual data in conventional visualization techniques (GAVIN, JEDIR, & NEFF, 2016)showed potential application in representing EEG data while visualization alone was challenging for users. Research suggested that sonification can be used to reduce overloaded visual content and lift cognitive load of the users while dealing with complex source of multivariate data on the screens in multimodal approach to EEG data presentation. Likewise, Integration of spatial sound and beeping cues(by using HTRFs) with music in mobile phone for navigation of destination to cyclist and pedestrian(Albrecht, Väänänen, & Lokki, 2016) allowing them to follow their destination without looking at map and enjoying music, two guidance system were developed; route and beacon. In experiment with pedestrian and cyclist in city center and suburban, both navigation was found useful and preferred subjectively depending in the navigation environment. Route guidance was found better preferred for unfamiliar and uncertain environment to track the destination while beacons was considered more non-interferent enjoyable options for familiar environment.

3.2.2 Mathmatics

Mathematics mostly use visual structures to represent diagrams, formulae and notations in explanation which creates major problems for visual impaired scholars and users to access and grasp the content out of mathematical equations, theories, calculation and applications(Karshmer, 2007). In an attempt to investigate and evaluated auditory cue to enhanced spoken mathematics for visually impaired users by reducing ambiguities and barriers introduced by traditional TTS while replicating mathematical equations with synthesized speech and braille (Murphy, Bates, & Fitzpatrick, 2010)a mixed implementation of auditory cues, spearcons along with binaural spatialization was surveyed online with 56 visually impaired and sighted users. Non-speech sound added to represent structure and scope of mathematical content were found intuitive and potential in increasing accessibility for screen reader users. Result outlined that participants has significant higher accuracy with slower spearcons used to represent fractions and superscripts and subscripts, visually impaired user has lower accuracy rate than sighted user indicating that impaired user face much difficulties in processing verbose lexical multiple-choice equations, spearcons supported higher accuracy for representing opening brackets in equation while beep like sounds dominates for closing brackets ,and higher pitch voice was preferred mostly for subscription and superscriptions and many others. Similarly, in a pilot evaluation of audio rendered mathematics (EAR) with 2 blind and 5 sighted participants (Kacorri, Riga, & Kouroupetroglou, 2014)shows that structural elements in matrix representation of data in compared to graph tree representation of reference data mathematical expressions had highest error rate. The study highlighted basic measures in integration of audio to represent different mathematical content have different usability requirements.

Application at representing and recognizing shapes and gestures of object through spatial representation of non-speech sound(Sanchez, 2010) includes use of different pitch and intensity corresponding to the different shape and size of object. And users without any visual access are found to be easily able to follow the sound reference and recognize object. Similar, auditory cues to enhance accessibility to mathematical materials and eliminates impurities induced by synthetic speech, where non-speech sound with modified speech were able to replicate mathematical formulae in auditory mathematics design (Murphy et al., 2010). Similarly, Integration of earcons and spearcons to represent structure of mathematical equations along with synthesize speech to deliver the content of mathematical equations were used in mathematical library, called SpatialMaths(Fitzsimons, Murphy, Mulwa, & Fitzpatrick)to increase the usability and accessibility of mathematics for blind users. The library use standard MathML (Mathematical Markup Language) for content rendering and synthesizer. Brackets and nested layer structures in equation were represented by using beep like sound earcons along with spearcons for fractions, subscription and superscriptions structures. In mathematical content rendering process, series of rendered strings were converted into audio object by synthesizer layer, and consequently passed to 3D audio layer for presentation. Moreover ,study for increasing effective understanding and usefulness of synthesized speech for mathematics (Frankel & Brownstein, 2016) , MathPlayer software tested for rendering algebraic math expression of secondary school with certain prosodic and lexicon modification such as adjustment in speech rate, pitch, volume and pause length.

Integration of audio with tactile feedback to increase accessibility in graph based educational content(DiGVis) to visually challenged students showed effectiveness of auditory cue in inclusive learning (Syal, Chatterji, & Sardana, 2016). Study with 28 visually impaired users indicated that DiGVis enabled them to recognize visual and logical layout and connectivity of edges, elements of graphical content in directed graphs and flowcharts. Method of audio combination with touch design found applicable to facilitate physically accessibility to virtual spatial workspace and multimodal interface to non-visually representation and comprehension of complex design and mathematical content to visually impaired users group. Similarly, an an exploratory study on auditory cue for improving navigation and comprehension of code in visual programming environment to visually impaired programmer (Ludi, Simpson, & Merchant, 2016)showed improvement in their performance based on the type of auditory cues. In 3 different trail test with 7 visually impaired convincing programmers, it is found that navigation time for a particular code in pre-generated test source code linearly increases with spearcons, speech and earcons auditory cues environment while best comprehension of code found with speech auditory cue followed by spearcons and earcons. Similarly, participants rated usability of auditory cues at navigation and comprehension of code in programing environment both in favor of speech followed by earcons and spearcons. Overall, study highlighted the application of non-speech sound in designing accessible programming platform for visually impaired programmers.

3.2.3 User Interface: Menu and visual icons

Studies had proved that auditory enhancement of visual icons in computer user interface improves their usability and performance in terms of navigation, accuracy, error recovery, easement and time demand. For example, structured sound earcons were found more usable and effective than non-structured sound in hierarchical menu experiments (S. Brewster, V.-P. Raty, & A. Kortekangas, 1996), auditory menu were found better than visual menu in a visual scanning system(S. A. Brewster et al., 1996), earcons based sonically enhanced buttons and scrollbars to overcome visual information load in display(Stephen A Brewster, 1997) improved usability by reducing subjective workload and time to recover from errors, sonically enhanced button of different size to reduce size of visual icons on small screen space of handheld device were fond effective and improved usability for both standard and small button sizes(Stephen A Brewster & Cryer, 1999).Moreover, study with sighted and blind user integration of single tone, double tone , alphabetical and proportional grouping of earcons in auditory scrollbar design resulted into best performance and positive subjective feedback(Yalla & Walker, 2008).Consequently ,favored proportional groupings for best auditory scrollbar design. The study highlighted auditory design for making visual menus accessible to the blind user and those who benefits from auditory feedback received during menu navigation.

Spearcons (Walker, Nance, & Lindsay, 2006a) in auditory menu design were found effective (Suh, Jeon, & Walker, 2012; Walker et al., 2006a) .They drove better performance and accuracy, as well as more flexible menu structures. They severed effective visual cue at improving navigation (Palladino, 2007) ,and auditory cue in mobile phone (Walker & Kogan, 2009) which comparatively improved performance than both text to speech(TTS) and visual menu . findings led to enhanced design in both visual and non-visual context. Likewise , in an empirical experiment at improving navigation in advance auditory menu(Walker et al., 2013), spearcons were found outperforming auditory icons and earcons in accuracy, efficiency and learnability. Participants found spearcons easier to learn and remember. Performance at TTS navigation in two-dimensional menu increased with spearcons than without spearcons although addition of sound increased the length of auditory menu. At the end, study suggested spearcons as effective and better auditory cue in menu based navigation designs than auditory icons, earcons and without auditory cue alone.

Effect of multimodal shared input user interface for mobile devices on menu selection in single-task and dual-task environments(Zhao, Brumby, Chignell, Salvucci, & Goyal, 2013), test with desktop single task setting and simulated driving dual task setting showed that auditory enhancement in visual output significantly reduced visual distraction and inattentions caused by visual output. Participants were instructed to complete a series of menu selection quickly and accurately in both single and dual task scenario while performing simulated driving of car setting 65 miles per hours. Effect of auditory enhanced menu only, visual menu only and auditory enhanced visual menu feedback and menu layout on participant's preference and performance in both task scenarios showed application environment and task situation dependencies. Visual menu was preferred over others in single task situation and delivered better performance while auditory enhanced menu was perceived well and found significant in menu navigation in dual task condition (while driving). Similarly, Spindex as auditory cue in advance auditory search and navigation enhanced interface in mobile phone (Gable, Walker, Moses, & Chitloor, 2013)for searching and scrolling through long lists of contact or music in dual task condition while driving vehicles was found lower down distraction. Test with 26 students in scrolling 150 lists of songs for search operation on mobile while driving and performing lane changing task resulted higher user eye sight flexibility, preference, significantly less subjective workload and reduced off lane attentions and distractions. Moreover, Auditory menu cue at improving menu navigation with in-vehicle technologies in dual task scenario(Jeon et al., 2015), spearcons and spendix were rendered as auditory menu in secondary task while participants were simulating driving by playing perceptual motor ball catching game as primary task. In all experiments, menu presented as visual only, with TTS and auditory menu, participants preferred auditory menu and rated lower workload than visual only and with TTS. It indicated that auditory cue enhanced menu offers user more flexibility and accessibility in in-vehicle technologies which consequently drives effective and efficient safe driving in multitasking contexts.

3.2.4 IMAGE

Auditory aid for virtual handicapped user to provide sense about visual aspect of environment through audio wave enabled user identify source and direction of sound in process of perceiving surroundings(Stanton, 1982). Auditory patterns representation of time-multipixel invertible image(Meijer, 1992) found preserved visual information to some promising extent to the blind people. They study simulates image to sound and inverse sound to image real time conversion of image up to 64 by 64 pixels with 16 gray tones per pixel with designed focused on development of low cost portable system for visual aids. The study reveals convincing real time image to sounds comprehension mapping as an auditory representation of image to substitute limited vision. Similar, online and real time audio translation of image based on edge detection and graded resolution (Capelle, Trullemans, Arno, & Veraart, 1998)with personal computer connected to mounted camera and headphone for optimizing auditory substitution to blind people demonstrate feasibility to some extent and found usefulness for patterns recognition .

Sonification of image to support navigation and object detection such as location and size to blinds users, Navigation Assistance for Visually Impaired (NAVI) (Nagarajan, Yaacob, & Sainarayanan, 2002)maps visual information into acoustic streams for blind users. NAVI uses fuzzy clustering algorithms in features extraction and clustering to produce environmental noise free more clear representation of image content. It incorporates understanding of human vision for clear distinction of object identity and background. The processed image is then converted into audio patterns. Auditory translated of pre-processed image depends on direct mapping of intensity of image pixel into frequency of audible range of 20 Hz to 20kHz. Low frequency used for the mapping because of higher sensitivity to human ear than higher frequency where sound pitch was used to map vertical position and loudness of sound to map intensity of pixels. In training and test with blind users, NAVI direct mapping to auditory cue found usable to substitute some amounts of visual information to aid them in free environment navigation. Similarly, Stereo Vision based Electronic Travel Aid(SVETA) (Balakrishnan, Sainarayanan, Nagarajan, & Yaacob, 2005)used built-in toolkit for pre-processing before signifying visual information into musical tone based on direct mapping and musical octave method. The direct mapping method based on mapping amplitude of sound directly proportional to intensity of image pixels whereas musical octave method uses audible range of musical tone octave to create logarithmic musical tone. The device consists of stereo camera for real-time video streams and headphone for delivering visual information through designed structured auditory tone. SVETA uses pre-computing visual information rather than direct mapping of pixel intensity into sounds loudness in order to enhance objects properties and filtering environmental noise from the object information. Consequently, it helped to produce more pleasant sounds and overcome users from extensive training and perpetual demand to distinguish objects aside background information. In user test with blind and sighted user to simulate collision free navigation of environmental objects among several 12 users group musical octave method is found effective in term of performance and compatibility to identify object location and properties.

Others are system for seeing using auditory feedback (Tkacik, 2011) which translate the video image into audio signal, representing visual content (i.e. spatial layout) via audio and haptic interfaces a technical diagram access tool TeDUB (King & Evans, 2007) found successful at representing electronic circuit and UML(Saldhana & Shatz, 2000) diagrams in both hierarchy and connected graph to blind people. Added non speech sound to speech interface(King & Evans, 2007) allowed user to explore and identify graph and tables with significant decrease in subjective workload, reduced temporal demand and errors as compared to speech interface and haptics graphs. Other significant studies with blind people in representing image includes converting image to sound via method of edge detection(Krishnan, Porkodi, & Kanimozhi, 2013), converting extracted image information into haptic environment(Nikolakis, Moustakas, Tzovaras, & Strintzis, 2005). Similar advance auditory mapping of images to blinds people includes acoustic feedback on touch screen (Banf & Blanz, 2013)provides direct access to image. The multi-level system overcomes limits of manual acoustical object recognition by incorporates machine learning algorithm for object recognition and classification, and maximizes auditory translation of image content such as color, edges, surface. The system enable blind user to explore image on touch screen and receives acoustic feedback associated to visual features of image (selected surface, color). It uses acoustic feedback sounds based on perpetual and semantic consideration rather than instrumental MIDI.

Moreover, auditory feedback was found superior over visual feedback in support vector regression(SVR) based Brain Computer interaction(BCI)(Roussel, Negishi, & Mitsukura, 2016). In study, 14 users were instructed to perform task based on both auditory and visual presentation conditioning, and after trainings and repetitive experiments auditory instructions were found to increase the accuracy and usability of BCI in comparison of visual conditions. This indicates replicating values of visual information with auditory representation in BCI.

4. Research methodology

Research in this paper is a within-subject design that includes both qualitative and quantitative data. Quantitative data include the workload perceived by participants when recognizing test images, and usability ratings of the test prototype. Qualitative data covers participants' comments while performing tasks, which was further calibrated with their post-test reactions and feedback. Observation data sheets comprised other aspects of user data and trends, such as time spent on each prototype and attempts made on each test image during the experiment phase.

Purpose of the study was to investigate and explore significance of non-speech sound, especially audemes to represent web image for screen reader users. Web image enrichment efficiency of audemes was assessed based on perceived subjective workload that participants felt while accessing test images none-visually through TTS application. NASA Task Load Index (NASA-TLX) tool was used for assessing subjective workload. Similarly, usability factor of audemes in encapsulating web images to increase non-visual accessibility was evaluated along with the System Usability Scale(SUS): How much usable the webpage user find while web images are enriched with audemes to increase accessibility for screen reader users? Usability ratings of test prototypes was expected to provide information about user satisfactions level with audemes enriched web images.

However, subjective workloads and SUS ratings yield statistical data for analysis; data alone won't be sufficient to conclude significance of the study over current state of art (alt text). Thus, experiment was conducted on behalf of audemes in comparison to alt text. Consequently, Data were collected on the behalf of both audemes and alt text in a comparative evaluation situation. Thereby, Conclusion was drawn based on comparative analysis in order to explore and reflect content enrichment significance of audemes in enriching web image for increasing accessibility to the blind people. Besides these, we collected observations data about test participants to reflect accuracy and error rate at identifying web image, learnability and mnemonic memory impact of methods used to represent web image.

4.1. NASA- Task Load Index (NASA-TLX)

The NASA-TLX is a widely used, subjective multidimensional workload assessment tool which was developed by the Human Performance Group at NASA's Ames Research Center. It measures perceived workload of task performer to assess complexity of the task, system and other aspects of user performance. It has been in use in a variety of domains including complex socio-technical domains since last 30 years. It is a familiar tool in the HCI community for accessing the subjective workload of the experiment. NASA TLX (Hart & Staveland, 1988) measures factors like workload demands, performance , time pressure, and frustrations by focusing on the tasks that have well defined objectives. . it is also concluded that TLX provides a sensitive indicator of overall workload as task varies by the nature of physical and mental demand.

Originally, it has two parts: the total workload is break down into six different subscales which serve as one part of the questionnaire, and second part deals with defining individual weight of these subscales. Subscales are:

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

There is a description for each of these subscales which can be found to help participants answering accurately:

Table 4.4.1 NASA TLX scale rating definition

Scales	Endpoints	Descriptions
Mental Demand	Low/High	How much mental and perceptual activity was required (for examples; thinking, deciding, remembering, searching and calculating)? Is the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/ High	How much physical activity is required (for examples; pushing, pulling, turning, controlling and activating control keys)? Is the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure does the user feel at the task? Is the pace slow and leisurely or rapid and frantic?
Effort	Low/High	How hard does user have to work (mentally and physically) to accomplish defined level of performance?
Performance	Good/ Poor	How successfully does the user accomplish the goal of task set in the experiment? What is the user satisfaction level in accomplishing goal?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed, does the user feel during task?

All subscales are rated on bipolar basis ranging from 1 to 100, Low and High single adjective anchored at both end. The whole scales of 100 points are divided with 5 point intervals. Generally, these ratings are then combined with the task load index. That is overall workload rating determined from a weight combination of individual scores on the six subscales.

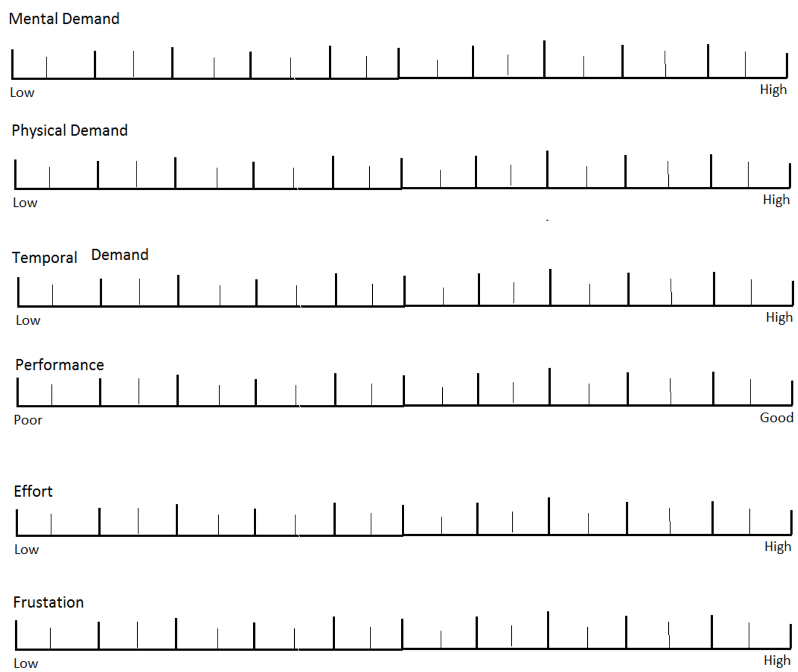

Figure 4.1 NASA TLX Rating Scales

The NASA TLX has several excellent features which makes it a good tool for evaluating productivity of the application or system. For example, it is easier to use, quickly to fill out workload questionnaire for participants. And importantly, it can produce both classified and a single overall rating of the workload. It is a standardized survey tool that also makes it appealing for researcher to use and report study results on the NASA TLX without explaining the details of the measurement. And, the reports are also remains meaningful to the others researcher as well. The study reported that significance of tested workload effects has high (+0.769) correlation (Battiste & Bortolussi, 1988; Hart & Staveland, 1988). Similarly, claims were made by study that NASA TLX is a valid and reliable measures of workload(Corwin et al., 1989).

Since NASA TLX has been used for obtaining workload estimation in a number of disciplines(Hart, 2006), studies had explored the relationship of tool with others numbers of performance factors. For examples, relationship between workload and subjective fatigue for a regularly rotating 12 hours work schedule (Baulk et al., 2007), relation of mental workload with subjective experience(Cao, Chintamani, Pandya, & Ellis, 2009),trust (Turner, Safar, & Ramaswamy, 2006) and with others such as effects of physical parameters (DiDomenico & Nussbaum, 2011), psycho-physiological measures(Di Stasi, Antolí, & Cañas, 2011), emotional experience(Jornet et al., 2008) and so on. There are also number of modification made to the original version of NASA TLX scales; one of them was an un-weighted rating know as raw TLX (RTLX). It became common and popular because studies had found high correlation between the weighted and un-weighted scores (Moroney, Biers, Eggemeier, & Mitchell, 1992). Other modifications are replacing original subscales and tailored with descriptions(Ihm et al., 1998), delaying reporting of workload ratings (Moroney et al., 1992) were all found high correlation with conventional NASA TLX procedure.

NASA TLX measures subject workload, generally by asking participants to describe the workload they experience while performing task. It does not measure the details about the nature and objective of task, it only focusses entirely on the participants' feelings about their workload. There are basically two types of subjective workload measurements methods:

- Subjective numerical measurements, and
- Subjective comparative measurements.

In subjective comparative measurement, participants experience tasks complexity, and scale comparatively workload pressure under different task circumstance or grounds or among the tasks such as subjective workload dominance (SWORD) technique. It believes that a more reliable evaluation of the workload experienced by the test participants in various task situation can be achieved when the task situation is compared with one another rather than evaluated in absolute term. This subjective comparative technique does not require participant to assign numerical or ordinal ranking or scaling to the workload questionnaire. Thus, eliminates all the problems associated with scale usage, and uses key words and phrases to describe numerical levels of workload. However, it is reliable and feasible for comparative workload evaluation of task among different task environment, lack of numerical data associated the test measures has some serious drawbacks as well. It provides a clear judgment between the comparators based on predefined phrases of workload evaluation sheets not subjective numerical data. So, with exact numerical values it is impossible to conclude that how much workload does the participants exactly experienced in certain task or task environment, evaluate acceptance of workload scores against standard workload value and amount of workload difference between tasks. Most importantly, as it does not provide statistical analysis individual workload factors and their impact on overall workload scores.

Therefore, subjective numerical measurement was chosen for the experiments where overall workload calculated based on individual scores of TLX subscales as well as individual workload factors are analyzed individually. Most importantly, it facilitates the invariant as well as multivariate statistical analysis of the workload factors against overall workload.

4.1.2. Analysis

Generally, second part of TLX deals with creating an individual weight of all individual subscales by letting participants compare those subscales pair wise based on their perceived importance. Then, weight of subscale is defined based on how many times the user chooses which measurement are more relevant to overall workload. This weight of individual subscales later used for multiplying individual raw scores in a process of creating workload task index. However, most recent common modification made to NASA-TLX was to eliminate the weighting process all together and analyzing each workload factor individually. The method has been referred to as Raw TLX (RTLX). It has gained popularity because it is simpler to apply. The ratings are simply averaged or added to create an estimate of overall workload. In the 29 studies in which RTLX was compared to the original version, it was found that RTLX is either more sensitive(Hendy, Hamilton, & Landry, 1993), less sensitive(Liu & Wickens, 1994), or equally sensitive(Bittner, Byers, Hill, Zaklad, & Christ, 1989).Moreover, while experimenting with one-dimensional and multidimensional measures of workload (Hendy et al., 1993) it is suggested that;

“If and overall measure of workload is required, then a single variant measure is as sensitive as an estimate of derived from multivariate. If a single variant measure is not available then a simple un-weighted additive method can be used to combine ratings into an overall workload estimate.”

In another study, it is also suggested and reported that the pre-rating weighting method is unnecessary since the correlation between weighted and un-weighted scores was +0.94(Bittner et al., 1989; Moroney et al., 1992).Therefore, RTLX was selected and found more reasonable because it facilitates subscales ratings analysis as well as overall workload scores. The subscales ratings help to pinpoint source of a workload and performance problem as well.

4.2 System usability scale

Moreover, appropriateness and usability of the purposed image representation method will be assessed by System Usability Scale (SUS)(Brooke, 1996). It describes a reliable, low-cost usability scale that can be used for standard assessment of the system usability. Usability is not a quality that exists in exact real or absolute sense. It can only be defined regarding the context in which the system is used. While SUS is only intended to measure perceived easement of user or operation. A recent study (Lewis & Sauro, 2009) showed that it provides a global measure of system satisfaction and sub-scales of usability and learn-ability. It is proved to be a valuable evaluation tool, being robust and reliable. It also correlates well with other subjective measures of usability such as SUMI (Kirakowski, 1996).However, there are not any clearly defined measures; they must be dependent on the way in which usability is defined. According ISO 9241-11, measures of the usability in the experiments will cover following characteristics of the system;

1. Effectiveness: it will define the ability of purposed method to represent image none visually
2. Efficiency: it will define how fast and good the participants will grasp and recognize the image information
3. Satisfaction: it will define the participants' reactions to using the purposed method to represent image

Benefits of using SUS include:

1. It is a very easy scale to administer to participants
2. It can be used on small sample sizes with reliable results

3. It is a valid method which can effectively differentiate and compare between usable and unusable systems

Additionally, it is very difficult to draw comparisons of usability across different system because a particular design feature has proved to be very useful in making one system usable does not necessarily mean that it will do so for another system in same context. But, we could avoid these possible misleading in generalization by presenting similar information in same designs. Thus, we kept an entire difference only on representation of image in non-visual context. Experiment will be conducted against two different prototypes which have test images from similar context in same page layout and same user interaction design.

In this comparative assessment of usability across two prototypes, subjective assessments will be obtained using questionnaires and attitude scales. A modified system usability scale worksheet was provided to the participants to fill out their experience after the test. Consequently, result was used to evaluate usability and draw comparisons between the purposed methods.

4.2.1 Scale

SUS is a Likert scale (Bangor, Kortum, & Miller, 2009)that generally stand on forced choice questions. Statements are made to be responded by the test participants. Answers are used to indicate the degree of agreement or disagreement with the statements on 5 to 7-points scale. It covers a variety of system aspect such as the need for support, training and complexity, and thus has high level of valid measurements for system usability. It has a set of 10 forced questions with 5 response options:

1. I think that I would like to use this webpage frequently.
2. I found this webpage unnecessarily complex.
3. I thought this webpage was easy to use.
4. I think that I would need assistance to be able to use this page.
5. I found the various functions in this webpage were well integrated.
6. I thought there was too much inconsistency in this webpage.

