LAHDEN AMMATTIKORKEAKOULU
Lahti University of Applied Sciences

Natural Interface to Improve Human-Computer Interaction for People with Upper Limb Disabilities

Exploring the Potentials of Voice Input and Hand Gestures in Application Development to Improve the Communication Possibilities of People with Motor Disorders

LAHTI UNIVERSITY
OF APPLIED SCIENCES
Degree program
in Business Information Technology
Bachelor's Thesis
Spring 2014
Alexey Salnikov

Lahti University of Applied Sciences
Degree Programme in Business Information Technology

SALNIKOV, ALEXEY: Natural Interface to Improve Human-Computer Interaction for People with Upper Limb Disabilities. Exploring the Potentials of Voice Input and Hand Gestures in Application Development to Improve the Communication Possibilities of People with Motor Disorders

Bachelor's Thesis in Business Information Technologies 47 pages, 00 pages of appendices

Spring 2014

ABSTRACT

According to a report from the Department of Justice on Enforcement of the Americans with Disabilities (2011), 36% of people with severe disability, aged between 15 and 64, used a computer, and 29% used the Internet at home. This low percentage of computer users, who are disabled at different levels, could be boosted by the development of a Natural User Interface. However, the study at hand considers focuses on building a development framework of a Natural User Interface for the physically impaired people, affected by upper limb disorders.

The present study explores the potentials of Text-To-Speech and Speech-To-Text technologies, as well as hand gestures' recognition, adapted for people with upper limb motor disorders, in the development of a Natural User Interface.

This research was conducted qualitatively, and it is explorative in nature. As the base for conclusion drawing is represented by a continuously developing application (STTK), based on people's needs (a tester) and feedback, the study was built based on the Design Science framework.

STTK, was tested twice by a person with motor disabilities caused by an upper limb disorder, and was developed based on their needs and feedback.

The resulting prototype makes the voice and Text-To-Speech input available and accessible for the focus audience, and its core functionality can be flexibly extended according to the type, severity and peculiarities of user's disorder.

The importance of an application's clear User Interface, which can be used from a great distance and which is permanently responsive, despite the number of applications running in the background, was pointed out as a result of the study. Furthermore, punctuation and other "details" as such, have proven to be of a great importance for application users. The STTK application should be further developed according to the findings of the study. A Natural Computer Interface can be implemented without custom, very expensive, or rare devices.

Key words: Hand Gesture Recognition, Text-to-Speech, Speech-to Text, Natural Computer Interface, Microsoft Kinect, Google Text-to- Speech API.

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

AAC	Augmentative and Alternative Communication
AI	Artificial Intelligence
API	Application Programming Interface
ASHA	American Speech-Language-Hearing Association
AT	Assistive Technology
CDC	Centers for Disease Control and Prevention
GUI	Graphical User Interface
IS	Information System
MMORPG	Massive Multiple Player Online Role Playing Game
NCI	Natural Computer Interface
NPC	Non-Player Characters – in-game character, not controlled by players
NUI	Natural User Interface
OCR	Optical Character Recognition
STTK	Speech-To-Text plus Kinect – The intermediate name of the application at the moment of writing
STT	Speech-to-Text
TrabHCI	Technologies to reduce access barrier in human-computer interaction
TTS	Text-to-Speech
UI	User Interface
WPF	Windows Presentation Foundation

1 INTRODUCTION

Regardless the position on the social scale, gender or any other issues that could cause discrimination, individuals have the right to communication. The National Joint Committee for the Communicative Needs of Persons with Severe Disabilities highlights the neglected right of persons with severe motor disorders to express themselves by any communication means, generally used by all other people. According to the above mentioned Committee, each person has the right to "request desired objects, actions, events and people; refuse undesired objects, actions, or events; express personal preferences and feelings; etc." (ASHA, 1992). Furthermore, the American Speech-Language-Hearing Association stresses on the importance of giving the physically impaired individuals the possibility to access AAC (augmentative and alternative communication) and other AT (assistive technology) services and devices at all times. (ASHA, 1992).