7. I would imagine that most people would learn to use this webpage very quickly. I found this webpage very cumbersome and awkward to use.
8. I felt very confident using this webpage.
9. I needed to learn a lot of things before I could get going with this webpage.

The SUS response format:

Strongly Disagree 1	2	3	4	Strongly Agree 5
○	○	○	○	○

Figure 4.2 System Usability Scale for measurement

Scoring method:

1. For odd number question: subtract one from the user scores.
2. For even-numbered question: subtract the user scores from 5
3. The score values range from 0 to 4 (with four being the most positive response).
4. Add up all converted scores for each user, and multiply that total by 2.5. The result will fall in the possible range from 0 to 100. The SUS score above a 68 would be considered above average and anything below 68 is below average.

5. Designing Experimental Prototype

To test the significance difference between alt text and audemes, two webpages were developed as test prototypes[1]. The prototypes have single home page for the experiments. They contain images as test objects along with information and instruction regarding test procedure. Structure of test page was divided in to two different parts. First part of pages contains information about the prototype and instructions on how to use whereas rest of second part lists set of images as test objects. Each prototype has exactly same design and layouts. Information in the pages was organized in exactly same chronological order. Both test page has exactly same number of test images which are organized in two rows in a set of four images per rows. The whole difference between the two prototype was kept only on the methods of enriching test images for screen reader user which was key to experiment. Consequently, after brief training with training samples participants were asked to take test and reflect their views on the paper based experiments' questionnaire. The entire difference between the test prototypes was meant to cover and present the participants' reflection on the significance difference between the techniques of non-visual represent of web images for screen readers' personalities.

5.1 Prototype A

This was a test prototype where images were encapsulated and explained with audemes enrichements. Each image was enriched with content and context riched musical non-speech audemes which simply explains the context and contains of the image to the users. Test images were randomly taken from the Google search engines, and appropriated audemes were developed in order to represent the image in auditory display context. The prototype was designed to access and operate from blind prospective where participants could reflect and perform test on behalf of blind or screen reader user. Thus, test page was developed and structured in a consideration with text to speech access environment. Textual content of the page was supposed to translate into speech by TTS in non-visual access context. Users had to accommodate themselves in non-

[1] https://audemes.000webhostapp.com/ratan/

visual context, and relies entirely on the screen reader all over the training and experiment sessions.

5.1.1. Design

Design of prototype focuses on two important issues. Easy keyboards navigation of webpage content and integration of audemes on test images. Navigation mainly deals with use of minimum, easy and well-known keys combination to navigated content of test pages for screen reader users whereas integration of audemes deals with triggering corresponding audemes audio file upon keyboard navigation of test images in the test prototype.

5.1.1.1 Navigation

Generally, users navigate internet content by using mouse than keyboards. But, some users prefer keyboard commands for efficiency, which is also helpful for the users with certain disabilities. For example, people with physical impairments have difficulty with the fine motor movements required to operate a mouse; blind users who rely on assistive technology such as screen reader can't see where to click by mouse.The big difference between the keyboard and the mouse is that when users navigate through the keyboard, the access to the content on the screen is sequential; users must tab through all the page content one by one before reaching to the particular content of interest. In contrast, a mouse user can inspect visual elements on the screen and move the cursor directly to the element he wants to click. Thus, the mouse allows direct access to the content on the screen. This is the reason why a sophisticated and wonderful website might be completely useless to someone who cannot interact and access their controls. In scenario, a keyboard-friendly websites design, which make these interactions possible for users who cannot use the mouse, not only increase the accessibility of websites for disable user but also makes them more usable for normal user as well.

Thus, prototype need to be keyboard accessible for simulating experiment on behalf screen reader users or blind perspective. Participants must access and take test from non-visual perspective. So, they have to depend on keyboard accessibility over entire experiment duration in order to access the test prototype. For the keyboard-friendly website design, WCAG 2.0 has defined success criteria for html "<tabindex>" tag[2]. Tags attribute value makes webpage element

[2] https://www.w3.org/TR/WCAG20-TECHS/F44.html

focusable on keyboard selection. The integer value of the attribute also defines the order of key board navigation of "<tabindex>" element which are as follows:

- A negative value (-1) means that the element should be focusable, but should not be reachable via sequential keyboard navigation;
- Zero value (0) means that the element should be focusable and reachable via sequential keyboard navigation, but its relative order is defined by the platform convention;
- A positive means should be focusable and reachable via sequential keyboard navigation. Its relative order is defined by the value of the attribute where the sequential navigation follows with the increasing value of the "<tabindex>". If several elements share the same "<tabindex>", their relative order follows their relative position in the webpage.

The prototype page design was divided into two layout class: "information" and "test content". The first information class "<div class="jumbotron" >" was again divided into three "<tabindex>" sections: "Name of prototype", "Introduction" and "instructions". Whereas "test content" class "<div class="container-fluid bg-3 text-center"> "organized eight test images in two "<tabindex>" rows. The value of all "<tabindex>" attribute was defined zero. That means all the "<Div tabindex=0>" sections are focusable and selectable through sequential keyboard navigation in an order information was presented on test pages. Participants could access test page in a sequential order from starting section (information sections) to the end (test sections in the experiments).

5.1.1.2 Audemes Integration

The intension behind prototype design was the process of enriching test images with audemes description where user should not need to deal with additional keyboard keys more than usual tab and shift keys combination for keyboard navigation and accessibility. Audemes as an audio clips; in that case, HTML has a defined "<Audio>" tag for audio or sound clip integration in webpage. The HTML "< Audio>" tag is defined to embed audio content in webpage which could reference one or more audio sources. It has "SRC" attribute or the "<source>" element for defining the source and address of the audio file. The other global attributes of the “<Audio>” tag[3] are as follows:

AutoPlay

It is a Boolean attribute (predefined “true” if specified), the audio will automatically begin playback as soon as it can do so, without waiting for the entire audio file to finish downloading.

Auto Buffer

It is also a Boolean attribute (if specified); the audio will automatically begin download, even if not set to auto-play. This continues until the media cache is full or the entire audio file has been downloaded, whichever comes first. This should only be used when it is expected that the user will choose to play the audio.

Buffered

This attribute determines which time ranges of the media have been buffered. This attribute contains a Time-Rage object.

Controls

This attribute allows the user to control audio playback such as volume control, seeking, and pause/resume playback.

Loop

It is a Boolean attribute (if specified), will automatically seek back to the start upon reaching the end of the audio.

Muted

It is a Boolean attribute which indicates whether the audio will be initially silenced. Its default value is false.

[3] https://www.w3schools.com/tags/tag_audio.asp

Played

It is a Time-Range object which indicates all the ranges of the audio that have been played.

Preload

This enumerated attribute is intended to provide a hint to the browser about the best user experience. It may have one of the following values:

- None: indicates that the audio should not be preloaded.
- Metadata: indicates that only audio metadata (such as length) is fetched.
- Auto: indicates that the whole audio file could be downloaded, even if the user is not expected to use it.
- The empty string: synonym of the auto value. If not set, its default value is browser-defined. The spec advises it to be set to metadata.

Volume

It defines playback volume which range from 0.0 (silent) to 1.0 (loudest).

Audio element also supports various events-handling while accessing html documents. So far, our goal was to integrate audemes file with image object so that upon keyboard navigation of the image element should play audemes file which basically explain or describe the image content to the users. However, all the features in “< audio>” tag are suitable for the purpose; it increases the complexity in controlling and accessing them through keyboard for screen reader applicants. User must go through multiple steps and use multiple keys combination for the use which increases barriers to the visually impaired uses. Thus, in order to overcome the shortcomings of the “<audio>” tag, extra java scripts were accommodated along with audio tag. Here is the sample of the code from the prototype:

```
<p>  <div tabindex="0"> First Picture </div> </p>

<div tabindex="0" onfocus="javascript:play_single1_sound(this);"
onfocusout="javascript:stop_single1_sound(this);">
```

```
<audio id="audiotag1" src="Kids_cycling_audemes.wav" preload="auto"></audio>

<image src="kids_Cycling.jpg" alt="" style="width:302px;Height:228px"/>

<script type="text/javascript">

var music1= document.getElementById('audiotag1');

function play_single1_sound()

{

music1.load();

music1.play();

}

function stop_single1_sound()

{

music1.pause();

}

</script>

</div>

</div>
```

The code automatically enables audemes on keyboard tab navigation of test image in the prototype. Similarly, audio stops if the user shifts navigation to other elements of the page. Shift of navigation temporally pause the audio, and on the consequence of regain of navigation browser reloads the audemes file and plays them from the beginning. Events related to keyboard navigation were used to handle and control the audemes audio file. Generally, tab navigation generated two kinds of events while user moves back and forth between the webpage elements. They are:

- OnFocus: The event is emerged when element gets navigation.
- OnFocusOut: The event is emerged when element lost navigation.

Following things are taken into consideration while integrating audemes file in test images:

1 Audemes should not interfere and overlap with the sounds of other audemes and screen reader
2 Audemes should run exactly on the navigation of image.
3 Shift of navigation should instantly stop the audemes, so there would not be simultaneous run of more than one audemes from different image in the prototype.
4 Regain of navigation should start play audemes from the beginning not from the previous pause position.
5 Volumes of audemes should be compatible with screen reader and test device.
6 Length of audemes should not be more than 4 seconds

Consequently, following functions were used to implements the considerations:

- Function play_single_sound() : The function comes into action when image in the prototype gets focus. It has two statements for two different purposes. They are: -

    1. music1.load(): it reinitialized and load the audio files irrespective of previous audio state. So, the audio start play from beginning on re-navigation.
    2. music1.play(): it generally triggers the audio file to run and audio start play instantly.

- Function stop_single_sound(): This function invokes when the image in the prototype lost focus. It contains single statement:

    1. music1.pause(): it stops the AUDEMES to run. It generally froze the audio in current state.

5.1.2. Designing audemes for the test image

Audemes are new category of semantically more flexible and multi-vocal (having more than single referenced meaning) non-speech sounds. Audemes could be generated from sounds occurring in nature as well as abstract and musical sounds, such as from songs. In other words, they could be combination of anything such as music snippets, even an unrecognized human voice, but a piece of non-speech sounds. It was described (Steve Mannheimer, Ferati, Bolchini, & Palakal, 2009a) as;

"Audemes are short, non-speech sound symbols, under seven seconds, and are comprised of various combinations of sound effects referring to natural and/or artificial, man-made context, abstract sounds and even snippets of popular music".

Meaning of audemes starts wide open and eventually narrow downs with the introduction of additional sounds. Consequently, multiple audemes finally merged into a atomic meaning. However, there are not any hard and fast rules to generated audemes. Creation of audemes entirely depends on semiotics (Pirhonen, Murphy, McAllister, & Yu, 2006) and modes of listening(Vickers, 2013). In terms of semiotics, for example when an audemes of "key juggling" is heard, the sound is recognizable and the object clearly defined. When the object is known, it is easier to develop a meaning. Similarly, meaning also depends on the causal and referential modes of listening. Causal mode of listening, first help to establish semiotics link to the object generating sounds, then reference help to sharpen and finalize the meaning of audemes across various state and situation of object generating sounds. Guidelines(M. A. Ferati, 2012c, p. 105) described in designing audemes for Advanced Support and Creation-Oriented Library Tool for Audemes (ASCOLTA) (M. A. Ferati, 2012c, p. 102)were followed while developing audemes for the experimental prototypes. Those are:

Guideline 1: Serially Concatenate Sounds: Sounds are searched based on a match with the image characters (contents or objects), then concatenated in a serial fashion to simulate characters in image are in motion and linked with the events.

Guideline 2: Mix Sound Types: Recorded sounds of real life phenomenon (such as crowds, people enchanting, street sounds) and those from the nature having natural association with the objects and events are mixed with primary sound those obtained from serial concatenation to simulate context of test image. Contextual description of image covers non-verbal, non-linguistic contents such as environmental, sensational and emotional information about the character and events in the image.

Guideline 3: Music First: Designed audemes, first starts with audio effects to simulate context of the image in order to established foundation of environmental and situational state where characters or object in image come to play. Then, followed by concatenated primary sound to convey information about the characters or objects of image. Finally closing of audemes again marked audio effect to simulate change in the context after character came into play in the context.

Guideline 4: Causal and Referential Sounds Have Priority: signature sounds related to the identity of objects (such as animals voice, abstract sounds for object like car, plane, natural sounds such as wind breeze, water flow, rain) were used for identification of the characters or objects in image first flowing by additional concatenated sounds to link character to the events in image for animated effect.

Audemes for prototype were also developed on the similar context of casual and referential mode of listening .Audemes files were developed on consideration that audemes symbols made up of 2-5 individual sounds lasting 3-7 seconds improves encoding and long-term memory (Steve Mannheimer et al., 2009a) and sequential concatenation of different sound yielded the highest meaning recognition of audemes (M. Ferati et al., 2012, p. 84). Necessary concepts and raw audemes were taken from the audemes dictionary[4]. The website contains samples of basic audemes which were further integrated with different sound cues from different music source to create final AUDEMES for the test images. Audemes were kept under

[4] https://audemes.org/dictionary_2014/dictionary_2016.html

4 seconds in a consideration to match the speech duration of alt text. Further, Audacity tool (Waseem & Sujatha, 2014)was used for creating, editing and sharpening AUDEMES.

5.1.3 Audacity[5]

It is a free, easy-to-use, multi-track audio editor and recorder for Windows, Mac OS X, Linux and other operating systems. It was started by Dominic Mazzoni and Roger Dannenberg in 1999 at Carnegie Mellon University, and later released as open-source software at SourceForge.net in May of 2000.The interface is translated into many language which can be used for:

- Recording real time audio streams
- Capturing computer playback on any Windows Vista or later machine.
- Digital translation of analog audio.
- Supports wide range of audio files such as WAV, AIFF, FLAC, and MP2, MP3 files
- Supports imports and exports of wide range of audio libraries.
- Editing and mixing audio files.
- Creating sound effects such as altering speed and pitch of recording and others.

It has following features:

Recording

It can record live audio through a microphone. It can also capture streaming audio.

Import and Export

It could import and exports sound files, edit them, and combine them with other files.

Accessibility

It can be fully operate using the keyboard and support large range of keyboard shortcuts. It supports JAWS(Scientific, 2014), NVDA(Access, 2010) and other screen reader on windows and for VoiceOver (Leventhal, 2005) on Mac.

[5] http://old.audacityteam.org/

Effects

It supports real-time preview, noise reduction, isolate vocals parser, limiter, reverse, truncate silence and many more functionalities.

Analysis

It also supports spectrogram view, plot spectrum, sample data export and contrast analysis mode for visualizing and selecting frequencies for analysis.

5.1.4. Test Content of prototype A

The prototype page contains following eight images as test objects and corresponding audemes for them:

Table 5.1 Test content of prototype A

Test Image	Source	Audemes
	www.e2sport.com	Kids_cycling_audemes.wav
	www.visitoslo.com	New_year_audemes.wav
	c1.staticflickr.com	Lion_Zebra_audemes.wav

Test Image	Source	Audemes
	i.ytimg.com	WAV River_flow_audemes.wav
	cdn.ussoccerplayers.com	WAV Stadium_soccer_audemes.wav
	www.clubedafotografia.com	WAV Seagull_beach_audemes.wav
	il3.picdn.net	WAV Rain_forest_audemes.wav

Test Image	**Source**	**Audemes**
	www.railpictures.net	 steam_train_audemes.wav

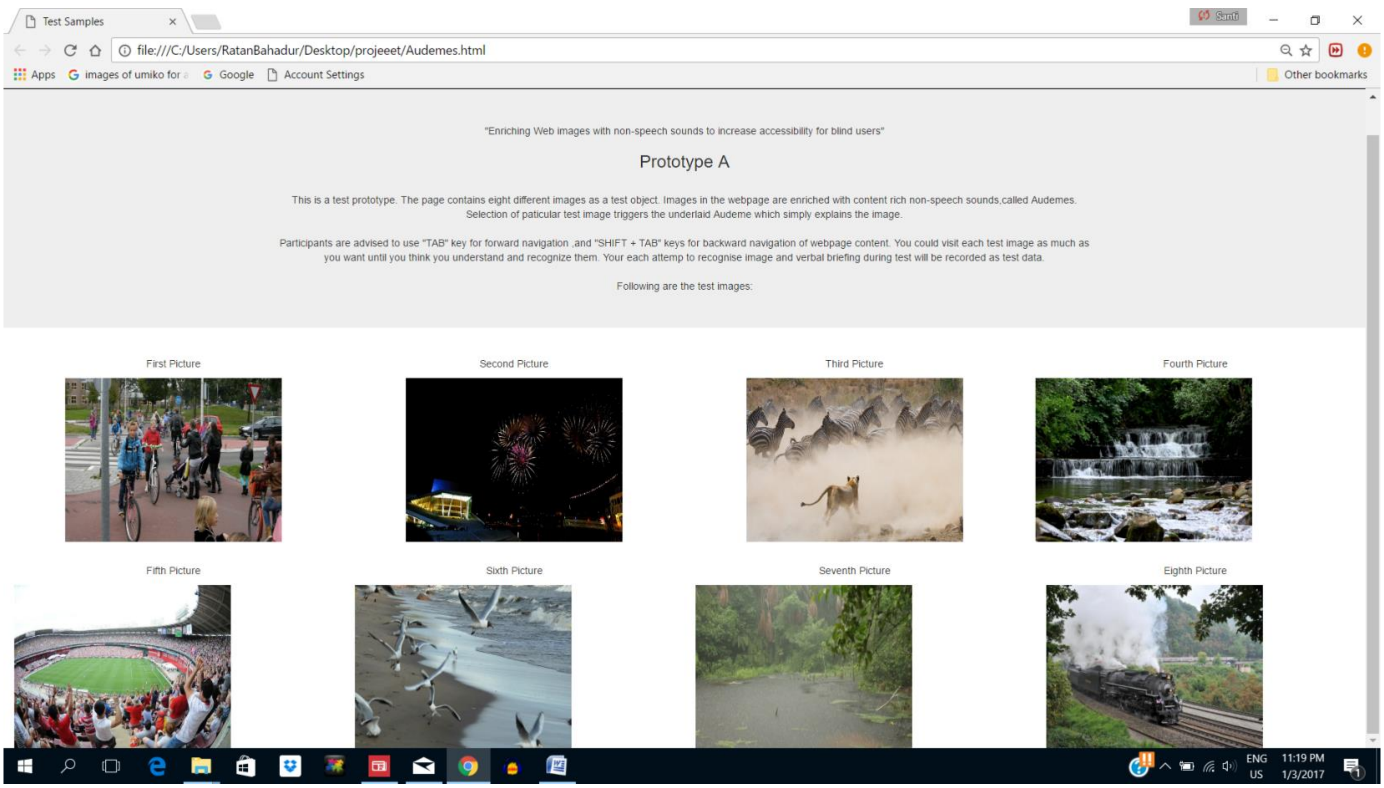

Figure 5.1 Screenshot of prototype A

5.2 Prototype B

Like prototype “A”, prototype “B” contains exactly same number of test images but are explained with ALT Text in place of audemes enrichment. Each image in test prototype includes short textual descriptions which simply summarize the image. Test images were randomly taken from the Google search engines. Whereas prototype contains different set of images with respect to “A” but from similar perspective. In that case, images which were used in first prototype could be used in second and vice versa with no difficulties in representing them through either of techniques. It was also designed to access and operate through screen reader where participants can reflect and perform test on behalf of visually impaired persons. They completely depend on screen reader and keyboard navigation while taking test with the prototype.

5.2.1 Design

Design of the page deals with two issues. Easy keyboards navigation of webpage content and integration of ALT TEXT with test images. Navigation focuses on the use of minimum, easy and well-known keys for accessing prototype whereas integration of ALT TEXT deals with TTS translation of corresponding textual description of the test image upon the keyboard navigation.

5.2.1.1 Navigation

HTML has “<tabindex >” global tag attribute for that purpose which makes webpage element keyboard navigable and focusable. Like first prototype, webpage is divided into two classes: “information” and “test content”.

- The first information block contains class “<div class="jumbotron" >” which is further divided into three “<tabindex>” sections: “Name of prototype”, “Introduction” and “instructions”.
- Whereas test block contains different class “<div class="container-fluid bg-3 text-center">“which organized eight test images into two by four matrix “<tabindex>” rows. The value of all “<tabindex>” attribute was defined zero. That means all the “<Div

tabindex=0>" sections content is focusable and selectable in sequential keyboard navigation order.

5.2.1.2 ALT TEX

Alternatives text describes the information or function presented by the image. This ensures web image' accessibility among people with various disabilities. How to provide appropriate text alternatives to image depends on the context and purpose of image in webpage[6] such as:

- Informative images: They graphically represent concepts and information such as typically pictures, photos and illustrations. The text alternative should be at least a short description conveying the essential information presented by the image.
- Images of text: Readable text is sometimes presented within an image. If the image is a logo, text in images should be avoided. However, if texts are mentioned as image, the text alternative should need to contain the same words as in the image.
- Complex images: such as graphs and diagrams: To convey data or detailed information, the text alternative should be a full text equivalent of the data or information provided in the image.
- Decorative images: The purposes of these images are to substitute visual decoration to the page, rather than to convey information that is important to understanding the page. So, text alternative for such images are not necessary.
- Functional images: These are used for a link or a button which describe the functionality rather than visual image. For example, printer icon to represent the print function or a button to submit a form. In this case, alternative text should need to explain the function of image not the image.
- Groups of images: If multiple images convey a single piece of information, the text alternative for one image should convey the information conveyed by the entire group.
- Image maps: The text alternative for an image that contains multiple navigable areas should provide an overall context for the set of images. Additionally, each individual

[6] http://webaim.org/techniques/alttext

clickable area should have alternative text that describes the purpose or destination of the link.

Moreover, text alternative needs to be determined by the author, depending on the usage, context, and content of an image. They are now serving only available medium to visual impairment, including people with low vision. Images are extensively in use in designing websites. They help to create attractive webpages and severs significant method at presenting information precisely. However, image could be a major informational barrier if were not accessible, especially for blind and visual impaired persons. Thus, alt text enrichment to web image have many advantages such as:

- People using screen readers or browsing speech-enabled websites**:** The alternative text can be read aloud or rendered as Braille
- People using speech input software**:** Users can put the focus onto a button or linked image with a single voice command. Alternative text removes other cluttering of interface and keeps them simple
- Mobile web users**:** Image can be turned off, especially for data-roaming and internet browsing speeding in case of available textual descriptions.
- Search engine optimization**:** Images become index-able by search engines if enriched with alternative textual description

For image integration in webpage, HTML contains “<IMG>” tag with ALT attribute which defines an alternate text for an image. Image in the case where image cannot be displayed or accessed. The alt attribute provides alternative information for an image if a user for some reason cannot view it such as slow internet connection, an error in the SRC attribute, or visual impairments. However, the main intension behind ALT TEXT is that computers and screen readers cannot process an image and determines what the image presents. It also helps to simplify content and function of the images within web content to the users. Alternative text can be provided in two ways:

- Within the alt attribute of the < IMG ALT=” ”> Tag.

- Surroundings of the image itself.

It means the alt text is not only method for defining the content and function of an image. This information can also be provided in text near to the image or within the page containing the image.

Moreover, Web Content Accessibility Guidelines (WCAG) 2.0 which covers a wide range of recommendations for making Web content more accessible to a wider range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech disabilities, photosensitivity or combinations of them. It defines following success criteria(W. C. W. W. W. Consortium, 2013) for the ALT TEXT guidelines which are closely related to the prototype:

1. 1.1.1 Non-text Content (Level A): All non-text content that is presented to the user has a text alternative that serves the equivalent purpose.

If a short textual description can serve the same purpose and present the same information as the non-text content: In this case following technical guidelines are recommended for textual description of graphical visual content in the webpage:

- ARIA6: Using aria-label to provide labels for objects
- ARIA10: Using aria-labeled to provide a text alternative for non-text content
- H2: Combining adjacent image and text links for the same resource
- H35: Providing text alternatives on applet elements
- H37: Using ALT attributes on IMG elements
- H53: Using the body of the object element
- PDF1: Applying text alternatives to images with the Alt entry in PDF documents
- SL5: Defining a Focusable Image Class for Silverlight

2. 1.4.5 Images of Text (Level AA): If the technologies being used can achieve the visual presentation, text is used to convey information rather than images of text except.
3. 1.4.9 Images of Text (Level AAA): Images of text are only used for pure decoration or where a presentation of text is essential to the information being conveyed.

Out of these three levels of criteria, the test images on this prototype only correspond to the first 1.1.1 Non-text Content (Level A).

5.2.2 Test Content of Prototype B

Table 5.2 Test Content of Prototype B

Test Image	Source	Alt Text
	www.thebetterindia.com	Students are playing basketball in a tournament organized by project KHEL.
	mundy.assets.d3r.com	Tourists are enjoying Amazon rainforest river ferry expedition by boat.
	wallpaperscraft.ru	Flocks of penguins are jumping in ice snow water in Antarctica.

Test Image	Source	Alt Text
	lh3.googleusercontent.com	Debris of residential area of 2011 Tsunami destruction in Japan.
	userscontent2.emaze.com	Typical wild African elephant family herd at a watering hole.
	pbs.twimg.com	Arial view of hundreds of sheep leaving the stall.
	k37.kn3.net/taringa	A group of wielders working in a bridge construction.

Test Image	**Source**	**Alt Text**
	www.rock-palace.net	Picture of a space rocket taking off from launch station.

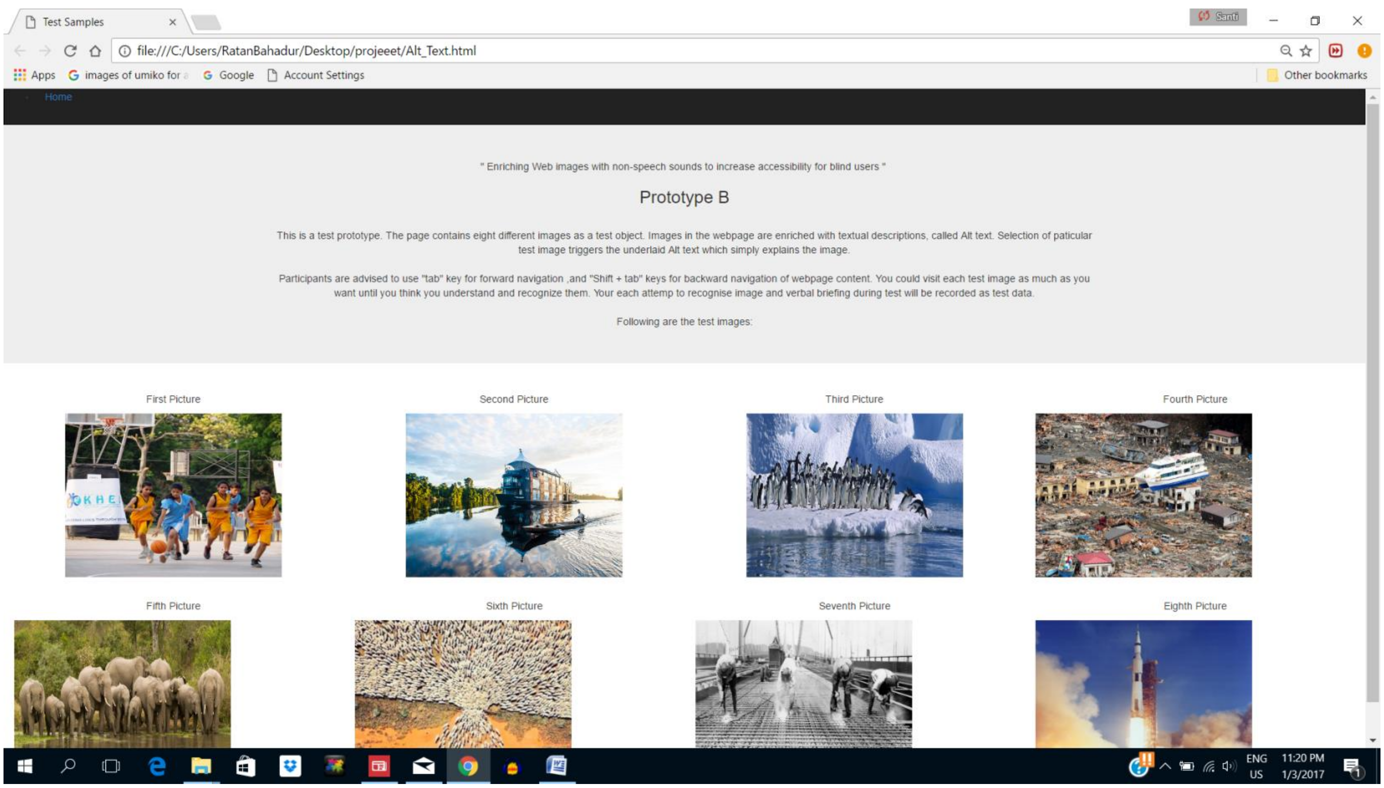

Figure 5.2 Screenshot of prototype B

6. Questionnaire

To quantify and test above mentioned hypotheses of the study, standardized questionnaire sheet was developed as a tool to collect users' reactions, feedback and reflection on test. While designing those questionnaire followings guidelines(Patten, 2016) were taken into consideration:

- Defining wording and order of questions to ensure that each respondent receives the same stimuli
- Defining each question to ensure that all participants handle questions consistently and can answer them undoubtedly.
- Developing response format to enable easy and rapid completion of the questionnaire during the experiment process.

Following consideration has been taken before developing questionnaires for the experiment:

- Questionnaires design should meet the research objectives.
- Use of standard research questionnaires such as standard SUS questionnaire in order to collect complete and accurate information as possible from the participants.
- Questionnaire should be easy and self-explanatory as much as possible. Necessary information and interpretation should be provided.
- Questionnaires should be brief and to the point of research hypothesis and intentions.

Consequently, three questionnaires were designed for the purposed experiments which are as follows:

6.1. Demographic Questionnaire

Demographics are characteristics and social behaviors of a population. This includes race, ethnicity, gender, age, education, profession, occupation, income level, and others. While designing an experiment, the research needs to assess who to test hypothesis and how to breakdown overall experiment response data into meaningful conclusion. Both assessments heavily rely on demographic considerations. This demographics questionnaire was designed to collect basic information about computer and internet literacy, and experience of the participants (Appendix: 16.2.1). Test was scheduled to focus in under-graduated and graduated university students from different disciplines. The questionnaire basically covers the following information about the participants:

- **Age**
- **Gender**
- **Education Literacy**
- **User Experience:**
    - ✓ Computer
    - ✓ WWW
    - ✓ Text to Speech Translator (TTS) Application
- **User Behavior**: Internet Access per hours a day and purpose

6.2. NASA TLX Questionnaire

NASA TLX questionnaire contains a set of six different questions related to the six different workload subscales measurement along with TLX complexity rating scale for individual assessment. Questionnaire covers the workload assessment of a pre-defined task across two prototypes for each participant. It measures the participants perceived subjective workload of performing same task across two different similar platforms. Participants are instructed to provide their workload ratings based on comparison between the techniques of representing test image in test prototypes. Two TLX measurement scales for each cross-ponding prototype were placed vertically right across the workload questions for the easement of comparison, and further colors was used for easy eye guidance while recording their scores in the questionnaire sheet (Appendix: 16.2.2).

6.3. System Usability Questionnaire

This is a standard SUS questionnaire which contains a set of ten usability question regarding each experimental prototype. The scale valued with “0” to “5” points was defined for accessing participants level of agreement or disagreement with usability questions regarding test object. Consequently, two SUS score scale for each test prototype was defined and aligned right across the set of SUS questions. Participants were advised to make comparison between the prototypes while marking their usability ratings for the corresponding prototypes (Appendix: 16.2.3).

7. Indirect Measures

Indirect measure was aimed to collect other aspect of users’ performance data while they are performing experiments such as time spent on each prototype, number of attempts made to each test image while identifying them and status of their effort at recognizing test image (accuracy and error). Along with these, we also noted participants verbal briefing and comments on test image while accessing them through screen reader application. We designed observation sheet for each prototype to collect these indirect measures. Accuracy and error data about each participant at identifying test image in observation sheet were further calibrated with their post-experiment reaction and feedback when prototypes were shown to them (Appendix: 16.3.1).

8. Participants

It is a legal right of research participants that they must be informed about important aspects of the research project, such as the research aims and objectives, potential outcomes and benefits for the participants themselves and others, burdens and inconvenience, handling of personal data, and others. Similarly, participants must be informed that participation is voluntary and based on consent. They also have the right to have information about significant changes in the research that might impact their data and willingness to participate. According to modern norms of research ethics(Diener & Crandall, 1978) and with the Declaration of Helsink

(Association, 2014), following rights of research participants were preserved while conducting experiments (Appendix: 16.3.1):

- Information about why the research is being done
- Information about what expected outcomes of the study
- Information about how private information of participants will be kept safe
- Information about whom to contact in case of any questions or concerns in future
- Has right to questions at any time
- Has right to decide whether to participate without being pressured
- Has right to withdraw from the research at any time without any reason, clarification or penalty
- keep a copy of the consent form

Fourteen participants (N=14) took participation in the research. Most of them had bachelor (N=9) or master degree (N=4) from the university. All have convincing computer and World Wide Web experience for the research experiment which ranged from 1 to 2 years to more than 15 years. The age of the participants ranged from age group 20 to 24 years to 35 to 39 years old. None of them have any visual impairment conditions; all of them were found comfortable with the notion of experiment, pilot test. The majorities of participants (N=12) were unfamiliar with screen reader applications, and consequently they were explained and taught to administrate screen reader application during training session with random web samples. Table below summarize details of demographics information about each participant:

Table 8.1 Demographics Details of Test Participants

Demographic Factors	**Classification (yrs=years)**	**Participants ID**	**Number of Participants**	
	20-24 yrs	00J	1	
	25-29 yrs	00A,00B,00D,00E,00H,00J,00L,00N	7	14
Age	30-34 yrs	00C,00F,00G,00I,00M	5	

Demographic Factors	Classification (yrs=years)	Participants ID	Number of Participants	
	35-39 yrs	00K	1	
Gender	Male	00A,00B,00C,00D,00E,00F,00G,00H,00I ,00K, 00L,00N	12	14
	Female	00J,00M	2	
Education level completed	Bachelor	00A,00B,00D,00E,00H, 00J ,00L,00N,00M	9	14
	Master	00C,00F,00G,00I	4	
	Doctor	00K	1	
Experienced with computer	1-2 yrs	00C	1	14
	3-5 yrs	00D,00E,00K	3	
	6-10 yrs	00A,00B,00I,00J,00N	5	
	11-15 yrs	00F,00G,00H	3	
	More than 15 yrs	00L,00M	2	
Experienced with www	1-2 yrs	00C	1	14
	3-5 yrs	00D,00E,00J	3	
	6-10 yrs	00A,00B,00F,00I,00K,00N	6	
	11-15 yrs	00G,00H,00L,00M	4	

Demographic Factors	**Classification (yrs=years)**	**Participants ID**	**Number of Participants**	
Internet access per hours a Day	3-5 hrs	00A,00B,00C,00D,00E,00J,00N,00M	8	14
	6-10 hrs	00F,00G,00H,00I,00K	5	
	More than 10 hrs	00L	1	
Purpose of using internet	More than single (Entertainment, Communication, Academic, professional, Commercial or Others) purpose	00A,00B,00C,00D,00E,00F,00G,00H,00I ,00J,00K,00L,00N,00M	14	14
Experienced with screen reader application	Yes	00A,00B	2	14
	No	00C,00D,00E,00F,00G,00H,00I,00J,00K, 00L,00N,00M	12	

9. Test Procedure

In absence of real screen reader user, we conducted a pilot test with sighted users. Pilot test is helpful in testing a hypothesis within a limiting resource such as cost and time. It can be adapted to perfect the larger experiment as well. First, participants are introduced and trained with Text to Speech (TTS) screen reader application to develop and gain some level of experience with non-visual access contexts before conducting actual test. Participants had to access each test prototype entirely based on screen reader and keyboard navigation. Blind perspective was created by disabling pointing inputs (such as touch pad and mouse) and displays of the test device. Screen of test device was completely covered with sheet of black paper and consequently participants were confirmed about complete absence of visual sight.

Participants were briefed one week in advance about the conditions of the test when they received the consent form. The experiment was scheduled approximately for one and half hours for each participant. On test day, it began with briefing experimental procedure and measurement scales used for collecting participant's response. Then, selected screen reader application was introduced to the participants for the test. They all went through a 30 minutes training phase to develop some level of experience with non-visual access of content. Initially participants were familiarized with the NVDA (Non-Visual Desktop Access)(Access, 2009) screen-reader application. They were taught how to administrate NVDA with eSpeak synthesizer[7] and Ava, US-English Premium High vocalizer[8] for webpage navigation[9]. Consequently, they were introduced to audemes with samples taken from an online audemes dictionary.

During the test, users were advised to access each test image as many times they wanted until they thought they understood it. They were encouraged to speak out what they thought about the test images while accessing them. The number of attempts each participant made when accessing the images along with their comments during the test was recorded in an observation

[7] http://espeak.sourceforge.net
[8] https://vocalizer-nvda.com/docs/en/userguide.html
[9] https://www.nvaccess.org/files/nvda/documentation/userGuide.html#toc29

sheet. After participants went through both prototypes, they completed a paper version of the NASA Task Load Index (TLX) to assess the workload they perceived in identifying test image. Two TLX measurement scales for each prototype were placed vertically right across the workload questions and participants were instructed to provide their workload ratings based on comparison. Finally, A System Usability Scale with ten question for each prototype was administered using 5-point Likert rating scale. Participants were advised to make comparison between the prototypes while rating the usability.

9.1 NASA TLX Scale

NASA TLX wording scales were used for assessing participants' perceived workload experiences across two different prototypes. NASA TLX scale is extremely useful, but its applicability suffers from the subjective ability to interpret them individually. For example, people might feel that mental or temporal demands are the essential aspects of workload regardless their extra effort and achieved performance. Others might feel that if they performed well then workload must have been low and if they performed badly it must have been high. Similarly, they might feel that effort or frustration levels are the most important factors in workload and so on.

Moreover, the difference in ability and skill among participants is another major factor. A skilled and smart user might feel task less complex than the averaged one user. Thus, NASA TLX scale and ratings is possibly most difficult and frequently misunderstand scales. The workload complexity scale is very hard to break down and administer from individual participants' perspective. Therefore, Bedford workload rating scale (Hart & Wickens, 1990) was used for further clarification and easement of the workload scale breakdown. The Bedford scale information was further modified and used only as an informative tool to explain the load of NASA TLX scale interval to the participants. This was a training sample for breaking down temporal demands on the TLX scales of 0 to 100.

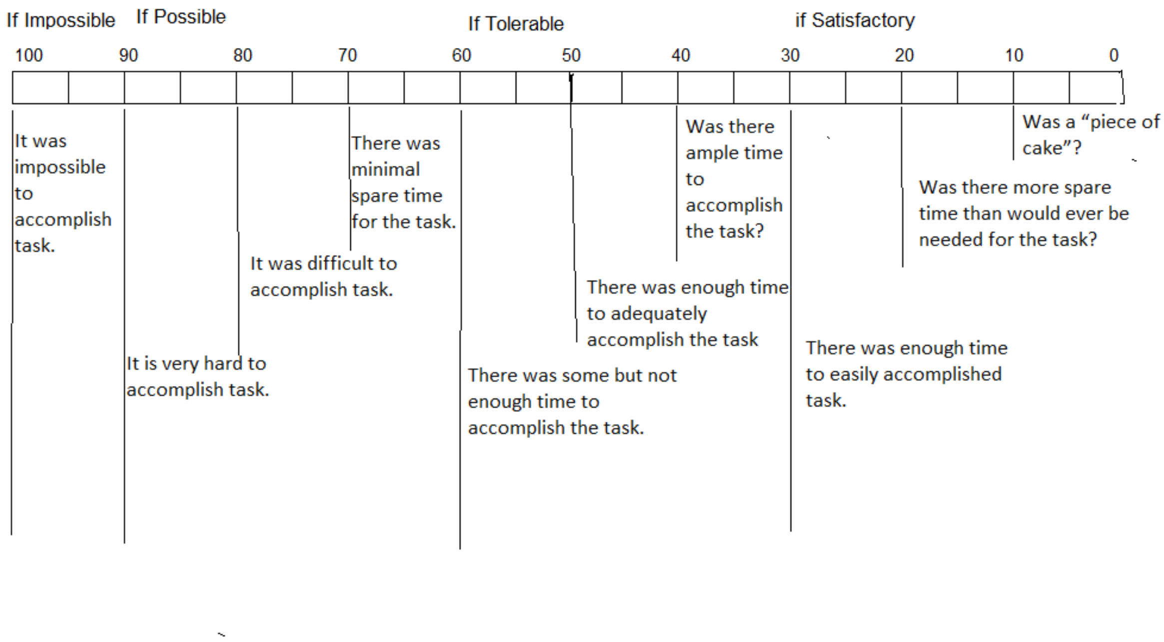

Figure 9.1 Breaking down NASA TLX scale for temporal demands

In the Raw TLX, rating scale is a horizontal line divided from 0 and 100 into intervals of 5 points. When participants check in a particular box, response is calculated on the basis of intervals value from the line. Following guidance was provided for scoring their responses to the participants.

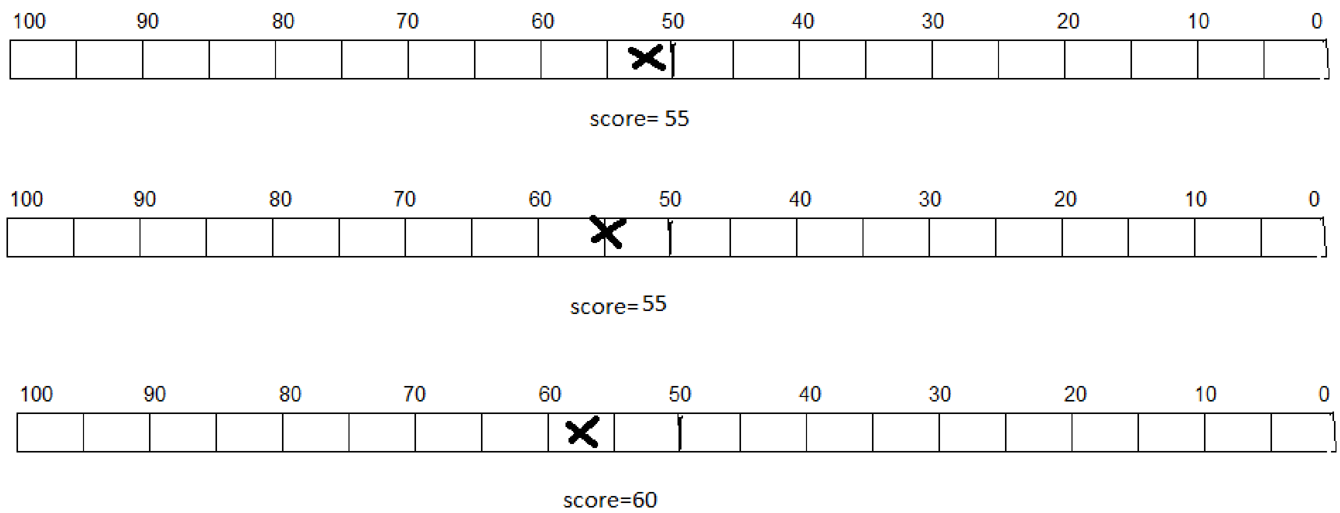

Figure 9.2 TLX Score validation samples

9.2 System usability Scale

SUS scale measures the positive and negative aspect of the prototype in a set of 10 questions. Questions are organized in odd and even number in relation to positive and negative remarks of test object. Negative questions are shades in questionnaire for more clarity to the participants. The scores on these two types of question act differently in defining the usability of the prototypes. For example, higher level of agreement on even number questions (negative one) leads to less usable the prototype, and the reverse was happened for the odd number (positive one). However, SUS was only designed to measure usability of the system it also provides a global measure of system satisfaction, and sub-scales of usability and learn-ability. For example, question number 4 and 10 in SUS questionnaire sets provides the learn-ability dimension and question number 8 provide the usable dimension. This means participants reflection scores can be tracked and reported on both subscales as well as whole in relation to usability.

Scoring usability factors, how to select right number in usability scales to express right level agreement and disagreement to the usability questions are another important issue of sus questionnaires. Generally, usability scales are divided into 7 points or 5 points to represent level of agreement and disagreement corresponding to sus questions which could be explained by using following adjective as follows:

Disagree Agree

1	2	3	4	5	6	7
Worst Imaginable	Awful	Poor	Ok	Good	Excellent	Best Imaginable

Figure 9.3 Adjective ratings breakdown for 7-ponts SUS Likert scale

Disagree Agree

1	2	3	4	5
Worst	Poor	Ok	Good	Excellent

Figure 9.4 Adjective ratings breakdown for 5-ponts SUS Likert scale

In this case, 7 points were found to be more computable for the participants as compared to 5 points SUS scales, especially for carrying out the comparison in between the prototypes while registering their response to the questionnaire. Thus, whole level of disagreement and agreement in sus runs in 5 points Likert scales, and selection of middle point value 3 will be interpret as neutral response to the questions. In this way, participants could reflect their response to the usability question in ratio to quantifiable number in the scale.

9.3 Text to Speech Application: NV Access

NVDA (Non-Visual Desktop Access) is an open source text to speech (TTS) screen reader application(Access, 2009). It assists blind and visually impaired people to access computers. It uses computerized voice, supports all accessibility criteria for disable people. It also assists

Braille display device. The main purpose of NVDA is to provide education and employment for blind people. It enables blind people to access social networking, online shopping, banking and news. Normally, all screen readers are expensive, making them commercial and unaffordable for most people whereas NVDA is free at cost open educational research project. It is available in 43 different languages. It compatible on both 32-bit and 64-bit editions of Microsoft Windows XP, and does not requires any additional hardware and software beyond those of the operating system.

NVDA has following features(Access, 2015; Preece, 2016):

- It supports all available web browsers, email, internet chatting, and office programs including Word and Excel.
- It has built-in speech synthesizer for more than 43 languages.
- It supports Braille and refreshable Braille displays devices.
- It is easy to use and administer.
- It supports common accessibility interfaces including Java Access Bridge.
- It supports Windows command prompt and console applications.

Reasons for selecting NVDA as a screen reader application for the experiments:

- It is available free at cost
- It is not commercial that makes it available and accessible everywhere despite trade and geopolitical sanctioned.
- It has extremely light and portable computer resource for the blind people

9.4 Training

Training session starts with a brief discussion about semiotics structure of audemes and its referential cue; how an audemes could represent things. Participants were also explained about mode of listening and their impact on audemes meaning derivation. Consequently, they are guided to samples of audemes for a short training before starting actual test. Secondly, participants were familiarized with NVDA screen reader application. They were taught to administer and use NVDA access for web surfing and object navigation. All over, each participant spent approximately 30 minutes in training sessions.

9.5 Task and Condition

However, NASA TLX tool does not measure the nature of task, it only measures subjective perception of difficulties while performing task. Thus, we defined a clear task to assess participants' perception of workload through NASA TLX. Task for the experiment was:

Task:" Identifying test images accurately in webpage while accessing non-visually."

9.6 Study Setup

Screen of laptop was completely covered with black paper sheet, and consequently participants were conformed. Volume and speed of screen reader speed was adjusted according to their preferences. The "tab" and "shift" keys on the keyboard were instructed to navigate the test prototypes content. When the user navigates test image corresponding audemes audio file starts play. Consequently, audemes for the test image paused when user move-out navigation to next elements of the page, and began to play from the beginning while navigating again. While in case of alt text, screen reader simple reads the textual description. Users were also advised to access each test object as much as they want until they think they understood them. They are encouraged to speak what they think about the test object while accessing them. Their each attempt to access test image and verbal briefing about the test image during test was recorded in observation sheet for further data analysis.

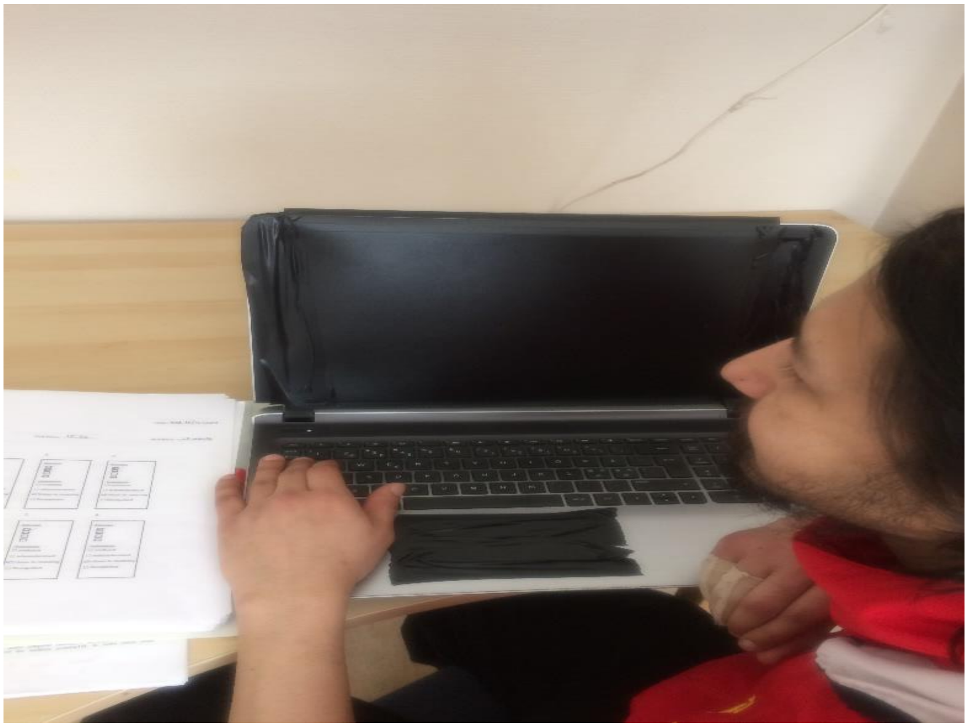

Figure 9.5 Participant accessing test prototype with NVDA

9.7 Components of study setups

HP Laptop

System Manufacturer: Hewlett Packard

System Type: x64-based PC

Operating System

OS Name: Microsoft Windows 10 Home

Version: 10.0.14393 Build 14393

Screen Reader

NVDA, Version: 2016.4

Synthesizer: eSpeaker

Vocalizer: Ava “en_us_English” Premium High

Web Browser

Name: Google Chrome

Version: 55.0.2883.87 m (64-bit)

10. Data collection

Right after the experiment, participants were guided to the paper and pencil version of questionnaires for collecting their response about experiment. They first completed SUS questionnaire, then NASA TLX. Later, whole data is reorganized in the relative scoring sheet for further calculations which are as follows:

10.1 NASA RTLX

The file contains raw TLX workload data, with completed set of ratings provided by individual participants. The table 4 has two headers for NASA TLX workload factors and participants ID. The whole data is organized under six different TLX subscale score corresponding to the fourteen participants into 6 by 14 matrix table cells. NASA Subscales value range from 0 to 100 with interval value of 5.