There are several definitions for the concept of communication that were developed by scientists and scholars. As the communication process is a very complex one, it is to be noticed, that all of them have a common factor: the "exchange of information". According to American Speech-Language-Hearing Association (ASHA) the process of communication represents "any act by which one person gives to, or receives from another person, information…" The Encyclopaedia Britannica confirms the meaning of communication as "the act or process" undertaken in order to "express or exchange information" (The Encyclopaedia Britannica, 2013).

Over the years, there were a few attempts of developing solutions that could enhance the possibilities of communication for people with severe motor disorders (e.g.: eye tracking systems, head mouse controllers, screen readers, etc.). These solutions, however, were found by their users to be superficial (Magee, 2011).

The present study focuses on the possibilities, in terms of communication, ATs offer to the physically impaired individuals. In a fast-paced world, where technology takes over every-day life actions, such persons are in the impossibility of using basic internet servers to pay bills, get access to social service related materials, and communicate with family and friends by, for instance, e-mail.

The research at hand presents the potentials of voice input and hand gestures in application development, to improve the communication possibilities of people with motor disorders. The existing ATs, such as eye tracking systems, head mouse controllers, screen readers, etc., are proven to be superficial and hurdle the natural movement. The present research will explore the usefulness of an application, combining voice input and hand gestures, in the communication process involving the physically disabled individuals.

2 RESEARCH BACKGROUND

The present chapter will guide the reader through the key concepts of the study, as well as through the way the research will be conducted, while presenting its motivational grounds and structure.

2.1 Definition of Key Concepts

The key concepts standing at the base of the current studies are: Motor Disability Hand Gesture Recognition, Text-to-Speech, Speech-to-Text, Natural Computer Interface, Microsoft Kinect, and Google Text-to- Speech API.
They will be defined in the following section of the paper at hand.

2.1.1 Motor disability

Motor disabilities, which refer to movement disorders, represent a group of neurological and musculoskeletal disorders that affect the motor and movement systems.

Some of these disorders can cause from extreme difficulty to move, up to excessive, uncontrolled movements. Not only that these impairments affect the movement, known by the majority as “normal”, but they can also involve a range of “behavioural, psychological, cognitive and mental impairments.” (Szente, Breipohl, 2003, 2).

Under the name of “motor disability” is gathered a wide range of conditions, such as: motor neurone diseases (e.g.: lower motor neurone disease, upper motor neurone disease, etc.), peripheral neuropathies, myopathies, inborn and acquired skeletal disorders.

The present study focuses on the motor disabilities which affect the upper limbs at different level; namely, on the disabilities which cause abnormal movements in shoulder, arms, wrists and hands.

2.1.2 Natural Computer Interface

Natural Computer Interface (NCI), commonly known as Natural User Interface (NUI) is a “user interface designed to use natural human behaviours for interacting directly with content.” The NUIs are designed according to user’s needs, the content to be handled, and the context in which it will be used. The main factors that make a Graphical User Interface (GUI) to be a NUI concern the means by which the user interacts with it — namely through human-natural behaviours such as touching, gesturing, and talking. (Joshua Blake, 2010).

2.1.3 Microsoft Kinect

Kinect is a motion sensor for Xbox 360 and Windows PCs, developed by Microsoft. This add-on allows the users to interact intuitively, and without the help of a traditional controller, making the user interface to be widely recognized as a Natural User Interface (NUI).

The Kinect system identifies individual players through face and voice recognition with the help of a depth 3D camera, which creates a skeleton image of a user, and a motion sensor that detects their movements. The speech recognition software allows the system to “understand” speech, and the gesture recognition feature enables the tracking of a user’s movements. (Rouse, 2011).

2.1.4 Hand Gesture Recognition Technologies

A hand gesture recognition technology uses a set of motion sensors incorporated into a camera to read the movements of the human body, communicating, afterwards, the gathered data to the computer. The data is used as an input to control a given application, and, through the application, other computer devices. For instance, a hand gesture recognition technology serves the physically impaired to interact with computers, which are programmed to recognize, and interpret the sign language. Furthermore, the gesture recognition technologies can be used for reading facial expressions and speech simulation (i.e.: lip reading), follow and react according to eye movements. (Yadav, 2006, 86).