As un-weighted workload, RTLX score is calculated by averaging subscales values. RTLX workload value is then rounded off or approximated into nearest TLX workload scale ratings. Similarly, mean value for individual subscale among 14 participants is calculated as well. Consequently, mean value of individual workload factors among 14 participants were also approximated into nearest TLX scale ratings.

Table 10.1 NASA RAW-TLX score sheet of prototype B: ALT TEXT

NASA TLX Factors	Participants ID														Mean score	Approx. score
	00A	00B	00C	00D	00E	00F	00G	00H	00I	00J	00K	00L	00M	00N		
Mental Demand	60	50	60	60	60	90	70	55	55	65	60	65	55	60	61.7	60
Physical Demand	10	15	45	50	70	50	50	45	25	35	25	25	50	30	37.5	40
Temporal Demand	50	85	60	65	65	50	70	65	70	50	65	65	65	70	63.9	65
Performance	15	50	70	70	60	40	80	55	40	55	65	55	55	70	55.7	55
Effort	50	40	70	55	65	85	80	55	50	65	65	60	55	55	60.7	60
Frustration	50	45	70	70	55	85	100	55	35	55	65	65	75	70	63.9	65
RTLX Score	39.5	47.5	62.5	61.6	62.5	66.6	75	55	45.8	54.1	57.5	55.8	60.8	59.1		
Approx. RTLX	40	50	65	60	65	65	75	55	45	55	60	55	60	60		

Table 10.2 NASA RAW-TLX score sheet of prototype A: AUDEMES

NASA TLX Factors	Participants ID														Mean score	Approx. score
	00A	00B	00C	00D	00E	00F	00G	00H	00I	00J	00K	00L	00M	00N		
Mental Demand	40	25	55	40	40	65	25	35	45	40	35	55	25	25	39.28	40
Physical Demand	10	15	40	50	25	50	50	50	25	35	30	25	50	30	34.65	35
Temporal Demand	50	35	45	35	30	50	25	30	35	45	45	45	30	40	38.5	40
Performance	15	25	25	30	30	65	20	35	25	35	30	55	30	40	32.5	35
Effort	30	20	40	35	55	30	20	35	45	40	40	50	35	30	34.64	35
Frustration	25	20	25	25	25	30	15	35	25	45	30	30	25	25	27.14	30
RTLX score	28.5	23.5	38.3	35.8	34.1	48.3	26	35	33.3	41.6	33.3	43.3	32.5	31.6		
Approx. RTLX	30	25	40	35	35	50	25	35	35	40	35	45	35	30		

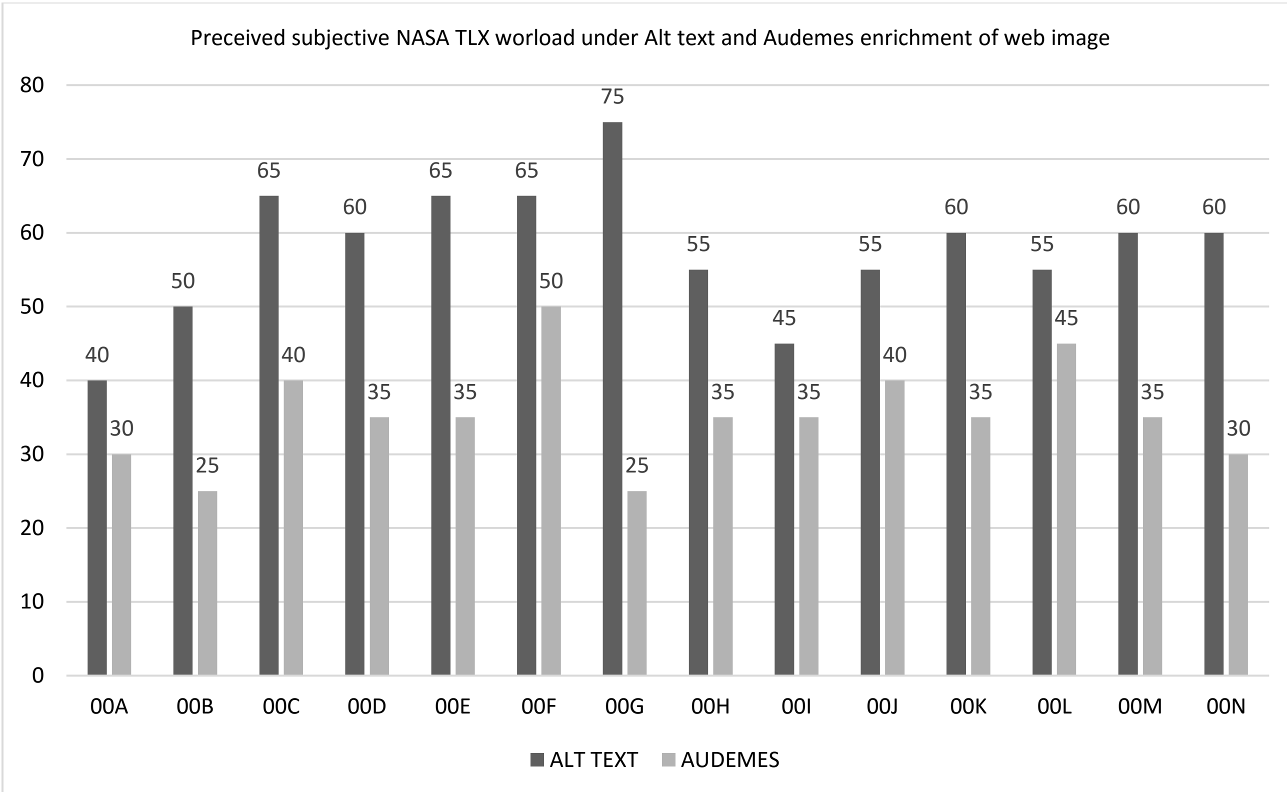

Figure 10.1 NASA TLX workload scores of 14 participants under Alt text and Audemes enrichment of web image

10.2 System Usability Scale

After testing prototypes, participants were asked to record their immediate response on SUS questionnaire without taking too much time on individual SUS questions. They were advised to mark at the middle of scale in case of confusion about the question. Finally, questionnaire sheet was checked for complete attempt of all questions by participants.

Score of individual question in questionnaire will range from 0 to 4. First, questions were identified with odd and even numbers. For odd number questions: 1,3,5,7, and 9; scores were calculated by subtracting one from the scale value whereas for even numbers: 2,4,6,8 and 10; scores were calculated by subtracting scale value from 5. Finally, sum of the scores was multiplied by 2.5 to obtain the overall value of SUS which have a range of 0 to 100. Data sheet contains two columns: one for SUS question and another for the Participants ID.

Table 10.3 System usability scale score sheet of prototype B: ALT TEXT

SUS Question	Participants ID													
	00A	00B	00C	00D	00E	00F	00G	00H	00I	00J	00K	00L	00M	00N
1.(ODD)	1	3	1	1	1	1	0	2	2	1	1	3	2	1
2.(EVEN)	3	2	2	0	0	0	2	1	2	3	1	4	2	2
3.(ODD)	2	2	1	2	1	0	2	3	2	1	1	1	2	2
4.(EVEN)	0	1	3	3	1	3	0	4	2	4	1	1	2	3
5.(ODD)	2	1	2	2	2	2	3	3	2	0	2	3	2	2
6.(EVEN)	2	2	2	0	2	2	1	0	2	4	4	3	2	2
7.(ODD)	1	2	2	2	0	0	3	3	2	1	1	1	1	4
8.(EVEN)	1	1	0	0	3	3	3	3	2	2	1	1	1	1
9.(ODD)	0	2	2	2	0	2	4	4	2	1	3	1	2	2
10.(EVEN)	1	3	1	1	0	2	0	4	2	2	3	1	1	2
Total	13	19	16	13	10	15	18	27	20	19	18	19	17	21
SUS score	32.5	47.5	40	32.5	25	37.5	45	62.5	50	47.5	45	47.5	42.5	52.5

Table 10.4 System usability scale score sheet of prototype A: AUDEMES

SUS Question	Participants ID													
	00A	00B	00C	00D	00E	00F	00G	00H	00I	00J	00K	00L	00M	00N
1.(ODD)	2	3	3	4	4	4	4	3	3	3	3	3	4	4
2.(EVEN)	2	3	3	4	4	4	4	3	2	2	1	3	3	4
3.(ODD)	3	3	4	4	3	3	4	3	3	4	3	3	2	4
4.(EVEN)	3	2	3	3	4	4	1	4	1	2	3	1	2	4
5.(ODD)	2	1	2	2	4	2	3	3	2	3	2	3	2	2
6.(EVEN)	2	2	2	1	4	2	3	2	2	3	4	3	2	2
7.(ODD)	3	3	4	4	4	3	3	4	3	3	3	3	4	4
8.(EVEN)	2	3	2	4	4	4	4	3	2	2	3	3	4	4
9.(ODD)	1	3	2	4	4	4	4	4	3	3	3	1	4	4
10.(EVEN)	1	1	1	1	3	2	1	4	1	1	3	0	4	4
TOTAL	21	24	26	31	38	32	31	32	23	25	28	23	31	36
SUS score	52.5	60	65	77.5	95	80	77.5	80	57.5	62.5	70	57.5	77.5	90

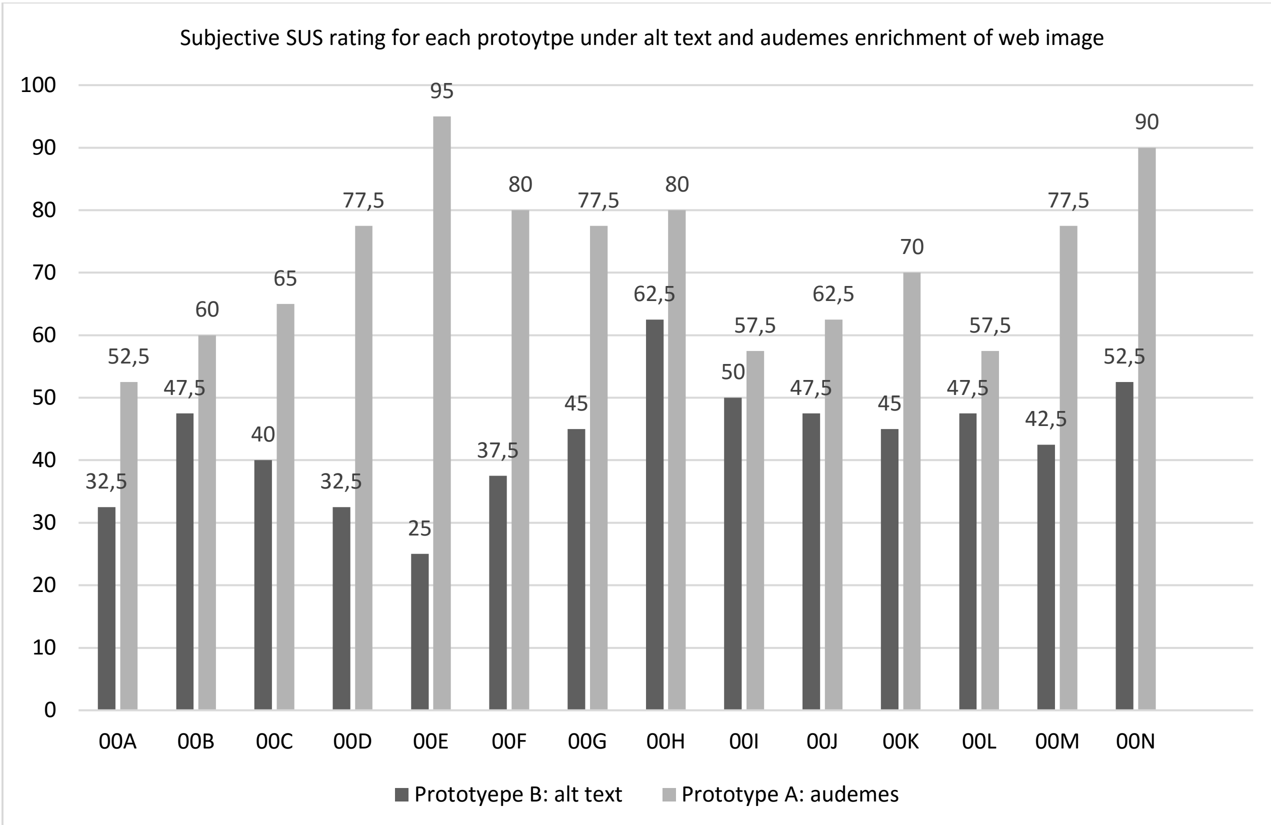

Figure 10.2 SUS rating provided by 14 participants for each prototype under alt text and audemes enrichment of web image

10.3 Observation sheet

While participants were performing test, we recorded other aspects of their performance data in observation sheet along with their comments. The observation sheet further organizes these participants' data into following dimensions:

10.3.1. Accuracy and Error

During test, participants were suggested to in-vision the test image while listening them, and explain what they think about the image after listening. Consequently, participants provided their explanation and comments on each test image while accessing them through NVDA, and accordingly we further categorized their performance at identifying test image into four categories depending upon their explanations and comments. The categories are: Recognized, Closer to meaning, Confused and Miss understood.

- Recognized: Participants explained image accurately
- Closer to meanings: Participants explanation is close to image but not accurately same as image
- Confused: Participants explained the image with confusion and said they were not sure about image.
- Miss understood: Participants explanation is completely different than image

These performance data were further calibrated with participants' responses (where they said what they had imagined about the image during test) when test images were shown to them after the test. Finally, these data were recategorized into two group; Accuracy and Error, for analysis.

Table 10.5 Participants' accuracy and error scores out 8 test image under alt text and audemes

Participants	Accuracy				Error			
	Recognized		Closer to meaning		Confused		Miss understood	
	Alt Text	Audemes	Alt Text	Audemes	Alt Text	Audemes	Alt Text	Audemes
00A	3	2	4	4	1	2	0	0
00B	4	2	2	4	2	2	0	0
00C	3	2	4	4	1	2	0	0
00D	3	3	3	4	2	1	0	0
00E	2	3	3	4	3	1	0	0
00F	4	4	2	4	2	0	0	0
00G	3	3	3	3	2	2	0	0
00H	2	3	2	4	4	1	0	0
00I	3	4	3	4	2	0	0	0
00J	3	4	4	4	1	0	0	0
00K	3	3	3	4	2	1	0	0
00L	3	3	3	4	2	1	0	0
00M	3	3	3	5	2	0	0	0
00N	3	3	3	4	2	1	0	0

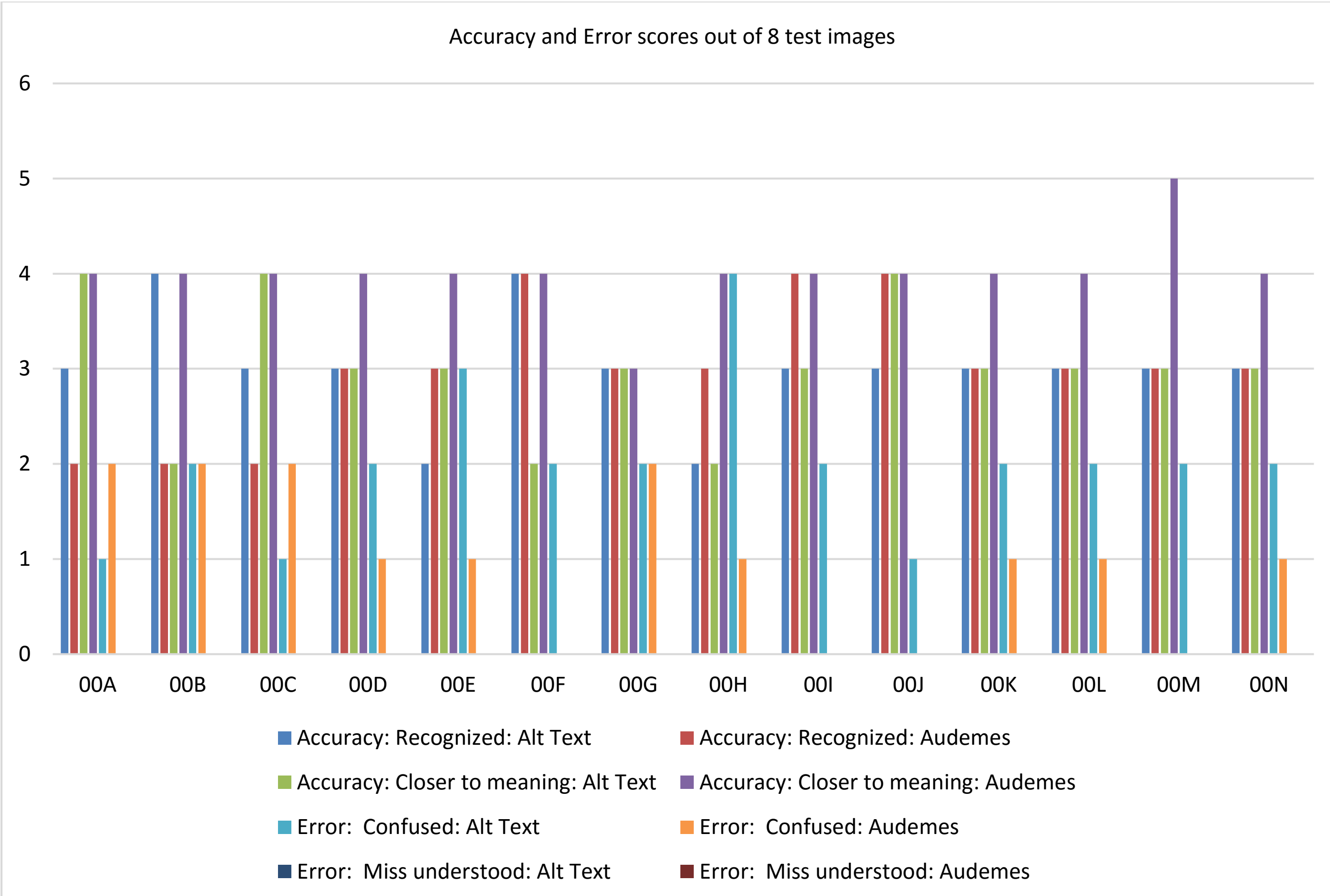

Figure 10.3 Participants' accuracy and error scores out 8 test image under alt text and audemes

10.3.2. Attempts to test images

While performing test, participants were suggested to access each test image as much as they want until they think they understood them. Consequently, we recorded participants' numbers of attempt on each test image on observation sheet. And, we found following cumulative attempt results in each prototype:

Table 10.6 Participants cumulative attempts to test image while identifying them with alt text and audemes enrichment

Participants	Cumulative attempts to all test images	
	Alt	Audemes
00A	20	15
00B	18	16
00C	20	15
00D	22	20
00E	21	20
00F	22	22
00G	19	16
00H	18	19
00I	18	23
00J	22	22
00K	18	20
00L	17	16
00M	18	22
00N	18	16

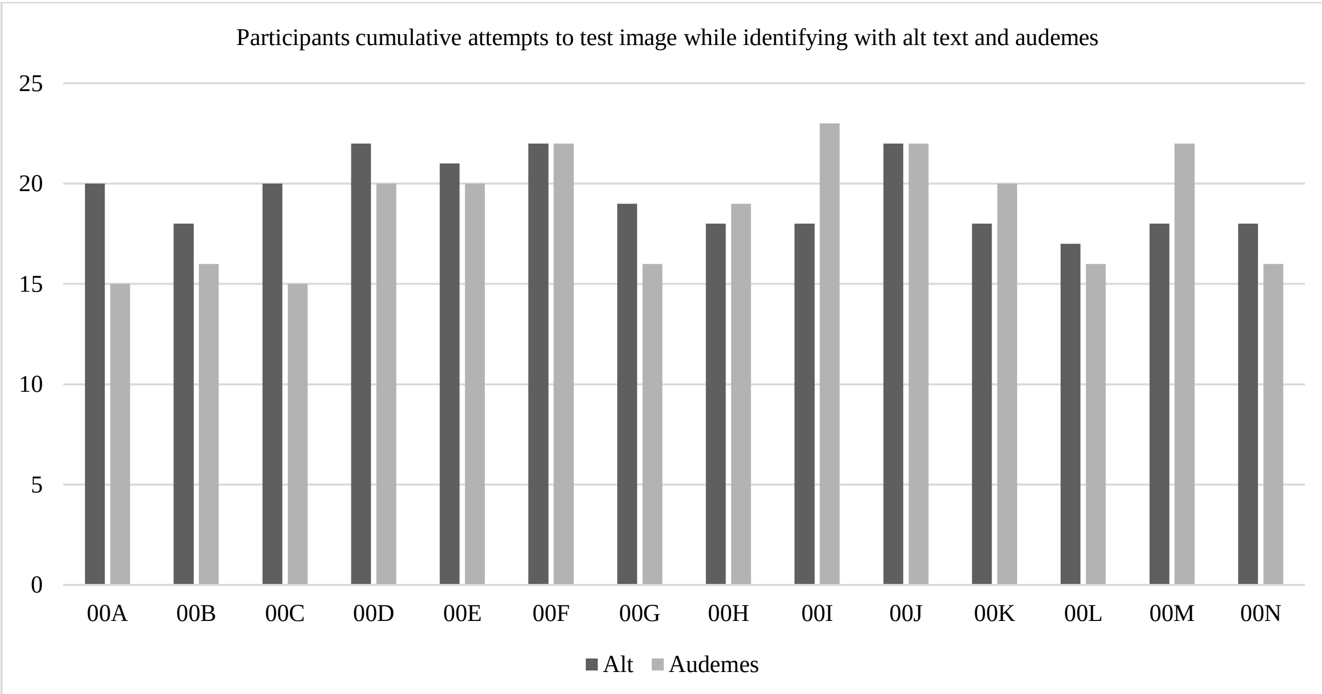

Figure 10.4 Participants cumulative attempts to test image while identifying with alt text and audemes

10.3.3. Time spent on prototypes

While participants were taking test, we also recorded total time duration they spent on each prototype. Table below shows the time spent on each prototype by each participant.

Table 10.7 Time spent on each prototype with alt text and audemes enrichment of web image

Participants	Time spent on each prototypes(minutes)	
	Alt	Audemes
00A	14	15
00B	13	14
00C	14	16
00D	15	16
00E	12	14
00F	15	21
00G	14	15
00H	11	13
00I	13	18
00J	15	19
00K	12	13
00L	14	16
00M	15	19
00N	13	14

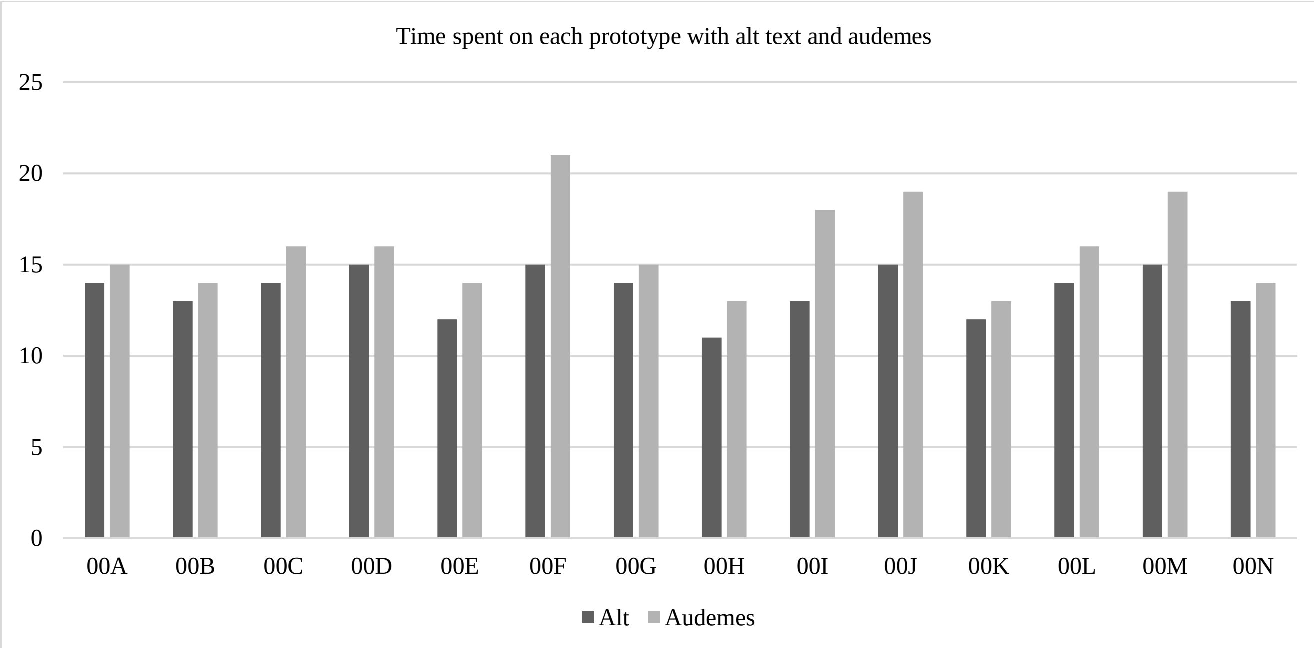

Figure 10.5 Time spent on each prototype with alt text and audemes enrichment of web image by 14 participants

11. Analysis and results

The examination of the data collected was initially conducted by ensuring its validity and suitability for statistical analysis. In all cases, we report means and standard deviations, while for the TLX data, and accuracy and error data, we also performed and report a paired t-test analysis to ensure significance. The TLX data analysis followed the approach in this study (Hart & Wickens, 1990).

11.1. Utilization of TLX scale

NASA TLX workload complexity scale has been a constant challenge to the researcher in assessing subjective workload of participants. Perception of subjective workload and utilization of TLX scale ratings for a task might be different among participants which often mislead the research findings. User experienced with task environment and ability to cope with the task has great influence in workload assessment. These factors were taken care in experiment procedure sessions before conducting test.

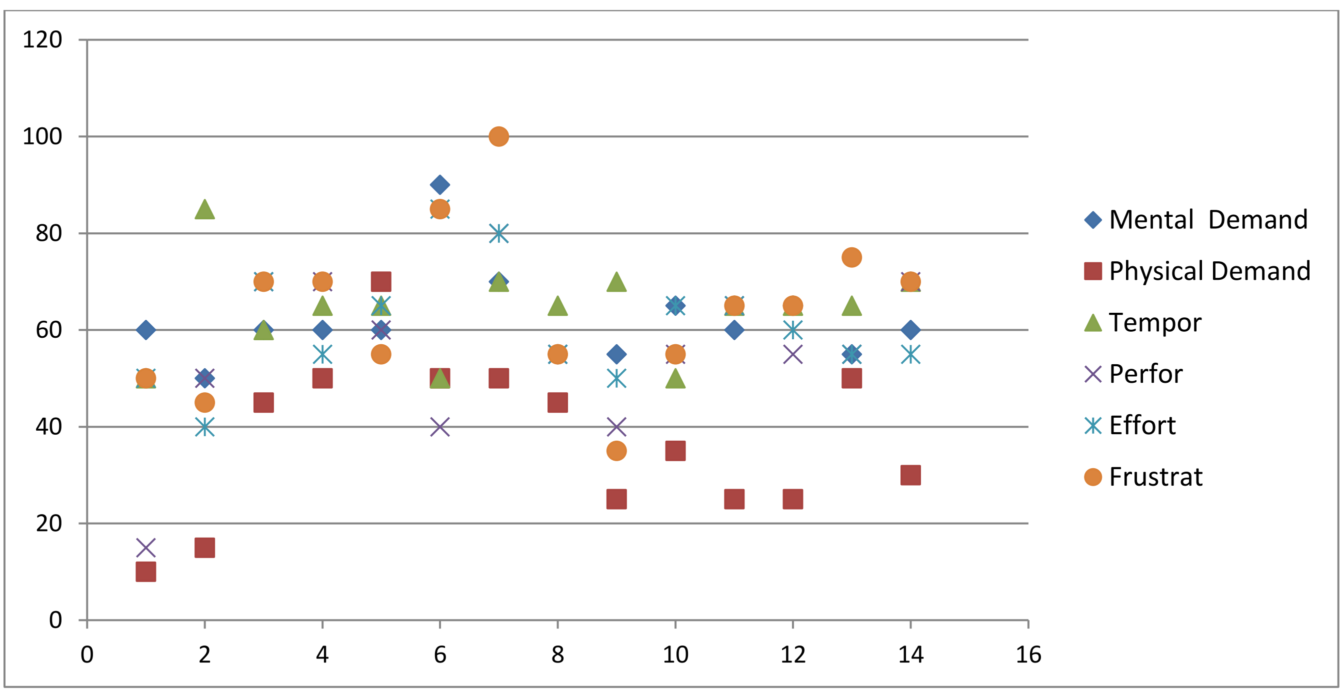

Figure 11.1 TLX Workload scale utilization by participants for alt text

After plotting scores in chart above we found that the participants used TLX scale value 10 to highest 100 for rating workload experienced while accessing web image with alt text. Higher the scale value used for rating workload indicates higher amount of workload experienced for TLX workload factors.

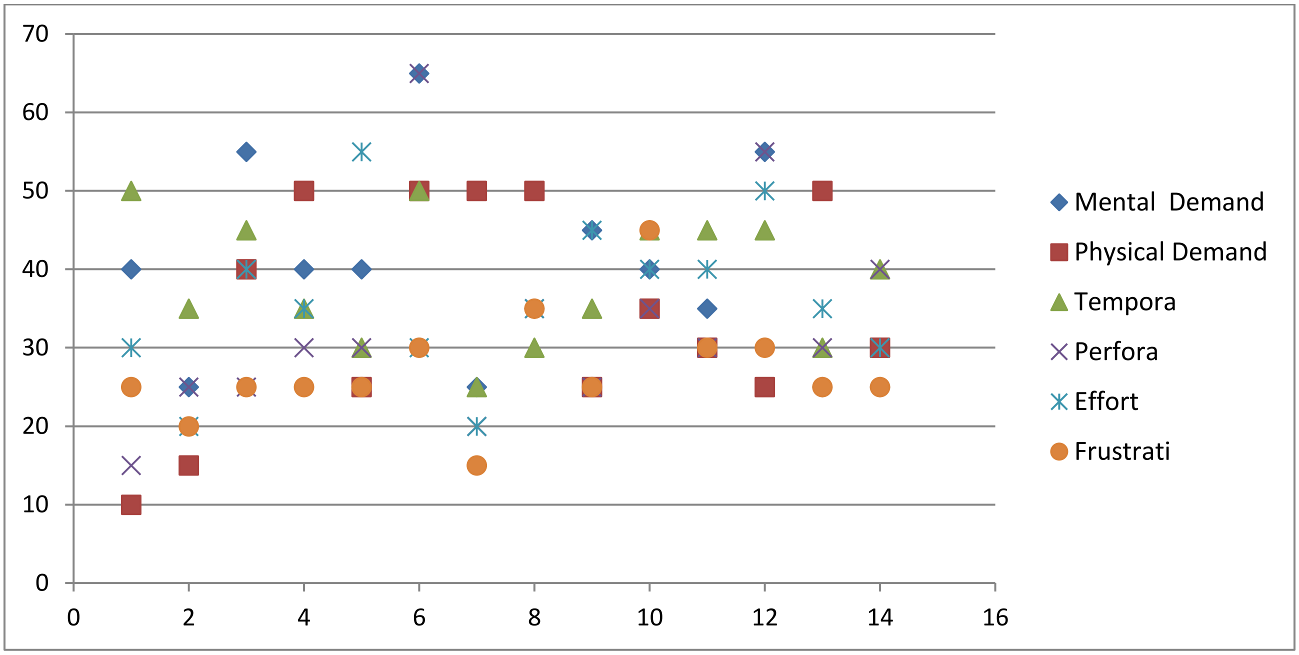

Figure 11.2 TLX Workload scale utilization by participants for audemes

Whereas, participants used TLX scale value 10 to highest 65 for rating workload while accessing web image enriched with audemes. Lower the TLX scale value used indicates lower the amount of workload experienced for TLX workload factors.

11.2 NASA TLX Workload

The overall findings indicate that participants experienced significantly lesser workload in identifying test images with audemes (M=35.35, SD=18.46) compared to alt text (M=57.85, SD=8.92), $t_{(13)}=7.87$, $p<0.05$.

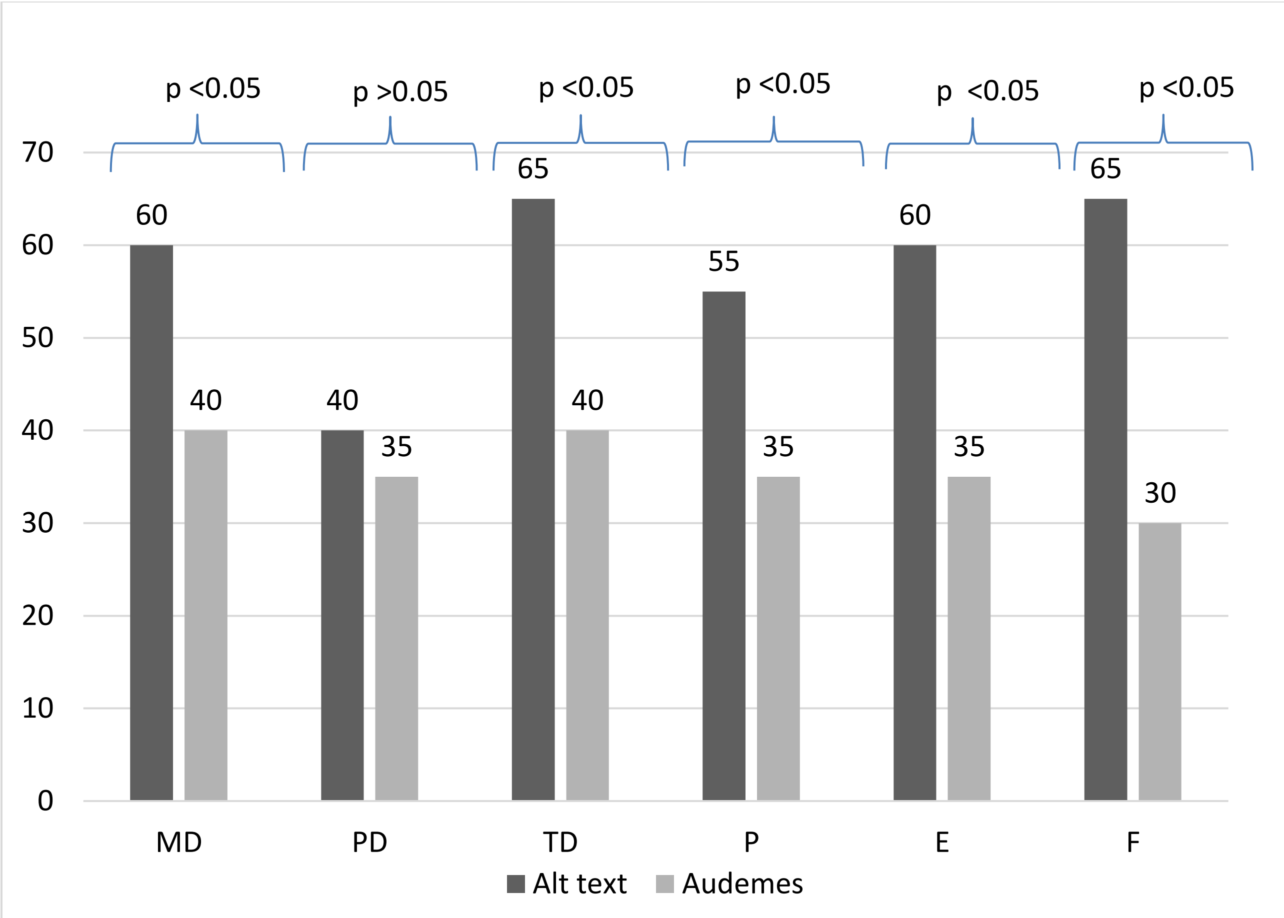

Figure 11.3 Paired t-test for each NASA-TLX subscales' workload (MD=mental demand, PD=physical demand, TD= temporal demand, P=performance, E=effort, F=frustration).

Looking at individual dimensions of the TLX, we saw significant decrease in mental demand required to recognize web images with audemes (M=39.28, SD=12.53) over alt text (M=61.78, SD=8.92), $t_{(13)}$=8.14, p<0.05. Similarly, there was a significant reduction in temporal demand with audemes (M=38.75, SD=8.18) over alt text (M=63.92, SD=16.61), $t_{(13)}$=5.97, p<0.05. In terms of physical demand, however, no significant difference was seen when comparing audemes (M=34.64, SD=13.93) with alt text (M=37.5, SD=9.53), $t_{(13)}$=0.86, p>0.05.

The results show that the task performance significantly increased when participants used the prototype with audemes (M=32.85, SD=13.25) compared to alt text (M=55.71, SD=9.44), $t_{(13)}$=4.03, p<0.05. Linked to this, the results also show that there was significantly less effort required to recognize images with audemes (M=36.07, SD=10.28) compared to alt text

(M=60.71, SD=16.39), $t_{(13)}$=5.94, p<0.05. Moreover, there was significant reduction in frustration level while accessing web images with audemes (M=27.14, SD=6.99) compared to alt text (M=63.92, SD=12.06), $t_{(13)}$=6.96, p<0.05. Figure 21 depicts these values.These results clearly indicate overall better results of the condition when audemes were used compared to alt text. Participants commented that the synthetic speech from the NVDA screen-reader proved stressful to concentrate and caused them to miss the alt text descriptions. On the other hand, the condition with audemes was found to be more pleasant, and did not make participants feel distracted and irritated even when visiting an image multiple time.

Table 11.1 Mean and standard deviation of NASA TLX workload score

NASA TLX workload factors	**Audemes**		**Alt text**	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
Mental demand	39.28	12.53	61.78	8.92
Physical demand	34.64	13.93	37.5	9.53
Temporal demand	38.75	8.18	63.92	16.61
Performance	32.85	13.25	55.71	9.44
Efficiency	36.07	10.18	60.71	16.39
Frustration	27.14	6.99	63.92	12.09

Additionally, looking at mean scores and standard deviation (Table 11.1), participants shows lowest 6.99 units to highest 13.9 units of subjective TLX workload perception difference in identifying web image with audemes. Subjective TLX workload perception difference increased with alt text from lowest 8.92 units to highest 16.39 units in identifying web image. Less subjective perception difference in audemes recognition among participants indicates high perception rate of audemes as a language (M. A. Ferati, 2012c, pp. 95-99). While larger subjective TLX workload perception difference in identifying image with alt text in compared to audemes indicates speech comprehension of alt text is more subjective than audemes.

The highest subjective difference in workload between audemes and alt text came under frustration factors where participants rated audemes less irritational and annoying than alt text by less than half unit of TLX. This also justifies participants' previous comments on alt text about being unclear and difficult to concentrate and audemes about being more musical and entertaining. This might have left huge impact in participants' perception of overall workload which need to be study further.

11.3 System Usability

In order to measure the usability level of the prototypes, we collected a System Usability Scale data and compared them. The result show that all participants rated the prototype with audemes to be more usable compared to the prototype with alt text (Figure 23). This goes in line with the results from the previous section showing that participants perceived less subjective workload in identifying test images using audemes, which influenced their perception of usability.

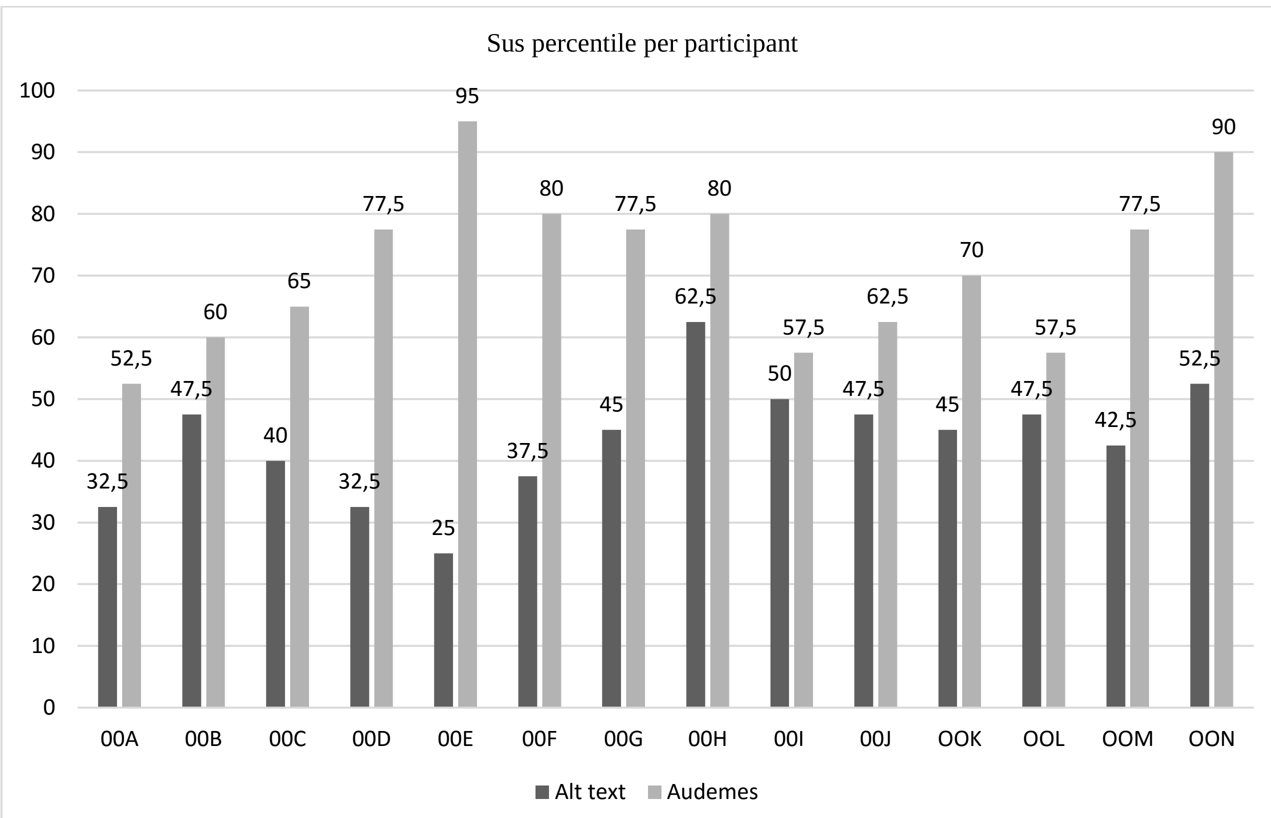

Figure 11.4 SUS Percentiles per participant

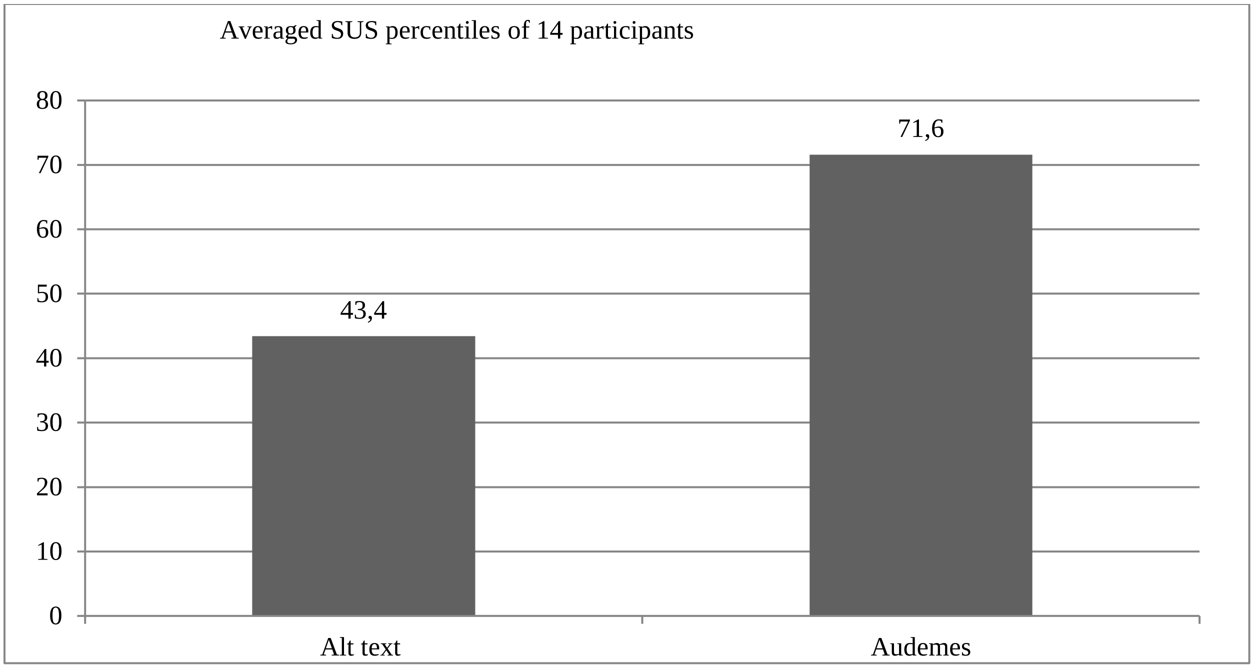

Figure 11.5 Mean SUS percentiles for each prototype provided by 14 participants

Specifically, the usability ratings given by all participants shows that the prototype with audemes scored 71.6 percentiles in comparison with the prototype with alt text, which scored only 43.4 percentiles (Figure 24). Following the ranking analysis offered in this study (Bangor, Kortum, & Miller, 2008), Table 11.2 lists the prototype A as acceptable compared to B.

Table 11.2 Grade scale, Acceptability and Adjective ranges of the prototypes

Test prototype	Acceptability ranges	Grade scale	Adjective ranges
Prototype A: Audemes	Acceptable	"F"	Excellent
Prototype B: Alt text	Not acceptable	"C"	Okay

This difference between the two prototypes according to participants' comments was that audemes were perceived as more entertaining and engaging than speech. Their additional comment, however, was that audemes were too short. They suggested that audemes should be longer to be more informational and easy to understand.

11.4 Results in Terms of Accuracy and Error

Part of the experiment was to measure participants' level of accuracy in terms of image recognition. For each image, we have recorded the level of accuracy using categories "Recognized" and "Closer to Meaning", and errors made using categories "Confused" and "Misunderstood". These responses were also further adjusted with participants' feedback when images were shown to them. We also recorded the number of attempts it took participants to identify each image and the time spent on each prototype.

From figure 11.6, the results show that participants' overall accuracy at identifying the images was significantly higher with audemes (M=7, SD=0.78) compared to alt text (M=6, SD=0.81), $t_{(13)}$=4.03, p<0.05. More closely, in both conditions participants recognized almost half of all images with the same accuracy, namely audemes (M=3, SD=0.60) compared to alt text (M=3, SD=0.55), $t_{(13)}$=0, p>0.05. However, using audemes (M=4, SD=0.74), participants were closer to meaning than using alt text (M=3, SD=0.67), $t_{(13)}$=4.76, p<0.05. In terms of error, in both conditions participants showed no signs of misunderstanding any of the images, however, they were significantly less confused when using audemes (M=1, SD=0.82) compared to alt text (M=2, SD=0.92), $t_{(13)}$=2.80, p<0.05. These values are depicted in the Figure 25.

Participants commented that the confusion with alt text was caused by the unclear speech generated by TTS and their inability to concentrate to listen to it continuously. Similarly, they thought audemes were more closely associated with the images. This indicates that compared to alt text, audemes improve image recognition as well as reduce errors.