2.1.5 Text-to-Speech (TTS) and Speech-to Text (STT) Synthesizers

A Text-To-Speech (TTS) synthesizer is a computer technology that reads out loud any given text, whether it is directly introduced into the computer, or scanned and submitted to an Optical Character Recognition (OCR) system. (Dutoit, 1996)

On the other hand, a Speech-to-Text synthesizer converts a vocal input into text, with the help of the speech recognition technology. There are two types of speech recognisers: Speaker dependent, whereas the system recognizes only one user's voice, and speaker independent voice recognisers, which allow multiple users to access a system using the voice input. (Beeks, Collins, 2001)

2.1.6 Google Speech API

The Google Speech-to-Text API was initially presented as part of Chrome accessibility technology and it is still under development. However the public access is available, for third-party solutions, via HTTP.

According to W3C's Web Speech API Specification final report (2012), the Web Speech API can support both server-based and client-based/embedded recognition and synthesis. It is designed to enable both brief (one-shot) speech input and continuous speech input. "The speech recognition results are provided to the web page as a list of hypotheses, along with other relevant information for each hypothesis." (Shiers, Wennborg, 2012)

All the provided tools are available only if the user has access to the speech-recognition server.

2.2 Motivation for the study

It is no secret that small and highly-precise devices, used in computer industry, are weakly adapted for the use of people with medium or strong motor dysfunctions.

The number of people suffering of various upper limb dysfunctions, injuries, chronic diseases, and impairments is extremely high. At the date of writing,

the National Center for Health Statistics in the United States reports an estimated range of 37 million to 56 million people living with a disability (up to 25% of total population of USA). Furthermore, the USA Center for Disease Control and Prevention shows that 50 million adults have self-reported, or doctor-diagnosed arthritis; these figures translating into nearly a quarter of USA's population (USA CDC, 2013).

Up to 40% of people with diagnosed arthritis notice an activity limitation, 30% of them experiencing even working limitation. According to Baker (Baker, 2009) the use of computers is significantly more difficult for individuals affected by arthritis than for the clinically healthy ones, due to the discomfort and the serious limitations caused by the disease. These issues were observed in up to 70% of the participants in the research.

According to the USA Census Bureau News (2013), there are 12 millions of people in age of 15, and older, who require the assistance of others in order to perform one or more activities of daily living, or instrumental activities of daily living (e.g.: bathing, dressing, doing housework, preparing meals ,etc.). Furthermore, an interpolation shows that there are 200 million people around the world who require assistance to perform handwork.

For there is no application which incorporates TTS, STT and gesture recognition, a team of international students have gathered in May, 2012 in TrabHCI 2013 intensive course to develop such an application, which would streamline the use of computer, by the persons with upper limb disorders.

Taking into consideration the high number of people with upper limb disorders and the lack of technologies customised for such users, the present study aims to specify the effectiveness of the developed application on a natural interface framework and determine the most efficient ways of further development, based on a purpose appropriate study case.

The present research is conducted under the supervision of rehabilitation department of Päijät-Hämeen Sosiaali- ja terveydenhuollon kuntayhtymä (PHSOTEY).

2.3 Research Question, Objectives and Scope

The study at hand is keen to understand the needs of motor impaired persons, therefore people with upper limb disorders, in the use of computers. The main aim of the study is to answer to the question: “How to design the NUI for people with upper limb disorders implementing speech recognition and hand gesture recognition?”

The present paper will explore the main and indispensable features of an application, which is meant to boost the quality of computer use, by the physically impaired.

The expected results of the study will include a brief analysis of the advantages and disadvantages of the technologies included in the STTK and the impacts they had on the computer based communication process undertaken by an upper limb disordered patient. Furthermore, the study will also result into a list of recommendations on what and how can such an application be further developed.