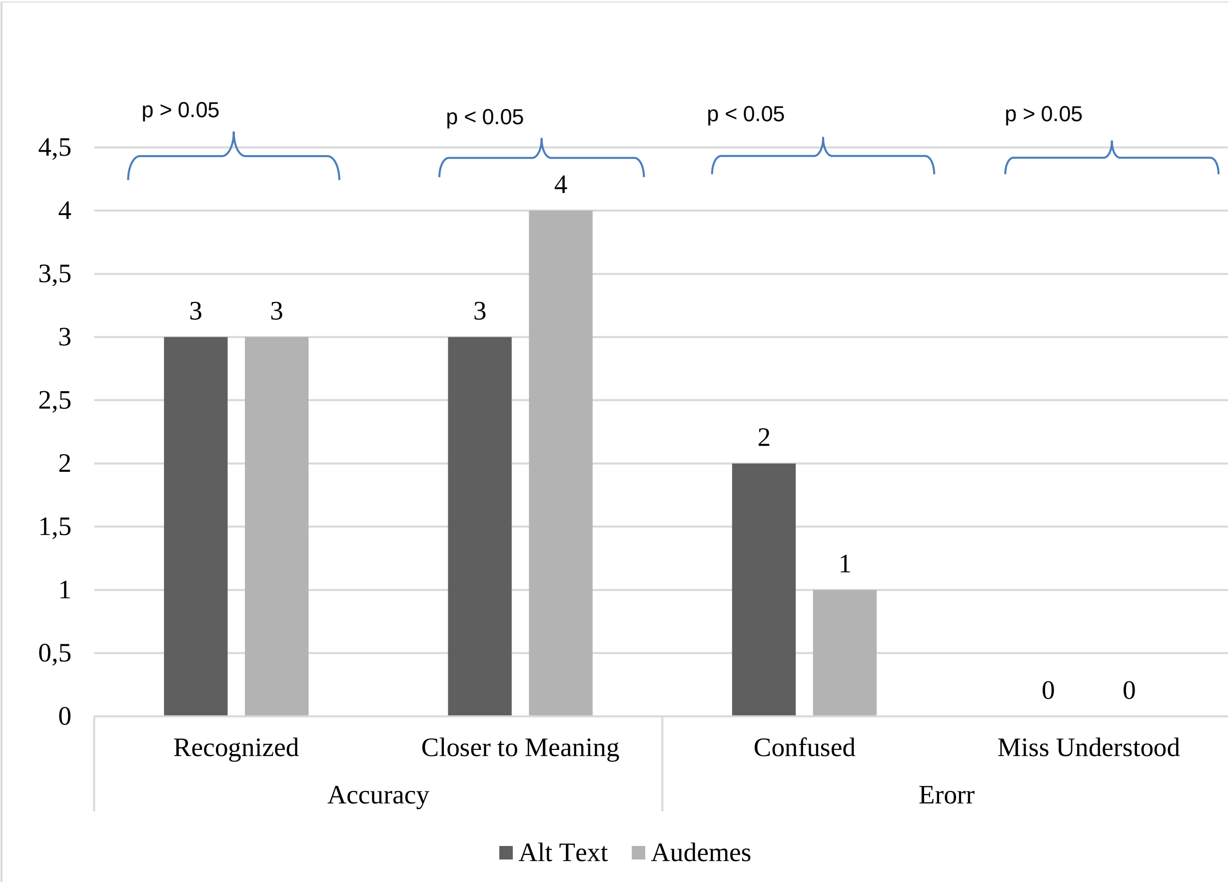

Figure 11.6 Accuracy and error rates means for image recognition

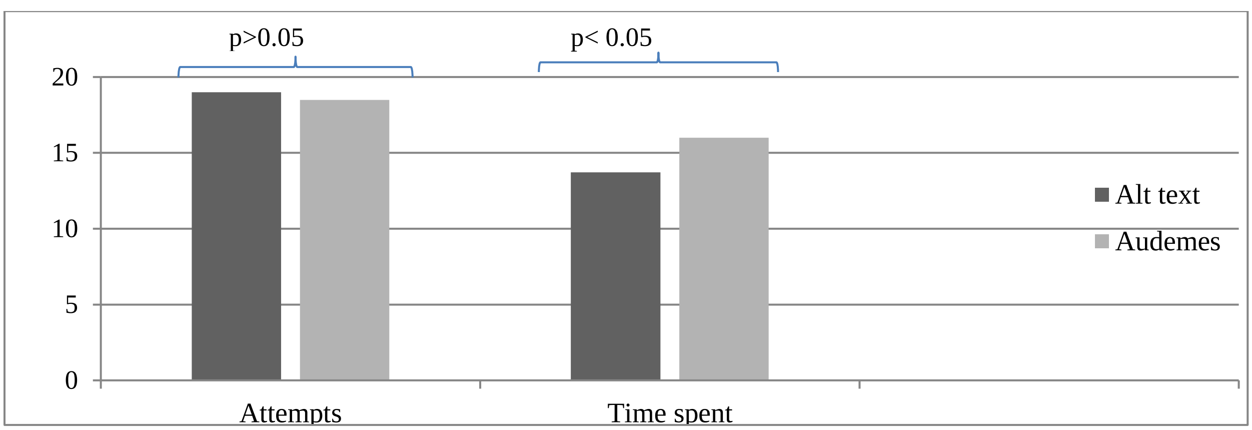

Figure 11.7 Attempts= mean number of times participants accessed test images in each prototype, Time spent= mean time spent by participants in each prototype in minutes

Figure 11.7 depicts two important measures, the cumulative number of attempts participants made to recognize the images, and the time spent in each prototype. We found significant difference for the later measurement, namely, participants spent more time with audemes (M=16, SD=2.44) compared to alt text (M=13.71, SD=1.38), $t_{(13)}=4.94$, $p<0.05$. In terms of attempts, the difference was not significant; audemes (M=18.5, SD=3.03) compared to alt text (M=19, SD=2.26), $t_{(13)}=0.07$, $p>0.05$.

In relation to these results, participants commented that they perceived audemes as richer in terms of information, which kept them think more and helped them visualize the images. Moreover, only in the condition with audemes we observed four participants who successfully identified all images, although it took them more attempts and more time than average. This indicates that audemes provided richer experience and helped participants immerse, and perform better than alt text.

12. Discussion

This paper shows that compared to alt text, enriching images with audemes significantly decreases the workload for screen reader users. Moreover, participants found it easier to recognize images using audemes as mental ability and perpetual activity required to perceive and recognize an image was diminished significantly. Audemes also offered a sense of enjoyment and no time pressure, which resulted in less effort and increased performance. Another important factor was the frustration users experienced when trying to understand the images. Participants found it difficult to process speech generated from alt text. On the other hand, they found listening to audemes to be pleasant.

The biggest workload perception difference between audemes enrichment of web image and alt text description came under frustration factors. Almost all the participants score audemes less irritational by half scales of TLX in compared to alt text. This shows that participants have difficult time in processing speech form of alt text description in compared to musical audemes representations. And, this went consistently with usability perception of the prototypes. Participants rated prototype with audemes enrichment of web image excellent and acceptable over prototype with alt text enriched web image as they perceived lower workload in identifying test image with audemes enrichment over alt text.

This article argues that the improvement of audemes compared to alt text, is attributed to the ability of the medium to communicate context in addition to content. Content includes the identity and properties of the object or events in the image, while context carries additional information about the content, such as, non-verbal cues, emotions, and environmental information. This article argues that, alt text is appropriate for communicating the content, while audemes also communicate context. Participants found the contextual experience missing with alt text, in contrast to audemes, which provided rich informational cues about the image. This contextual information further helped clarify the content of the image. Consequently, users' accuracy when identifying the images was increased. This is an indication that audemes provide richer experiences by communicating information about the content and context. While the content of an image may be simple to include in alt-text, contextual information is more difficult to describe in words. This claim, however, is based on an initial observation, and thus should be further investigated.

13. Limitations

13.1 Participants

Despite these results, this study, has several limitations. First, due to the difficulty in recruiting blind participants, the study simulated blindness with sighted participants, which influenced and introduced bias into the results. Blind people heavily depend on the auditory modality to substitute their lack of visual cues when interacting with the environment. This increases the effectiveness and efficiency with which they perceive and process auditory content compared with sighted people. Several studies report such claims, for example, compared to sighted people, blind people better utilize auditory information (Wk Niemeyer & I Starlinger, 1981; Starlinger & Niemeyer, 1981), they process auditory language stimuli faster (Röder et al., 2000), have enhanced congenitally processing of speech (Hugdahl et al., 2004), and they show better perception of degraded speech with equivalent hearing conditions (Gordon-Salant & Friedman, 2011). Hence, the outcomes of this study may differ with blind people.

Moreover, study with sighted participants might have induced other biased as well such as unfamiliar and unexperienced with TTS environment might have caused them difficult to perceive alt text descriptions. Irritation and frustration felt with synthetic speech might have led them to conclude heavier workload with alt text in compared to more musical audemes.

13.2 Test Image

The choice of images included in study make them more suitable for comparison between audemes and alt text. However, images found on web may vary and often depict content that is difficult to represent using non-speech sounds(Stephen A Brewster, 2002, p. 222). For example, it would be difficult to develop non-speech sounds to communicate an image containing numerical information, such as, prices, serial numbers, codes, dates and times. Also, it is difficult to represent color, size, structure and texture of various objects. This indicates the need to develop a complex audeme vocabulary and ontology that involves extensive training for the user. Additionally, although in this study we compared audemes and alt text in isolation, future research could test how and to what extent alt text and audemes complement one another.

13.3 Audemes Lengths

The length of the audemes for this study was set to four seconds to match the length of the speech representing the alt text. This length, however, might have influenced the results of the study, considering that other studies suggest audeme length to be up to seven seconds (Steve Mannheimer et al., 2009b). Moreover, the length of the audemes was a topic that was commented on by some of the participants.

13.4 Learnability Impact

All the participants who took part in experiment are not real screen reader user. This means their level of experience and expertise with text to speech environment and alt representation of image would have very inferiors to experienced screen reader user. Additionally, in training session participants were briefly introduced and trained with audemes but not with alt text. And, while conducting experiment participants were first tested with prototype with alt text representation of test images, then with the prototype enriched with audemes representation of test image. Since both the prototype has same structural layout where information and test image were organized in similar fashion, participants may have learned about that from the first tested prototype and that could have been found second prototype easy to explore than first one. This pre-acquired information could have played significant role in finding second prototype more usable.

13.5 Context of experiment

Image in webpage often have some natural and contextual association with other images around it and with the webpage itself. For example, name of webpage and genera or classification of contents has serious influence in image perceptions such as website of National Geography listing pictures of animals in animal genera, online grocery shopping listing images of green vegetables in vegetables genera. This contextual information might have helped to make up consideration and imagine pictorial content of the page that might have further influenced in comprehending alt text description of images to the screen reader users.

Moreover, providing appropriated alt text description to an image hugely depends on the context and surrounding of image itself. However, test image used in the study above were taken arbitrarily and did not has any association with test page and other surrounding test images which might have influenced comprehending alt text description to the participants.

13.6 Subjective Impact

Measurement scale used in the experiment to measure workload at identifying image and usability of the experimental prototype are subjective scales. This means collected data might have subjective influence. Though further subjective analysis of data shows low subjective impact on result, there might be huge subjective and perspective difference between sighted and real screen reader users. Further, more importantly, audemes for the test images were designed and used in semantic consideration close to test image, they were not approved by user test. Thus, design and use of audemes to represent test images in the test prototypes have subjective impact of experimenter consideration.

13.7 Complexity of measurement scales

Another limitation of this study comes along with the complexity of measurement scales used in this study. NASA TLX workload measurement scales often considered one of the complex scale to administrate and use. Since both measurement scales used in the study depend on subjective perceptions of participants, there might be huge difference in perception of measurement scales among the participants. A skilled and experienced user might find task easy to perform while others might have find same task tedious and difficult to perform in same situation. This could have been influenced the outcomes of the study.

14. Conclusion

In this study, we investigated the performance of audemes as compared to alt text for accessing images on the Web using screen readers. Overall, the findings indicated that audemes performed better across many dimensions. Specifically, they required lower mental and temporal demand, and resulted in less effort and frustration and better task performance. Moreover, audemes contributed to participants recognizing images with higher accuracy and lower errors. They were also perceived as more engaging compared to alt text delivered using synthetic speech. Additionally, audemes were found to deliver richer information by communicating the context in addition to the content of the image. These factors influenced the website’s usability, as the prototype with audemes was rated more usable than the prototype with alt text.

For future work, this article suggests investigating audemes in terms of delivering content and context. Additionally, this article recommends measuring the effect of audeme length on the process of image recognition. Finally, this article suggests replicating the study using blind participants to increase validity.

15. Study relevance to universal design

Universal design is a concept of universally accessible systems, a design concerns in addressing special needs of various user groups. People have special needs associated with their impairment and environmental factors(Nicolle & Abascal, 2001) such motor, cognitive, hearing, visual and many mores. Summing up all concept of usability, universal accessibility, universal design and inclusive design issues related to the requirement of feasible design to cover needs of all user groups shares relatively common grounds in human computer interaction. Where, our study highlighted numbers of significant factors associated with application of audemes, a non-speech sound, to increase accessibility in diverse area of design. The analysis of study suggests effective application of audemes in efficient design to overcome overloaded visual information in large display, design for small screen handheld devices, multimodal HCI design, Inclusive design for visually impaired and blind people, and so on.

15.1 Auditory display

Auditory display (Kramer, 1993)enables eyes-free human computer interaction for visually impaired as well as sighted users in multi task situation. Consequently, an attribute of hearing can contribute to the effectiveness of an auditory display even when vision is available. In this context, study above would have a significant impact in efficient auditory design to represent and process pictorial content. As audemes are found highly content and context rich non-speech sound to cover broader aspect of image in the web, it would be a new way to represent all other aspect of non-linguistic, non-verbal content in human computer interaction, a new prospect of advancing more natural human computer interactions.

15.1.1 Image comprehension

Picture are more memorable than words(Defeyter, Russo, & McPartlin, 2009) because human mind is extremely sensitive to symbolic presentation of information(Yuille, 2014). In other hand, picture facilitates deep level of information coding, and can carries huge amount of information. Thus, understanding image might take longer than words and comprehension is likely to be more subjective and abstract in user hands. Regarding this, selection and use of image to deliver or illustrate a piece of information or concept in web are also subjective to author knowledge and intention whose comprehension might be different among users. Alt text might be helpful in preserving author sole concept behind the image and maintaining uniform comprehension among the screen reader users. But, alt text used for describing image in web are found insufficient to cover all the content and contextual aspect of image to screen reader user in compared to audemes in study above. Therefore, audemes auditory enhancement of web image could be a solution. Audemes enhancement of web image could offers several advantages over alt text. First, audemes enhancement of web image offers multimodal advantage that users can listen image while they are engaged in secondary task or while image is unavailable (due to low internet bandwidth or disabled by browser). Second, as we found in study above that audemes are superior in encapsulating non-verbal, non-linguistic environmental and emotional contextual information in compared to alt text while representing image to screen reader users, this surplus amount of pre-coded non-intrusive informational enhance could make image more interactive and lively to the users and consequently could help them comprehending image content. The audemes enhancement might also serve as non-speech guidance to users in understanding image and could help in image comprehension process among all groups of web users that they all have exactly same understanding of image. Importantly, it ensures accessibility to the screen reader user. Besides these, paintings and portrayals have different place in depicting and emitting intricated and sensational message to the community. A single portrayal could carry huge amount information and present them in livelier and story forms such as historic battle portrayals, portrayals of culture and many mores. In this context, audemes could play a significant role where text fails to capture realistic impression and interpretation of portrayals.

15.1.2 Decorative Image

Decorative image could be a part of page design or text link or adjacent to textual description. Image as text link increase the accessibility by making link more visible and clickable to screen reader users. The iconic representation of link also provides short visual hint about the link itself. Image adjacent to textual description adds more clarity to the adjacent text. It serves as topic or overview for adjacent text. Third, decorative image as part of page design used for adding up ambience or visual attractiveness to the page. In all case as a link or adjacent to textual description or page design part, decorative image plays a vital role in supplement or setting up context for the linked content. However, WCAG 2.0 (success criteria 1.1.1) guidelines ignore inclusion of alternative text to decorative image in consideration that they do not add any information to the content of page, and as a question of repetition of information already present in adjacent text and link itself. In contrary to this, guidelines itself debate on declaring image as decorative or informative on same ground of use entirely depends on author decision and motives of use. In either case, whether decorative image might not add any content to the page, it carries surplus amount of contextual information as a theme to content of page that might help user engaged to page and retain focused in contents. In this case, audemes might play a significant role in presenting decorative image as a musical theme to the content which might further helpful to user in mediating through the content by reducing stress level and adding up contextual information.

15.1.3 Group of Image

Collection of imagery icons (such as multiple star symbols used for representing “ratings”) or image itself (such as collection of images that represent a theme or a piece of information, for examples wallpaper images, weather or climate or scenery or close related images) were often used in webpage as an informative symbol or picture. Alternative textual description of such group of images are often lengthy and incapable to complement informational enrichment of image groups. In this case, audemes could be a short and precise musical representation to whole groups of images.

15.1.4 Animated web Content

Animated content of webpage includes all the blinks, flash, flicker, add-on applets or scripts and pop up visual displays such as advertisements, alerts, extended auto-run and event triggered contents. Basically, animated content is widely used in webpage design to give more visual polishing or signing effects to the webpage. It gives webpage stylist and attractive look as well as provides effective interaction design. In certain cases, it makes presentation shorts and save screen space too. Animated content in webpage are always available with option to switch in deactivated mood for the users. However, animated content makes webpage look good, they cause significant amount of distraction to the users as well. Sudden pop ups and auto-run animated content in webpage increases the visual load and creates confusion to the cognitive disable users. In addition, these animated contents become complex to blinds or visually impaired user to access and comprehend. Animated contents are specially designed to give a quick and brief information to uses such alerts, ads and news. Blind user either ignored them as to overcome visual clutters (Gupta & Kaiser, 2005) and use simplified textual version(Parmanto et al., 2005) or missed them due to quickly changeable behavior. To ensure accessibility to these content, generally alternative text is used for blinking and flickering images, animated text or video are scripted and others clutter animations were advised to avoid or scape with non-animated methods.

While animated content of webpage improves user interaction by making page more dynamic, operative, visually attractive along with reducing long textual description through short visual presentation. Simultaneously, it increases difficulties in accessibility of page to cognitive and blind users. To the perplexity of using animated content in designing visually engaging and stylist webpage and ensuring accessibility at the same time, audemes enhancement could be a reasonable solution. Audemes could be equally informative as animated content and might have tendency to carry equal visual effects as well.

15.1.5 Visual stickers

Visual stickers are now popular among instant messenger users and so far, gained much popularity in use. They offer a new way to express non-verbal human expressions and intimacy effects complementing long text typing or using multiple emoji(Novak, Smailović, Sluban, & Mozetič, 2015). Stickers could be a picture of simple combination or complex blends of emotions and words or animation effects that can be used in message alone or along with other textual content to add more emphasis and emotional expression to communication. Basically, they add visual moods and reaction to the communication by incorporating image, emoticons and phrases that connects people at an expressive and emotional level. Thus, they make conversation more funny, enjoyable and expressive by delivering simulation of visual depiction of emotions, gestures and intension. However, visual sticker offers great intuition in expressive communication, they are equally inaccessible and unusable to the visually impaired users. Since text itself suffers from lack of expressiveness, alternative text used in covering accessibility of stickers are incapable of complement visual stickers. In this context, audemes enhancement of visual stickers could be helpful in increasing usability and accessibility of communication platforms as audemes found significant and effective at representing web image in compared to alt text.

15.1.6 Enhancing navigational maps

Navigational system or maps requires high level of visual attention while navigating through unfamiliar highly urbanized area whether in on-board in-vehicle or mobile phone based portable system. Speech provides an easy way to navigate through navigational system in dual task scenario where user could devote their sight in more demanding primary task such as driving or walking through street. Consequently, speech also ensure accessibility to blinds and visually impaired user to navigate their destination and surroundings on the way to destination with voice command and instruction such as device(Golledge, Klatzky, Loomis, & Marston, 2004) and product NavCue(Chen, Plaza-Leiva, Min, Steinfeld, & Dias, 2016). Beside these, speech over visual map also filters large volume of visual data into silkiest information only required to navigate destination. Users generally misses most of others information such as place and object which falls around the path towards destination. On the other hands, however speech enables non-visual access and minimizes overloaded visual information to the users, speech could be annoying to the users over listening everything up to the destination and could distress and distracts user from their primary tasks due limited human speech processing ability. For example, tracking speech enabled user interface and having conversation at the same time would be a difficult task(Wenzel, 1992). Similarly, it is found that speech has degraded quality at presenting information and lower user acceptance ratings(B. Q. Tran, 2000).

In this scenario, auditory enhancement of visual navigational systems or map would be a better option to increase the usability of the system and secure accessibility to the blinds and visually impaired users than standalone speech(Loomis, Golledge, & Klatzky, 1998). Because, Speech are hard to localize than non-speech sound in virtual environment(T. V. Tran, Letowski, & Abouchacra, 2000). Moreover, speech is generally lengthier and need to hear sentence or phrases completely before understanding them clearly. Consequently, leads to informationally clutters message when listening over multiple things like directions and objects over very short pause(Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Non-speech sounds were found non-intrusive to the primary information and enabled user to navigate through virtually environment by increasing usability and accessibility of the product such as Swan(Wilson et al., 2007). However, there has been extensive use and experiment with development of acoustic enhanced navigation map along with integration of speech, structural and sematic boundaries of all available non-speech sounds (Earcons, Spearcons, Auditory Icons, Musicons) somehow limiting the usability of full flex acoustic enhanced navigation system. In this context, findings from the above study suggest that audemes could delivery acoustic cues necessary for full flex auditory enhanced navigational system where all others non-speech sounds fail to provide.

15.1.7 Image processing

Image processing has significant importance and application in remote sensing, autonomous guidance and intelligent navigation system, autonomous surveillance and mores. The recent advancement in image processing algorithm especially feature extractions and edge detection, optical character processing, 3D and geometry, neural nets have extensive importance in assisting and increasing accessibility to impaired users. For examples, auditory mapping of real time video streams with camera, auditory translation of billboards or printed characters, real time language translation, eye free and free movement navigation system for blinds and impaired users, indoor and outdoor travelling assistance and many mores.

Blinds and visually impaired users are primarily relying on haptics and speech assistance modalities for accessing information and communication system. But accessing visual information such as images and video streams is limited to the provided textual description such as captions for the videos and alt text for the images. However, text cover important aspect of information besides its lack of expressiveness (Barker & Manji, 1989)and lengthier description, presentation of information in the form of speech adds extra difficulties to the users. For examples, even at as low intensities speech are more disruptive than non-speech sounds when recalling information(SalamÉ & Baddeley, 1990) and accuracy of recognition decreased with presentation rated of speech(Aldrich & Parkin, 1989). Others are annoyance and sounds clutters produced in auditory feedback and interface design. Importance of non-speech increased since the studies found that non-speech sounds are easily to habituate, suitable for sonification of continuous data and non-intrusive effect in user interface enhancement. Consequently, found effective in mapping video stream line into audio presentation(Martínez, Villegas, Sánchez, de Jesús Ochoa Domínguez, & Maynez, 2011), image segmentation and feature detections, computer vision and GIS to navigate building for blind peoples (Bhowmick, Prakash, Bhagat, Prasad, & Hazarika, 2014) and image features classification based in sound(Dennis, Tran, & Li, 2011). In this context, audemes could be a new way to index and sort all web image as it is found highly perceivable than alt text in study above. Audemes representation of web image could help maintain database and information mining in different level than before. As image is a mostly used source of data presentation and coding in academics and business, it accumulates large amount of highly classified and analyzed economics and scientific data, auditory representation of those data will result into lower cost in data storage, retrieval, analysis and utilization across all platforms. Moreover, internet and its whole application, sites are now full of images, pictures and graphics. Auditory representation of those pictorial content will significantly boost the bandwidth and lower down the computing resource at user machines such as auditory representation of graphics content requires low screen resolution and little space on screen, support lower battery consumption, less processing power, offers multimodalities operation and increase accessibility to blinds and visually impaired persons and many mores.

Image processing specially for computer vison such as in robotics is highly costly in terms of computation. it requires highly sophisticated algorithm and instruments like as camera and sensors for mapping objects and their properties in real environment. This cost will be overcome with auditory mapping such as video surveillance. Surveillance is important in an aspect of monitoring and maintaining security, and has been implanted and operating in all most everywhere from public place video surveillance such as street, traffic, stations, airport to private home, business, properties. This generates huge of amount of data around a day, week, month and year. Data is generally store and process in routinely basis for monitoring and security inspections, and this process cost huge amount of resource from accumulating, storing to processing. This could be hugely minimized with auditory representation of the content, and among speech and others non-speech audemes proves to be suitable as claims for web image in above study.

15.2 Inclusive Design

Inclusive design (Jenny, 1993)with concept that usability can only be defined in terms of the specific user group for which the system is being developed(Nielsen, 1994), have proved to be very useful for special needs user groups. This user centered design strategy for people with disability within user groups ensures that users are the heart of the design process. In regard this, image which has been a ubiquitous source of information all over the internet constantly posing informational barriers to the blind and visual impairments users. we investigate and evaluate a new alternative auditory approach to represent web image to these user in compared to existing methods, alternative text. In our experiment of non-speech sound enrichment of web image to increase accessibility to the screen reader user establishes a new research paradigm in representing web image in auditory display. The outcomes of study highlighted the significance of audemes non-speech sound in encapsulating content and contextual information of web image in compared to alternative text, and we found audemes are highly usable and significant at improving web image accessibility, user accuracy at identifying image has been significantly improved, they felt lower workload pressures and annoyance. As the audemes enrichment of web image tends to enhance the usability of webpage and increase the informational enrichments, this indicates that audemes could be a new design methods to represent image for blinds and visual impairments users. Thereby, it is suggested to be extremely valuable in inclusive design as well as design under suboptimal conditions(Nicolle & Abascal, 2001).

15.2.1 Sonic presentation

Audemes could be an acoustic medium to ensure accessibility of real time streams data to blinds and visual impaired user such as weather forecasting information, camera based mapping of visual information and sonification of others health and fitness related data and notification in smart mobile phones. Since, audemes could complement textual information(Steven Mannheimer et al., 2009) it could be applicable for providing musical summery of textual content which might be applicable in assisting blinds user in mail and message reading and corresponding. Study above demonstrated that audemes are significant at encapsulating content and contextual data and retains the potential to compliment text as well. Thus, potential further might be applicable in large platform such as text mining and analytics. Besides these, audemes enrichment of web image introduced another research dimension in image indexing, sorting and mining in internet. Allover, these could be able to broaden the scope of accessibility to blinds and visually impaired user by widening limited scopes imposed by screen reader application. Also, by offering more entertaining and musical way to perceive visual information.

15.2.2 Computer User Interaction

Most blind and visually impaired people rely on screen reader software for locating information on screens and reading them out with synthetic speech. Consequently, screen readers offer easy and precise access to the digital content (especially textual content) to blinds and visually impaired users. Besides some problems with slow and serial nature of information and annoying factors associated with speech, application provides accessibility to the digital content. Similarly, refreshable braille display(Schmidt, Lisy, Prince, & Shaw, 2002) devices, an integration of traditional braille into refreshable system where user can access content by haptics or tactile perception ensures accessibility to deaf-blind people as well. Others are voice recognition software that allow user to use computer or mobile phones by controlling through voice commands such as data inputs(Ballagas, Borchers, Rohs, & Sheridan, 2006). It provides an alternative method over traditional keyboard to blind people but application suffers from sophistication of voice recognition algorithm as well as proximity and situation of usage such as difficult to use in noisy and public environment. Moreover, vibrotactile(Wellman & Howe, 1995; Wiker, Vanderheiden, Lee, & Arndt, 1991) application were emerged for assisting blind people in physical and abstract information delivery(Choi & Kuchenbecker, 2013). But, subjective influence in vibrotactile perception among users and technical complexity in vibrotactile medium makes it difficult and less feasible for designers to implement and design for inclusive needs. Apart from these, importance of non-speech sound in Human computer interaction is increasing such as auditory enhancement of computer user interface, sonification of continuous streams data and many mores. Auditory enhancement of user interface offers flawless non-interfering pleasant user interactions and increases usability of the system alongside ensuring accessibility to blinds and visually impaired users.

Non-speech sound offer several advantages over speech like removing cultural dependency, language barriers, delay and frustration in message interpretation. Others distinctive characteristic includes highly localizable, easy to habituate and suitable for multimodal systems. Non-speech sounds have been used for several decades to represent information in a computer user interface. Several types have been designed depending on the specifics of the information intended to be delivered. Mostly used non-speech sounds to represent brief objects or information on a user interface are earcons, auditory icons, spearcons and their altered version for more specification and elaboration of use such as Lyricons, auditory scrollbar, spindex. Earcons represent abstract sound we hear when we receive an email. Musical grammar of earcons makes it difficult to understand for non-musician, and induce hard learning curve to user because of its arbitrary meaning mappings rather than any logically meaning association between sound and the object it represents. While an auditory icon is the sound of a crumpling piece of paper when deleting a file. Auditory icons are suitable for user interactions, alerts, and helpful for navigation. Its true application is based on the direct representation of an associated concept. It is, however, very difficult to accurately classify or create an auditory icon for every word or concept. Considering that auditory icons are based on the natural sound an object makes, there is an intuitive link between the sound and the object or concept that it represents. In other words, it leverages the knowledge people have of natural events and uses the same sounds to represent an object or event in a user interface. Over earcons and auditory icons, spearcons contains acoustic advantages. Acoustically it sounds like speech which generally produced by compressing output of text to speech translation of screen reader application. Close acoustic association with speech provides fast and clear acoustic cue about the text which has been found more perceived in acoustic enhanced menu navigation in user interface, presenting mathematical symbols and navigational cues in navigation system or maps. This object representation and message encoding scope of non-speech sound has further increased by the addition of auditory emoticons, spemoticons, anthropomorphic auditory icons to encapsulate and represents human emotions.

In contrary to these, study above demonstrates that audemes have the both potential to convey long message and complex human and environmental contextual data. It generally sums ups the all the non-speech sounds and fill ups the gap present on their semantic structures. This allows us to develop more naturally and user friendly human computer interaction. It would be applicable in presenting human notion of communication in artificial artifacts.

15.2.3 Communication

Communication platform now incorporate complex feature beyond text messaging. Now people are more get to use with visual and emotional content in communication to imitate their deep sense of emotions and make their conversation more expressive such as emoticons, subjective emoji, animated clips, doodling, stickers, pictures and cartoons. In scenario, platform is becoming meaner, complex and inaccessible for blinds and visually impaired users. Thus, audemes as being semantically more flexible multi-vocal content rich non-speech sound could be an acoustic enhancement for tool for the communication channel that will make platform more expressive and lively as well as accessible for blinds and visually impaired users.

15.2.4 Educations

Educating blind children generally base on tactile and auditory exploration of universe. They mostly rely on their hearing sensation to compensate visual perception of surroundings and environment around them. Their learning and exploring world begins with sensational happening hearing and touching at once. Braille has been a basis for introducing them to educational material where they began to explore letters, number and words. Other one is speech, they receive most of the communication and interaction through speech. Audio materials and braille content all together plays a sole role in education. Recent digitalization of information, they greatly rely on text to speech screen reader application for almost everything from education to communication. Digital device ranging from handheld smart cell phone to specially navigating tool allows them to free movements and roaming around environment. But, blind user higher suffers from the text to speech assistive tool as the tool mostly produce synthetic speech which is quite difficult to cope for long time in compared natural human voice. Unarticulated natural synthetic speech does not encapsulate human emotions, and lacks the expressiveness of information that further degrades the quality of information. Similarly, platform incompatibility, limited access to the graphical content, annoyance, sequential nature, delay and hard to operate in noisy environment makes it less usable to the users.

Non-speech auditory enhancement of computer user interface already confirmed the usability of sound in human computer interaction but syntactic and semantics boundaries of the sound limiting their information enrichment to short message and alert signal. They are not suitable for longer information coding and message. But, Audemes provide a novel category of non-speech sounds that can be used to represent complex content. They were initially invented and tested with blind and visually impaired users to convey thematic content(M. Ferati et al., 2012). The design of audemes is based on empirical knowledge, which often results in the creation of sounds derived from the personal preference of the sound designer. It could contain sounds alone or in combination from natural events or abstract, musical tones. A study shows that audemes significantly improve and increase the recognition of concepts for blind and visually impaired participants (Steven Mannheimer et al., 2009). Similarly, audemes for content navigation were tested on a touch-screen interface and they were found to be easy to learn, memorable and navigable for visually impaired teenagers (M. Ferati et al., 2009). In terms of information retaining, audemes were found helpful in reducing memory erosion and even after five months, the content was better remembered with audemes than without them (Steve Mannheimer et al., 2009b). Along with study above which shows that audemes can complement alt text in encapsulating content and contextual information of web image with reduced workload and increased usability to screen reader users. Studies above indicate that audemes can plays a significant role in educating blind children. I might deliver same textual educational materials in more entertaining, expressive and lively format then speech and braille. Audemes could be integrated with haptics and vibrotactile interactive learning environment to educated and assist blind children in working environment as well. Audemes as being non-speech with more musical and multi-vocal semantic structure have tendency to depicts large source and diverse genera of information. it could be used for providing medication information along with notification to elder people, educating blind children and others as well. Specially, graphical and pictorial diagrams, figures of learning materials in school which are difficult to represent with speech and explain to blinds student could be enriched with audemes. Moreover, learning problems faced with mathematical contents such as mathematical equations and annotations which are even hard for sighted children to grasp easily, could be signified with audemes vocabular to keep equations representation short and simple than speech. Metaphoric or mnemonics impact of audemes tends to be higher in learning and memorization which can help blind children to get over long textual description or explanation with short audemes play. Audemes not only delivers the content, it also makes content more engaging and moving by incorporating non-verbal, non-linguistic sensational emotional context and subjective expression to the message. This suggests audemes could be an alternative non-visual method to represent animated school material to the blinds and visually impaired childrens.

15.3 Multimodal Design

Multimodal design is a characteristic of system to facilitate multi modal user interaction which includes enriching human computer interaction with human natural behaviors or discourse for more natural mode of communication(Bourguet, 2003). The multimodal design of system concentrate on multimodal interfaces for multiple human perception (Turk & Robertson, 2000) which consequently improves system accessibility for different users' groups and application contexts. It intensifies systems robustness, expressiveness, and efficiency of communication(Flickner, Darrell, , & Oviatt, 2004). In this context, Audemes representation of web image could enables a more flexible and natural mode of communication, interfacing users with non-speech auditory representation of pictorial content. It could offer a more efficient and usable environment allowing users to perceive image just by listening them. As audemes are found well perceived more entertaining and engaging musical representation of web image, auditory enhancement of image with audemes preserve a significant importance in multitask scenario where user could devote their visual attention more priority task and get information about image by listening semantically structured musical patterns.

Nowadays, small handheld device and wearable technologies rapidly taking market place compared to desktop and laptop system which have limiting usability due to restricted input output modalities and portability. These small devices hugely incorporate multimodal user interaction system to overcome small screen display and limited space for visual content. Similarly, excessive visual content on the screen could easily leads to visual clutters and reduce usability of the device. Overloaded visual information on screen also pose accessibility problems to elderly and cognitive impaired user as well. Multimodality in user interaction not only helps to reduce overloaded visual information it also facilitates eye free human computer interactions. For example, voice commands and speech recognition novel applications which even improved for noisy environment as well. Face detection and finger print sensors for more secure biometric security, haptics and vibrotactile smart graphical icons for enhancing accessibility to special need users, multimedia communication platforms where user could adopt any of text or audio or video modalities for interactions and many mores. Multi modal design plays significant role in accessibility of the devices when users are travelling such as receiving message and establishing communication, accessing control to others on broad vehicle systems like music players, navigation maps. It offers interactivity to the device even when users are busy in primary task such as walking or driving or working, communication could be established by audio or audio visual surveillance which improves safety in multitask scenario as well as increase productivity. In regards of visual display which requires extensive visual attention and obtrusive in situational application such as outdoor daylight, multimodal designs offer more dynamic and user interactive interface to the sighted and impaired users as well.

Importance of non-speech sound and research in multimodal sonifying application has recently been increased because of some limitation exist with speech, text and graphical presentation of information in human computer interactions. Limitation of display screens in small portable system, length and lack of expressiveness of textual content, serial medium and annoyance related with speech feedbacks are some of the reasons why research and application of non-speech sound gained so much importance. Other characteristics of non-speech sounds are easy to habituates, non-intrusive, highly localizable, high acoustic association with compressed audio, best for presenting continuous trends data, unattended effects unlike speech, musical grammar for coding highly complex information, natural association, less annoyance than speech, melody entertaining nature, high pitch and intensity than speech, and many mores.

Among others non-speech sound, audemes could be used to enhanced graphical user interface, mobile computing and sonification. The difference in application of audemes lies on its semantic structure to address complex and longer message than other non-speech sounds. Whereas other sounds are best for warning and notifying users about events and interaction with system, providing short message and auditory cue in navigation to the users, audemes could maps even more longer and complex data in multimodal interaction and designs. For examples, musical summarization of textual content where user could listen the short version of musical description of long textual contents , could assist in mail ready by adding contextual climax such as emotion and expressions, expressing contextual information of data as background play might be applicable in data presentation, auditory representation of complex multimedia content such as pictures and paintings, auditory representation of illustrative content such as simulation and animated materials and many mores. Audemes could be used in guiding and assisting platforms along with others visual modalities likes signs and videos. For example, training peoples in different activities like as sports, exercise, computer gaming, mass management and hazard emergency situations. Others application of audemes in multimodal perspective would be in auditory surveillance and process administration where system could track the unusual and sudden changes in the environment, and could inform user about issues along with necessary measures in the form of non-speech guidance.

16 Closing remarks

This thesis investigates the information enrichment of audemes to encapsulate and represent web image to the screen reader user. First, document evaluate effective audemes to cover content and contextual aspect of information of web image in comparative study with alt text. Consequently, confirms the hypothesis of the study that audemes enrichment of web image significantly reduce the workload demand and frustration level as well as increase the usability of web page in compared to alt text. Second, based on finding it highlights the potentials of audemes in universal design of information and communication technology.

16. Appendices

16.1 Appendix A

16.1.1 Consent form

Participant ID: _______

CONSENT FORM

Research Title: "Enriching Web images with non-speech sounds to increase accessibility for blind users"

Researcher: Ratan Bahadur Thapa

Contact address: Okern Torgvei 92
0589, Oslo
Rb_thapa@hotmail.com
+47 465 85 561

Please Tick!

☐ I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.

☐ I understand that my participations is voluntary and I am free to withdraw at any time, without giving any reason.

☐ I agree that data collected as a consequence of my participations may be stored anonymously and used for research purposes. Upon my withdrawal data collected as a consequence of my participation will be destroyed.

☐ I do not have any physical and mental health problem which could become aggravated by my using computer device.

☐ I agree to not reveal the details of this study to anyone as this research is still ongoing.

____________________	____________________	____________________
Name of Participant	Signature	Date
____________________	____________________	____________________
Name of Investigator	Signature	Date

16.2 Appendix B

16.2.1 Demographic questionnaire

AUDEMES

Participant ID:________ Date:__________

Demographic Survey

Instructions: Please do not write your name on the form. All the responses to this survey are kept confidential. This survey gathers general information about you and your knowledge of computer, the Internet and World Wide Web.

Please check the box which best describes you.

1. Age:
   ☐ Less than 20 yrs ☐ 20 – 24 yrs ☐ 25 -29 yrs
   ☐ 30 – 34 yrs ☐ 35 – 49 yrs ☐ 40 yrs or Older
2. Gender:
   ☐ Male ☐ Female
3. Education Level completed:
   ☐ Basic school ☐ High school ☐ Bachelors
   ☐ Master Degree ☐ Doctorate
4. Experience with computer:
   ☐ Less than a year ☐ 1 – 2 yrs ☐ 3 – 5 yrs
   ☐ 6 – 10 yrs ☐ 11- 15 yrs ☐ More than 15 yrs
5. Experience with World Wide Web:
   ☐ Less than a year ☐ 1 – 2 yrs ☐ 3 – 5 yrs
   ☐ 6 – 10 yrs ☐ 11- 15 yrs ☐ More than 15 yrs
6. Internet Access per hours a day:
   ☐ Less than an hour ☐ 1 – 2 hrs ☐ 3 – 5 hrs
   ☐ 6 – 10 hrs ☐ More than 10 hrs
7. Purposes of using Internet:
   ☐ Entertainment ☐ Communication ☐ Academic
   ☐ Professional ☐ Commercial ☐ Others ....
   ☐ More than single purpose

AUDEMES

8. Experience with Text To Speech (TTS) translator application:

   ☐ Yes ☐ No

   8A. Application used on:

   ☐ Smart Phone ☐ Laptop ☐ Desktop ☐ Others ....
   ☐ More than single devices

   8B. Purpose of usage:

   ☐ Reading documents or files ☐ Browsing internet ☐ Others ....
   ☐ More than single purpose

16.2.2 NASA TLX workload questionnaire

Date: ___/___/___

Participant ID: _____

NASA TLX (Task Load Index)
Subjective Numerical Measurement Techniques

Instructions: Participants are requested to make relative comparison of the task with respect to two different test scenario (prototypes) by using TLX complexity scale of 0 to 100, by increment of 5. Mark on the right of the TLX scale indicates increase whereas on the left of the TLX scale indicates decrease in the workload complexity experienced by the participants .

TASK: " Identifying test images accurately in webpage while accessing non-visually."

	Prototype A : AUDEMES	**Prototype B : ALT TEXT**
	Very high Very low	Very high Very low
	100 90 80 70 60 50 40 30 20 10	100 90 80 70 60 50 40 30 20 10
1. How mentally demanding was the task(e.g. thinking, deciding, remembering etc)?		
2. How physically demanding was the task(e.g. pushing, pulling, controlling etc)?		
3. How hurried or rushed was the pace of the task?		
4. How successful were you in accomplishing what you were asked to do ?		
5. How hard did you have to work (mentally and physically) to accomplish your level of performance?		
6. How insecure, discouraged, irritated, stressed and annoyed were you?		

16.2.3 **System Usability Scale questionnaire**

Date: ___/___/___

Participant ID: _____

System Usability Scale

Instructions: For each of the following statements, mark one box that best describes your reactions to the respective test prototypes. If you don't know how to respond , simply check box "3".

		Prototype A: AUDEMES					Prototype B: ALT TEXT				
		Strongly disagree				Strongly agree	Strongly disagree				Strongly agree
1.	I think that I would like to use this webpage frequently.	1	2	3	4	5	1	2	3	4	5
2.	I found this webpage unnecessarily complex.	1	2	3	4	5	1	2	3	4	5
3.	I thought this webpage was easy to use.	1	2	3	4	5	1	2	3	4	5
4.	I think that I would need assistance to be able to use this page.	1	2	3	4	5	1	2	3	4	5
5.	I found the various functions in this webpage were well integrated.	1	2	3	4	5	1	2	3	4	5
6.	I thought there was too much inconsistency in this webpage.	1	2	3	4	5	1	2	3	4	5
7.	I would imagine that most people would learn to use this webpage very quickly.	1	2	3	4	5	1	2	3	4	5
8.	I found this webpage very cumbersome and awkward to use.	1	2	3	4	5	1	2	3	4	5
9.	I felt very confident using this webpage.	1	2	3	4	5	1	2	3	4	5
10.	I needed to learn a lot of things before I could get going with this webpage.	1	2	3	4	5	1	2	3	4	5

16.3 Appendix C

16.3.1 Observation sheet

AUDEMES

Participant ID: ____________ Date: ____________

Prototype: ____________

Start time: ____________ End time: ____________ Duration: ____________

Test Image

1.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning
☐ Recognised

2.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning
☐ Recognised

3.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning
☐ Recognised

4.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning
☐ Recognised

5.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning

6.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning

7.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning

8.

Attempt:
☐
☐
☐
☐

Outcomes:
☐ confused
☐ missunderstood
☐ closer to meaning

17 References

Access, N. (2009). NonVisual Desktop Access (NVDA).

Access, N. (2010). NVDA: Retrieved.

Access, N. (2015). NVDA features.

Albers, M. C. (1994). *The Varese System, hybrid auditory interfaces and satellite-ground control: Using auditory icons and sonification in a complex, supervisory control system.* Paper presented at the Proceedings of the Second International Conference on Auditory Display.

Albrecht, R., Väänänen, R., & Lokki, T. (2016). Guided by music: pedestrian and cyclist navigation with route and beacon guidance. *Personal and Ubiquitous Computing, 20*(1), 121-145.

Aldrich, F. K., & Parkin, A. J. (1989). Listening at speed. *British Journal of Visual Impairment, 7*(1), 16-18.

Alseid, M., & Rigas, D. (2011). *The role of earcons and auditory icons in the usability of avatar-based e-learning interfaces.* Paper presented at the Developments in E-systems Engineering (DeSE), 2011.

Alty, J. L. (1995). *Can we use music in computer-human communication?* Paper presented at the BCS HCI.

Andò, B., Baglio, S., La Malfa, S., & Marletta, V. (2010). Innovative smart sensing solutions for the visually impaired. *Handbook of Research on Personal Autonomy Technologies and Disability Informatics*, 60-74.

Andò, B., Baglio, S., Marletta, V., & Pitrone, N. (2009). *A mixed inertial & RF-ID orientation tool for the visually impaired.* Paper presented at the Systems, Signals and Devices, 2009. SSD'09. 6th International Multi-Conference on.

Association, G. A. o. t. W. M. (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists, 81*(3), 14.

Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to Perceive with a Visuo—Auditory Substitution System: Localisation and Object Recognition with 'The Voice'. *Perception, 36*(3), 416-430.

Balakrishnan, G., Sainarayanan, G., Nagarajan, R., & Yaacob, S. (2005). *Stereo image to stereo sound methods for vision based ETA.* Paper presented at the Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering, 2005. CCSP 2005. 1st International Conference on.

Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The smart phone: a ubiquitous input device. *IEEE Pervasive Computing, 5*(1), 70-77.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of experimental psychology: human perception and performance, 19*(2), 250.

Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89*(3), 499-517.

Banf, M., & Blanz, V. (2013). *Sonification of images for the visually impaired using a multi-level approach.* Paper presented at the Proceedings of the 4th Augmented Human International Conference.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies, 4*(3), 114-123.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction, 24*(6), 574-594.

Barker, P., & Manji, K. A. (1989). Pictorial dialogue methods. *International Journal of Man-Machine Studies, 31*(3), 323-347.

Battiste, V., & Bortolussi, M. (1988). *Transport pilot workload: A comparison of two subjective techniques.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Baulk, S. D., Kandelaars, K. J., Lamond, N., Roach, G. D., Dawson, D., & Fletcher, A. (2007). Does variation in workload affect fatigue in a regular 12-hour shift system? *Sleep and Biological Rhythms, 5*(1), 74-77.

Bhowmick, A., Prakash, S., Bhagat, R., Prasad, V., & Hazarika, S. M. (2014). *IntelliNavi: Navigation for blind based on Kinect and machine learning.* Paper presented at the International Workshop on Multi-disciplinary Trends in Artificial Intelligence.

Bittner, A. C., Byers, J. C., Hill, S. G., Zaklad, A. L., & Christ, R. E. (1989). *Generic workload ratings of a mobile air defense system (LOS-FH).* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Blattner, M. M., & Dannenberg, R. B. (1992). *Multimedia interface design*: ACM.

Blattner, M. M., & Greenberg, R. M. (1992). Communicating and learning through non-speech audio *Multimedia interface design in education* (pp. 133-143): Springer.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human–Computer Interaction, 4*(1), 11-44.

Bleicher, J., & Bleicher, J. (1980). *Contemporary hermeneutics: Hermeneutics as method, philosophy and critique*: Routledge & Kegan Paul London.

Bly, S. (1982). Sound and computer information presentation: Lawrence Livermore National Lab., CA (USA); California Univ., Davis (USA).

Bolea, Y., Grau, A., & Sanfeliu, A. (2003). *Non-speech sound feature extraction based on model identification for robot navigation.* Paper presented at the Iberoamerican Congress on Pattern Recognition.

Bourbakis, N. (2008). Sensing surrounding 3-D space for navigation of the blind. *IEEE Engineering in Medicine and Biology Magazine, 27*(1).

Bourguet, M.-L. (2003). *Designing and Prototyping Multimodal Commands.* Paper presented at the INTERACT.

Brewster, S. (1998). The design of sonically-enhanced widgets. *Interacting with computers, 11*(2), 211-235.

Brewster, S. A. (1998). *Using earcons to improve the usability of tool palettes.* Paper presented at the CHI 98 Cconference Summary on Human Factors in Computing Systems.

Brewster, S. A., Wright, P. C., & Edwards, A. D. (1994). *The design and evaluation of an auditory-enhanced scrollbar.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Brewster, S., Leplâtre, G., & Crease, M. (1998). *Using non-speech sounds in mobile computing devices.* Paper presented at the Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices.

Brewster, S., Raty, V.-P., & Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy *People and Computers XI* (pp. 169-183): Springer.

Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces.* University of York York, UK.

Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Displays, 17*(3-4), 179-189.

Brewster, S. A. (1998). Using nonspeech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction (TOCHI), 5*(3), 224-259.

Brewster, S. A. (2002). Non-speech auditory output. *Human Computer Interaction Handbook, Lawrence Erlbaum Associates, Mahwah, NJ*, 220-239.

Brewster, S. A., & Cryer, P. G. (1999). *Maximising screen-space on mobile computing devices.* Paper presented at the CHI'99 extended abstracts on Human factors in computing systems.

Brewster, S. A., Raty, V.-P., & Kortekangas, A. (1996). *Enhancing scanning input with non-speech sounds.* Paper presented at the Proceedings of the second annual ACM conference on Assistive technologies.

Brewster, S. A., Wright, P. C., Dix, A. J., & Edwards, A. D. (1995). The sonic enhancement of graphical buttons *Human—Computer Interaction* (pp. 43-48): Springer.

Brewster, S. A., Wright, P. C., & Edwards, A. D. (1993). *An evaluation of earcons for use in auditory human-computer interfaces.* Paper presented at the Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems.

Brewster, S. A., Wright, P. C., & Edwards, A. D. (1995a). *Guidelines for the creation of earcons.* Paper presented at the Adjunct Proceedings of HCI.

Brewster, S. A., Wright, P. C., & Edwards, A. D. (1995b). Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies, 43*(22), 153-175.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4-7.

Brown, L. M., & Brewster, S. A. (2003). *Drawing by ear: Interpreting sonified line graphs*.

Buxton, W. (1989). Introduction to this special issue on nonspeech audio. *Human Computer Interaction, 4*(1), 1-9.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior research methods, 41*(1), 113-117.

Capelle, C., Trullemans, C., Arno, P., & Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE transactions on biomedical engineering, 45*(10), 1279-1293.

Chan, A. H., & Ng, A. W. (2009). Perceptions of implied hazard for visual and auditory alerting signals. *Safety science, 47*(3), 346-352.

Chen, K., Plaza-Leiva, V., Min, B.-C., Steinfeld, A., & Dias, M. B. (2016). *NavCue: Context immersive navigation assistance for blind travelers.* Paper presented at the The Eleventh ACM/IEEE International Conference on Human Robot Interaction.

Chisholm, W. V., & Jacobs, G. (1999). I.(editors)(1999) Web Content Accessibility Guidelines (WCAG). *W3C Recommendation, 5*.

Choi, S., & Kuchenbecker, K. J. (2013). Vibrotactile display: Perception, technology, and applications. *Proceedings of the IEEE, 101*(9), 2093-2104.

Consortium, W. C. W. W. W. (2013). Web content accessibility guidelines 2.0 (WCAG 2.0).

Consortium, W. W. W. (2008). Web content accessibility guidelines (WCAG) 2.0.

Corwin, W. H., Sandry-Garza, D., Biferno, M., Boucek, G., Logan, A., Jonsson, J., & Metalis, S. (1989). Assessment of crew workload measurement methods, techniques, and procedures: Process, methods and results. *Resport n WRDC-TR-89-7006*.

Cowling, M., & Sitte, R. (2002). Analysis of speech recognition techniques for use in a non-speech sound recognition system.

Csapo, A. B., & Baranyi, P. (2010). *An interaction-based model for auditory substitution of tactile percepts.* Paper presented at the Intelligent Engineering Systems (INES), 2010 14th International Conference on.

Deatherage, B. H. (1972). Auditory and other sensory forms of information presentation. *Human engineering guide to equipment design*, 123-160.

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development, 24*(3), 265-273.

Delogu, F., Palmiero, M., Federici, S., Plaisant, C., Zhao, H., & Belardinelli, O. (2010). Non-visual exploration of geographic maps: does sonification help? *Disability and Rehabilitation: Assistive Technology, 5*(3), 164-174.

Dennis, J., Tran, H. D., & Li, H. (2011). Spectrogram image feature for sound event classification in mismatched conditions. *IEEE Signal Processing Letters, 18*(2), 130-133.

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: an index for detecting mental workload variation in complex tasks. *Applied ergonomics, 42*(6), 807-813.