2.4 Research Methodology, Framework and Structure

Because the aim of the study is to explore the potentials of voice input and hand gestures in application development, and their impact on the quality of computer-based communication undertaken by upper limb disordered, the study will be conducted using the qualitative research approach.

The qualitative research studies focus on certain aspects of social life and aim to understand the experiences and attitudes of people, as well as the factors that influence them. (Patton, Cochran, 2002).

As earlier mentioned, the study at hand aims to understand what a NUI means for people with upper limb disorders, while focusing on their experiences and feelings, concerning the issue.

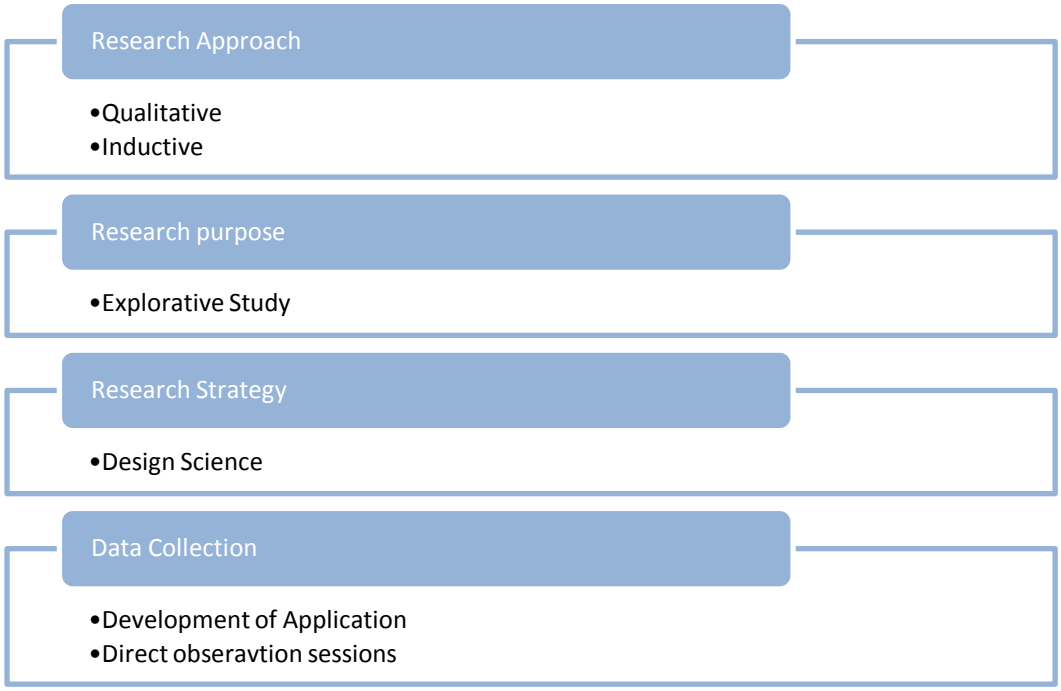

Figure 1. Research Methodology

Based on the same criteria, the study at hand is to be considered an inductive – exploratory study. As given by Trochim (2006) on the Research Methods Knowledge Base website, an inductive study is "by its very nature open-ended and exploratory".

This paper aims to explore of potentials of voice input and hand gestures in application development, based on a real application development and the feelings expressed by an upper limb disordered person who tests the application. The result of the study will be presented as an attempt to theorise on what is seen to be an NUI by upper limb disordered. Therefore this study qualifies as an inductive exploratory study.

Due to the fact that the present research will include, as an important phase, the development of an application, which would have solve a real-world problem of a person (the study case to be chosen by the researcher), the Design Science Framework will be used. "Design Science supports a pragmatic research paradigm that calls for the creation of innovative artefacts to solve real-world problems", while focusing "on the IT artefact with a high priority on relevance in the application domain". (Hevner and Chatterjee (2010, 261) citing Simons (1996)).