DiDomenico, A., & Nussbaum, M. A. (2011). Effects of different physical workload parameters on mental workload and performance. *International Journal of Industrial Ergonomics, 41*(3), 255-260.

Diener, E., & Crandall, R. (1978). *Ethics in social and behavioral research*: U Chicago Press.

DiGiano, C. J., & Baecker, R. M. (1992). *Program auralization: sound enhancements to the programming environment.* Paper presented at the Proceedings of the conference on Graphics interface'92.

Dingler, T., Lindsay, J., & Walker, B. N. (2008). *Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech*.

Dobrucki, L. W., & Sinusas, A. J. (2010). PET and SPECT in cardiovascular molecular imaging. *Nature Reviews Cardiology, 7*(1), 38-47.

Dombois, F. (2001). Using audification in planetary seismology.

Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagné, J.-P., Leclerc, C., & Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental brain research, 160*(2), 194-202.

Edwards, A. D. (1997). *Using sounds to convey complex information.* Paper presented at the Contributions to Psychological Acoustics: Results of the Seventh Oldenburg Symposium on Psychological Acoustics, Oldenburg.

Ekman, P. (1993). Facial expression and emotion. *American psychologist, 48*(4), 384.

Esposa Jr, J. I. (2008). How to write a good alt text.

Evett, L., & Brown, D. (2005). Text formats and web design for visually impaired and dyslexic readers—clear text for all. *Interacting with computers, 17*(4), 453-472.

Fabiani, M., Dubus, G., & Bresin, R. (2010). *Interactive sonification of emotionally expressive gestures by means of music performance.* Paper presented at the Proceedings of the 3rd interactive sonification workshop.

Fagerlönn, J. (2011). *Making auditory warning signals informative: examining the acceptance of auditory icons as warning signals in trucks.* Paper presented at the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Fagerlönn, J., & Alm, H. (2010). Auditory signs to support traffic awareness. *IET Intelligent Transport Systems, 4*(4), 262-269.

Fehr, L., Langbein, W. E., & Skaar, S. B. (2000). Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey. *Journal of rehabilitation research and development, 37*(3), 353.

Ferati, M., Mannheimer, S., & Bolchini, D. (2009). *Acoustic interaction design through audemes: experiences with the blind.* Paper presented at the Proceedings of the 27th ACM international conference on Design of communication.

Ferati, M., Mannheimer, S., & Bolchini, D. (2011). *Usability evaluation of acoustic interfaces for the blind.* Paper presented at the Proceedings of the 29th ACM international conference on Design of communication.

Ferati, M., Pfaff, M. S., Mannheimer, S., & Bolchini, D. (2012). Audemes at work: Investigating features of non-speech sounds to maximize content recognition. *International Journal of Human-Computer Studies, 70*(12), 936-966.

Ferati, M. A. (2012a). *Design foundations for content-rich acoustic interfaces: Investigating audemes as referential non-speech audio cues.* faculty of the University Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in the School of Informatics, Indiana University.

Ferati, M. A. (2012c). *Design foundations for content-rich acoustic interfaces: Investigating audemes as referential non-speech audio cues.* faculty of the University Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in the School of Informatics, Indiana University.

Fernström, M., & Brazil, E. (2004). *Human-computer interaction design based on interactive sonification-hearing actions or instruments/agents.* Paper presented at the Proceedings of the 2004 International Workshop on Interactive Sonification.

Ferrari, E., Robins, B., & Dautenhahn, K. (2009). *Therapeutic and educational objectives in robot assisted play for children with autism.* Paper presented at the Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on.

Fitzsimons, C., Murphy, E., Mulwa, C., & Fitzpatrick, D. SpatialMaths: a Library for Conveying Content and Structure of Equations.

Flickner, M., Darrell, , T., & Oviatt, S. (2004). Multimodal interfaces that flex, adapt, and persist. *Communications of the ACM, 47*(1), 30-33.

Frankel, L., & Brownstein, B. (2016). An evaluation of the usefulness of prosodic and lexical cues for understanding synthesized speech of mathematics. *ETS Research Report Series, 2016*(2), 1-19.

Frémiot, M. (1999). *De l'objet musical aux unités sémiotiques temporelles.* Paper presented at the Ouïr, entendre, écouter, comprendre après Schaeffer.

Froehlich, P., & Hammer, F. (2004). *Expressive Text-to-Speech: A user-centred approach to sound design in voice-enabled mobile applications.* Paper presented at the Proc. Second Symposium on Sound Design.

Fuller, J. A., Adams, G. G., & Buxton, B. (1989). Atrial fibrillation after coronary artery bypass grafting. Is it a disorder of the elderly? *The Journal of Thoracic and Cardiovascular Surgery, 97*(6), 821-825.

Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013). *Advanced auditory cues on mobile phones help keep drivers' eyes on the road.* Paper presented at the Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-computer interaction, 2*(2), 167-177.

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human-computer interaction, 4*(1), 67-94.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology, 5*(1), 1-29.

Gaver, W. W. (1997). Auditory interfaces. *Handbook of human-computer interaction, 1*, 1003-1041.

GAVIN, M., JEDIR, R., & NEFF, F. (2016). Sonification playback rates during matching tasks of visualised and sonified EEG data.

Golledge, R., Klatzky, R., Loomis, J., & Marston, J. (2004). Stated preferences for components of a personal guidance system for nonvisual navigation. *Journal of Visual Impairment & Blindness (JVIB), 98*(03).

Gomez, J. D., Bologna, G., & Pun, T. (2012). *Non-visual-cueing-based sensing and understanding of nearby entities in aided navigation.* Paper presented at the Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility.

Gomez Valencia, J. D. (2014). *A computer-vision based sensory substitution device for the visually impaired (See ColOr).* University of Geneva.

Gordon-Salant, S., & Friedman, S. A. (2011). Recognition of rapid speech by blind and sighted older adults. *Journal of Speech, Language, and Hearing Research, 54*(2), 622-631.

Graham, R., Hirst, S., & Carter, C. (1995). *Auditory icons for collision-avoidance warnings.* Paper presented at the Intelligent Transportation: Serving the User Through Deployment. Proceedings of the 1995 Annual Meeting of ITS America.

Gupta, S., & Kaiser, G. (2005). *Extracting content from accessible web pages.* Paper presented at the Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A).

Harada, S., Landay, J. A., Malkin, J., Li, X., & Bilmes, J. A. (2006). *The vocal joystick:: evaluation of voice-based cursor control techniques.* Paper presented at the Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility.

Harada, S., Landay, J. A., Malkin, J., Li, X., & Bilmes, J. A. (2008). The Vocal Joystick: evaluation of voice-based cursor control techniques for assistive technology. *Disability and Rehabilitation: Assistive Technology, 3*(1-2), 22-34.

Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology, 52*, 139-183.

Hart, S. G., & Wickens, C. D. (1990). Workload assessment and prediction *Manprint* (pp. 257-296): Springer.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: when is one scale better than many? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(4), 579-601.

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook*: Logos Verlag Berlin.

Hugdahl, K., Ek, M., Takio, F., Rintee, T., Tuomainen, J., Haarala, C., & Hämäläinen, H. (2004). Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. *Cognitive brain research, 19*(1), 28-32.

Hussain, I., Chen, L., Mirza, H. T., Chen, G., & Hassan, S.-U. (2015). Right mix of speech and non-speech: hybrid auditory feedback in mobility assistance of the visually impaired. *Universal Access in the Information Society, 14*(4), 527-536.

Hussain, I., Chen, L., Mirza, H. T., Wang, L., Chen, G., & Memon, I. (2016). Chinese-Based Spearcons: Improving Pedestrian Navigation Performance in Eyes-Free Environment. *International Journal of Human-Computer Interaction, 32*(6), 460-469.

Ihm, C.-G., Park, J.-K., Hong, S.-P., Lee, T.-W., Cho, B.-S., Kim, M.-J., . . . Ha, H. (1998). A high glucose concentration stimulates the expression of monocyte chemotactic peptide 1 in human mesangial cells. *Nephron, 79*(1), 33-37.

Jacko, J. A. (1996). The identifiability of auditory icons for use in educational software for children. *Interacting with computers, 8*(2), 121-133.

Janata, P., & Childs, E. (2004). Marketbuzz: Sonification of real-time financial dataa.

Jenny, P. (1993). A Guide to usability-human factors in computing: Wokingham, England, Addison-Wesley.

Jeon, M. (2013). *Lyricons (Lyrics+ Earcons): Designing a new auditory cue combining speech and sounds.* Paper presented at the International Conference on Human-Computer Interaction.

Jeon, M. (2015). *Exploration of Semiotics of New Auditory Displays: A Comparative Analysis with Visual Displays.* Paper presented at the Proc. of the 21st Int. Conference on Auditory Display (ICAD2015).

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009). *Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies.* Paper presented at the Proceedings of the 1st international conference on automotive user interfaces and interactive vehicular applications.

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction, 31*(1), 1-16.

Jeon, M., & Walker, B. N. (2009). *"Spindex": Accelerated initial speech sounds improve navigation performance in auditory menus.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Jeon, M., & Walker, B. N. (2011). Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Transactions on Accessible Computing (TACCESS), 3*(3), 10.

Jornet, J., Esquius, F., Seren, M., Tuset, D., Solaz, J. S., & de Rosario, H. (2008). HMI: An Emotional Experience. *ATZautotechnology, 8*(1-2), 50-54.

Kacorri, H., Riga, P., & Kouroupetroglou, G. (2014). *EAR-math: evaluation of audio rendered mathematics.* Paper presented at the International Conference on Universal Access in Human-Computer Interaction.

Karshmer, A. (2007). Access to mathematics by blind students: A global problem.

Katz, B. F., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., . . . Jouffrais, C. (2012). NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality, 16*(4), 253-269.

King, A. R., & Evans, D. S. (2007). *Re-presenting visual content for blind people*: University of Manchester.

Kirakowski, J. (1996). The software usability measurement inventory: background and usage. *Usability evaluation in industry*, 169-178.

Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K., & Nakano, M. (2010). *Artificial subtle expressions: Intuitive notification methodology of artifacts.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces*: Perseus Publishing.

Kramer, G. (1994). *An introduction to auditory display.* Paper presented at the Auditory Display: Sonification, Audification and Auditory Interfaces, SFI Studies in the Sciences of Complexity, Proceedings.

Kramer, P. D. (1994). Listening to prozac.

Krishnan, K. G., Porkodi, C., & Kanimozhi, K. (2013). *Image recognition for visually impaired people by sound.* Paper presented at the Communications and Signal Processing (ICCSP), 2013 International Conference on.

Krokstad, A., Svean, J., & Ramstad, T. A. (1994). Programmable hybrid hearing aid with digital signal processing: Google Patents.

Larsson, P., & Niemand, M. (2015). Using sound to reduce visual distraction from in-vehicle human–machine interfaces. *Traffic injury prevention, 16*(sup1), S25-S30.

Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *The Journal of the Acoustical Society of America, 126*(6), 3147-3155.

Leung, Y. K., Smith, S., Parker, S., & Martin, R. (1997). *Learning and retention of auditory warnings*.

Leventhal, J. (2005). Not what the doctor ordered: A review of Apple's voiceover screen reader.

Lewis, J. R., & Sauro, J. (2009). *The factor structure of the system usability scale.* Paper presented at the International Conference on Human Centered Design.

Liu, Y., & Wickens, C. D. (1994). Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. *Ergonomics, 37*(11), 1843-1854.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments, 7*(2), 193-203.

Ludi, S., Simpson, J., & Merchant, W. (2016). *Exploration of the Use of Auditory Cues in Code Comprehension and Navigation for Individuals with Visual Impairments in a Visual Programming Environment.* Paper presented at the Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility.

Machida, W., & Itoh, T. (2011). *Lyricon: A visual music selection interface featuring multiple icons.* Paper presented at the Information Visualisation (IV), 2011 15th International Conference on.

Mannheimer, S., Ferati, M., Bolchini, D., & Palakal, M. (2009a). *Educational sound symbols for the visually impaired.* Paper presented at the International Conference on Universal Access in Human-Computer Interaction.

Mannheimer, S., Ferati, M., Bolchini, D., & Palakal, M. (2009b). Educational sound symbols for the visually impaired. *Universal Access in Human-Computer Interaction. Addressing Diversity*, 106-115.

Mannheimer, S., Ferati, M., Huckleberry, D., & Palakal, M. J. (2009). *Using Audemes as a Learning Medium for the Visually Impaired.* Paper presented at the HealthINF.

Mansur, D. L., Blattner, M. M., & Joy, K. I. (1985). Sound graphs: A numerical data analysis method for the blind. *Journal of medical systems, 9*(3), 163-174.

Martínez, B., Villegas, O., Sánchez, V., de Jesús Ochoa Domínguez, H., & Maynez, L. (2011). Visual perception substitution by the auditory sense. *Computational Science and Its Applications-ICCSA 2011*, 522-533.

Matthews, T., Fong, J., & Mankoff, J. (2005). *Visualizing non-speech sounds for the deaf.* Paper presented at the Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility.

McGee-Lennon, M., Wolters, M., McLachlan, R., Brewster, S., & Hall, C. (2011). *Name that tune: musicons as reminders in the home.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

McGee-Lennon, M. R., Wolters, M., & McBryan, T. (2007). Audio reminders in the home environment.

McLachlan, R., McGee-Lennon, M., & Brewster, S. (2012). *The sound of musicons: Investigating the design of musically derived audio cues*.

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE transactions on biomedical engineering, 39*(2), 112-121.

Meyer, C., Ahsan, S., Gupta, V., Parham, A., Patel, H., & Qureshi, M. (2014). Audemes: Exploring the Market Potential of a Sound Based Educational Tool.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). *A comparison of two scoring procedures with the NASA task load index in a simulated flight task.* Paper presented at the Aerospace and electronics conference, 1992. NAECON 1992., proceedings of the IEEE 1992 national.

Murphy, E., Bates, E., & Fitzpatrick, D. (2010). *Designing auditory cues to enhance spoken mathematics for visually impaired users.* Paper presented at the Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility.

Mynatt, E. D., & Weber, G. (1994). *Nonvisual presentation of graphical user interfaces: contrasting two approaches.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Nagarajan, R., Yaacob, S., & Sainarayanan, G. (2002). *Fuzzy clustering in vision recognition applied in NAVI.* Paper presented at the Fuzzy Information Processing Society, 2002. Proceedings. NAFIPS. 2002 Annual Meeting of the North American.

Nees, M. A., & Best, K. (2013). *Modality and encoding strategy effects on a verification task with accelerated speech, visual text, and tones*.

Németh, G., Olaszy, G., & Csapó, T. G. (2011). *Spemoticons: Text to Speech Based Emotional Auditory Cues*.

Nicolle, C., & Abascal, J. (2001). *Inclusive design guidelines for HCI*: CRC Press.

Nielsen, J. (1994). *Usability engineering*: Elsevier.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness II. Central functions. *Audiology, 20*(6), 510-515.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. II. Central functions. *Audiology, 20*(6), 510-515.

Nikolakis, G., Moustakas, K., Tzovaras, D., & Strintzis, M. (2005). Haptic representation of images for the blind and the visually impaired. *CD-ROM Proc. HCI International*.

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PloS one, 10*(12), e0144296.

Palladino, D. K. (2007). *Efficiency of spearcon-enhanced navigation of one dimensional electronic menus.* Citeseer.

Palladino, D. K., & Walker, B. N. (2007). *Learning rates for auditory menus enhanced with spearcons versus earcons*.

Papastavrou, J. D., & Lehto, M. R. (1996). Improving the effectiveness of warnings by increasing the appropriateness of their information content: some hypotheses about human compliance. *Safety science, 21*(3), 175-189.

Parente, P., & Bishop, G. (2003). *BATS: the blind audio tactile mapping system.* Paper presented at the Proceedings of the ACM Southeast Regional Conference.

Park, S. H., Kim, S., Lee, S., & Yeo, W. S. (2010). *Composition with path: Musical sonification of geo-referenced data with online map interface*: Ann Arbor, MI: Michigan Publishing, University of Michigan Library.

Park, Y., Heo, H., & Lee, K. (2015). Enhanced auditory feedback for Korean touch screen keyboards. *International Journal of Human-Computer Studies, 73*, 1-11.

Parmanto, B., Ferrydiansyah, R., Saptono, A., Song, L., Sugiantara, I. W., & Hackett, S. (2005). *AcceSS: accessibility through simplification & summarization.* Paper presented at the Proceedings of the 2005 international cross-disciplinary workshop on web accessibility (W4A).

Parseihian, G., & Katz, B. F. (2012). Morphocons: A new sonification concept based on morphological earcons. *Journal of the Audio Engineering Society, 60*(6), 409-418.

Patten, M. L. (2016). *Questionnaire research: A practical guide*: Routledge.

Pirhonen, A., Murphy, E., McAllister, G., & Yu, W. (2006). Non-speech sounds as elements of a use scenario: a semiotic perspective.

Poláček, O., Sporka, A. J., & Slavik, P. (2012). A Comparative Study of Pitch-Based Gestures in Nonverbal Vocal Interaction. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 42*(6), 1567-1571.

Preece, A. (2016). An Overview of NVDA Remote Access, a Free Add-on for the Non Visual Desktop Access Screen Reader.

Ramloll, R., & Brewster, S. (2002). *A generic approach for augmenting tactile diagrams with spatial non-speech sounds.* Paper presented at the CHI'02 Extended Abstracts on Human Factors in Computing Systems.

Ramloll, R., Yu, W., Brewster, S., Riedel, B., Burton, M., & Dimigen, G. (2000). *Constructing sonified haptic line graphs for the blind student: first steps.* Paper presented at the Proceedings of the fourth international ACM conference on Assistive technologies.

Ramsey, H. R., & Atwood, M. E. (1980). *Man-computer interface design guidance: State of the art.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Röder, B., Rösler, F., & Neville, H. J. (2000). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia, 38*(11), 1482-1502.

Rømen, D., & Svanæs, D. (2012). Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users. *Universal Access in the Information Society, 11*(4), 375-385.

Roth, P., Petrucci, L. S., Assimacopoulos, A., & Pun, T. (2000). Audio-haptic internet browser and associated tools for blind users and visually impaired computer users.

Roussel, P., Negishi, A., & Mitsukura, Y. (2016). *Presentation of an SMR-based BCI using auditory feedback based on pitch.* Paper presented at the Mechatronics (MECATRONICS)/17th International Conference on Research and Education in Mechatronics (REM), 2016 11th France-Japan & 9th Europe-Asia Congress on.

Roy, D., Pappu, N., & Prabhakar, T. (2004). *Parichaykrama–an exploratory interface of indian classical music using experiential framework.* Paper presented at the International Symposium on Computer Music Modeling and Retrieval.

SalamÉ, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society, 28*(6), 540-542.

Saldhana, J., & Shatz, S. M. (2000). *Uml diagrams to object petri net models: An approach for modeling and analysis.* Paper presented at the International Conference on Software Engineering and Knowledge Engineering.

Sanchez, J. (2010). *Recognizing shapes and gestures using sound as feedback.* Paper presented at the CHI'10 Extended Abstracts on Human Factors in Computing Systems.

Sánchez, J., Lumbreras, M., & Cernuzzi, L. (2001). *Interactive virtual acoustic environments for blind children: computing, usability, and cognition.* Paper presented at the CHI'01 Extended Abstracts on Human Factors in Computing Systems.

Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design*: McGRAW-HILL book company.

Sanderson, P. M., Liu, D., & Jenkins, S. A. (2009). Auditory displays in anesthesiology. *Current Opinion in Anesthesiology, 22*(6), 788-795.

Schaeffer, P., Mâche, F. B., Philippot, M., Bayle, F., Ferrari, L., Malec, I., & Parmegiani, B. (1967). *La musique concrète*: Presses universitaires de France.

Scherer, K. R. (1994). Affect bursts. *Emotions: Essays on emotion theory, 161*, 196.

Schmidt, R. N., Lisy, F. J., Prince, T. S., & Shaw, G. S. (2002). Refreshable braille display system: Google Patents.

Schmitz, M., Fehringer, B. C., & Akbal, M. (2015). *Expressing Emotions With Synthetic Affect Bursts.* Paper presented at the Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play.

Scientific, F. (2014). JAWS for Windows® Screen Reading Software. *Freedom Scientific*.

Shoval, S., Borenstein, J., & Koren, Y. (1998). Auditory guidance with the navbelt-a computerized travel aid for the blind. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 28*(3), 459-467.

Slatin, J. M. (2001). The art of ALT: toward a more accessible Web. *Computers and Composition, 18*(1), 73-81.

Slowiaczek, L. M., & Nusbaum, H. C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 27*(6), 701-712.

Smither, J. A.-A. (1993). Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology, 12*(6), 330-335.

Stanton, A. N. (1982). Virtual sound system for the visually handicapped. *Acoustical Society of America Journal, 72*, 1347.

Starlinger, I., & Niemeyer, W. (1981). Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness I. Peripheral Functions. *Audiology, 20*(6), 503-509.

Suh, H., Jeon, M., & Walker, B. N. (2012). *Spearcons improve navigation performance and perceived speediness in Korean auditory menus.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*(3), 917-928.

Sun, Y., & Jeon, M. (2015). *Lyricon (Lyrics+ Earcons) Improves Identification of Auditory Cues.* Paper presented at the International Conference of Design, User Experience, and Usability.

Syal, P., Chatterji, S., & Sardana, H. K. (2016). DiGVis: A system for comprehension and creation of directed graphs for the visually challenged. *Universal Access in the Information Society, 15*(2), 199-217.

Tkacik, P. T. (2011). System for seeing using auditory feedback: Google Patents.

Tran, B. Q. (2000). Speech recognizer: Google Patents.

Tran, T. V., Letowski, T., & Abouchacra, K. S. (2000). Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics, 43*(6), 807-827.

Turk, M., & Robertson, G. (2000). Perceptual user interfaces (introduction). *Communications of the ACM, 43*(3), 32-34.

Turner, C. W., Safar, J. A., & Ramaswamy, K. (2006). *The effects of use on acceptance and trust in voice authentication technology.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Valero, X., & Alias, F. (2012). Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. *IEEE Transactions on Multimedia, 14*(6), 1684-1689.

Vickers, P. (2013). Ways of listening and modes of being: Electroacoustic auditory display. *arXiv preprint arXiv:1311.5880*.

Walker, B. N., & Kogan, A. (2009). Spearcon performance and preference for auditory menus on a mobile phone *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments* (pp. 445-454): Springer.

Walker, B. N., & Kramer, G. (1996). Mappings and metaphors in auditory displays: An experimental assessment.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. *Ecological psychoacoustics*, 150-175.

Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP), 2*(4), 407-412.

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*(2), 265-278.

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors, 55*(1), 157-182.

Walker, B. N., Nance, A., & Lindsay, J. (2006a). Spearcons: Speech-based earcons improve navigation performance in auditory menus.

Walker, B. N., Nance, A., & Lindsay, J. (2006b). *Spearcons: Speech-based earcons improve navigation performance in auditory menus*.

Waseem, M., & Sujatha, C. (2014). Speech Synthesis System for Indian Accent using Festvox. *International Journal of Scientific Engineering and Technology Research, ISSN*, 2319-8885.

Wellman, P., & Howe, R. D. (1995). Towards realistic vibrotactile display in virtual environments. *Proc. ASME DSCD, DSC, 57*(2), 713-718.

Wenzel, E. M. (1992). Localization in virtual acoustic displays. *Presence: Teleoperators & Virtual Environments, 1*(1), 80-107.

Wersenyi, G. (2008). *Evaluation of user habits for creating auditory representations of different software applications for blind persons*.

Wersenyi, G. (2009). *Evaluation of auditory representations for selected applications of a graphical user interface*.

Wersényi, G. (2010a). Auditory representations of a graphical user interface for a better human-computer interaction *Auditory Display* (pp. 80-102): Springer.

Wersényi, G. (2010c). Auditory representations of a graphical user interface for a better human-computer interaction. *Auditory Display*, 80-102.

Wiker, S. F., Vanderheiden, G., Lee, S., & Arndt, S. (1991). *Development of tactile mice for blind access to computers: importance of stimulation locus, object size, and vibrotactile display resolution.* Paper presented at the Proceedings of the Human Factors Society Annual Meeting.

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007). *Swan: System for wearable audio navigation.* Paper presented at the Wearable Computers, 2007 11th IEEE International Symposium on.

Won, Y.-J., & Hey, J. (2005). Divergence population genetics of chimpanzees. *Molecular biology and evolution, 22*(2), 297-307.

Xu, J., Fang, Z. G., Dong, D. H., & Zhou, F. (2010). *An outdoor navigation aid system for the visually impaired.* Paper presented at the Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference on.

Yalla, P., & Walker, B. N. (2008). *Advanced auditory menus: design and evaluation of auditory scroll bars.* Paper presented at the Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility.

Yoshida, T., Kitani, K. M., Koike, H., Belongie, S., & Schlei, K. (2011). *EdgeSonic: image feature sonification for the visually impaired.* Paper presented at the Proceedings of the 2nd Augmented Human International Conference.

Yuille, J. C. (2014). *Imagery, Memory and Cognition (PLE: Memory): Essays in Honor of Allan Paivio* (Vol. 27): Psychology Press.

Zhao, S., Brumby, D. P., Chignell, M., Salvucci, D., & Goyal, S. (2013). Shared input multimodal mobile interfaces: Interaction modality effects on menu selection in single-task and dual-task environments. *Interacting with computers, 25*(5), 386-403.