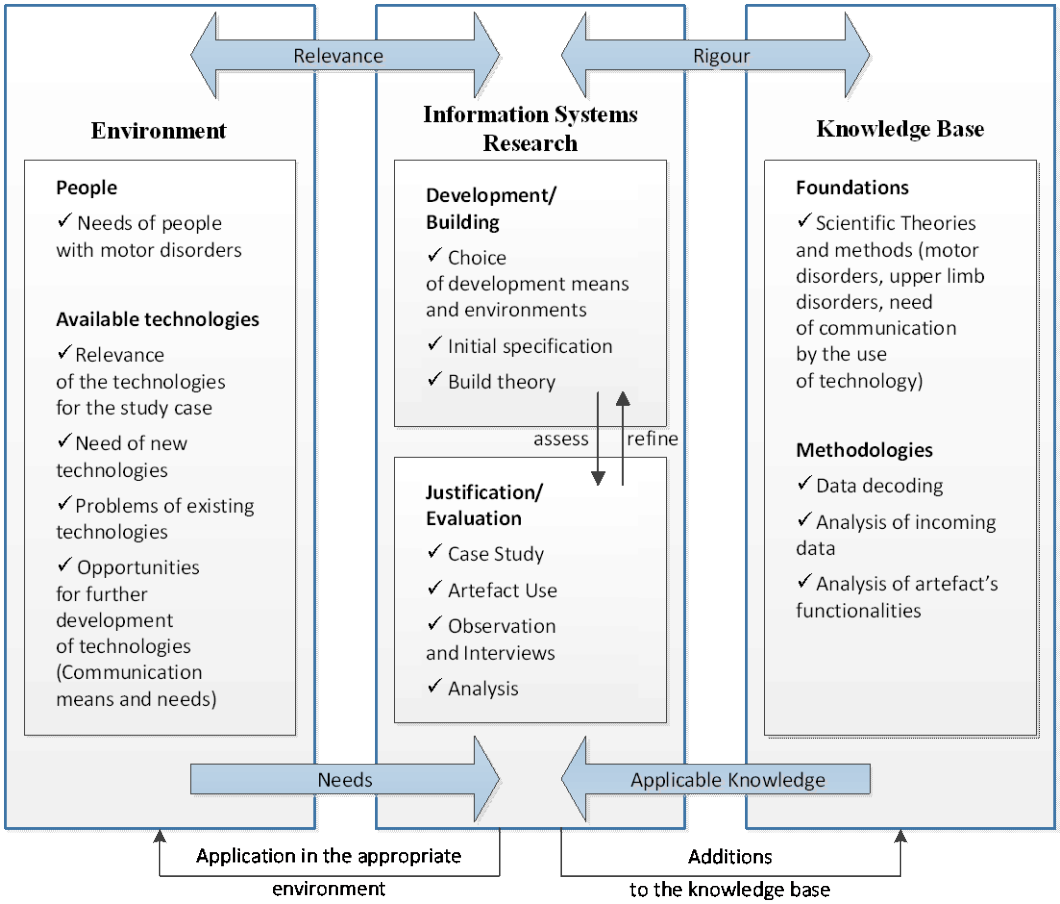

Figure 2. The Design Science Model Adaptation for the Present Research.
Original figure: Information Systems Research Framework,
Hevner, March Park, Ram, 2004, 80

Figure 2 illustrates the way the Design Science framework will be applied to the present study, while describing each phase's content. If compared against the design science principles established by Hevner et al. (2004, 80), the figure above represents a high-fidelity match; whereas it includes the three main elements of Design Science: the Environment — the problems encountered by people, the need of a new artefact, which would solve the discovered problems; the Information System Research — a new application's development, its evaluation, based on the results of an analysed study case; and the Knowledge Base, which will be created based on the application's (Information System's) development and evaluation. These phases are interdependent and represent a cycle which supports continuous development of the created application (IS / artefact) for providing optimal solutions for the initially identified problems.

Based on the principles of Design Science, the study will be structured as shown in Figure 3.

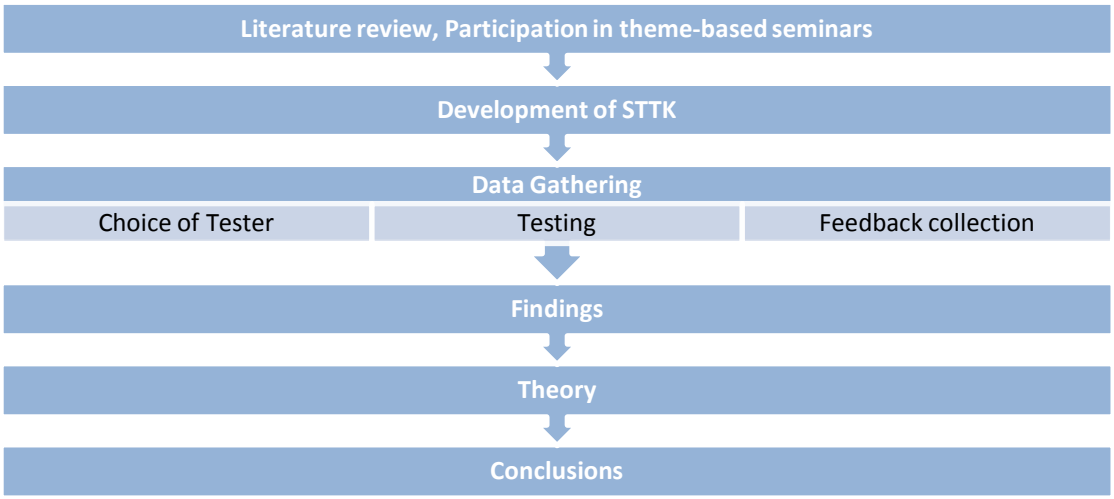

Figure 3. Research Framework

In order to identify the exiting problems and the need of a new application, which would facilitate the computer-mediated communication, the researcher will perform a well-structured literature review on motor disabilities and their impacts on HCI, as well as on the available technologies for upper limb disordered persons. Furthermore, information on available NUIs and their usability will also be gathered during the same phase. The researcher will participate in seminars, arranged as a part of the TrabHCI programme. This phase has a great role in building the understanding of the researcher on key issues for the study to be undertaken, as well as it will give the development team a solid base for the application development phase (Figure 2. Information Systems Research).

The development of the application (STTK) will take place in an organised environment, namely the TrabHCI intensive course, to be held in Rome, in July 2012. The STTK application and general information about its development is available under Chapter 6 of the present paper.

The application will be tested by a person who suffers of an upper limb disorder, who will represent the study case for the present research. The case will be chosen based on both, researcher's own reasoning and the general symptom description of "upper limb disordered", as given by Szente et al. (2003, 2). The tester will test

the application twice: once during the early phase of application's development, and second time, after the application will be modified according to the first set of provided feedback. This approach matches the Design Science Framework's structure (Figure 2), where the IS Research is strongly connected to the creation of the Knowledge base, which, again, will support the improvement of the developing IS (the application).

As the study at hand will be conducted qualitatively, the feedback will be openly provided by the tester, so that the developers could get a high diversity of development improvements. According to Saunders et al. (2007, 117–119) the gathered data should be open ended in nature, the richness of the data should be of high extent.

The data will be analysed and decoded by the researcher, in collaboration with application's development team, and it will serve as a knowledge base for application's further development.

As shown in Figure 3, after the two testing and feedback sessions, the gathered feedback will support a theory on the potentials of voice input and hand gestures in application development, and their impact on the improvement of communication possibilities of people with motor disorders. More precisely, the research aims to provide a solider base for further NUI development to be used by motor disabled individuals. Due to the fact that the application to be developed will be tested by a motor disabled person, as well as the strong technical know-how of application's developers, the researcher considers the data to be reliable and valid. This supposition is also supported by the qualitative research's generalisation logic: "To generalize is to claim that what is the case in one place or time, will be so elsewhere or in another time. Everyday social life depends on the success of actors doing just this" (Payne, Williams, 2005, 296).

2.5 Limitations of the Study

The study aims to develop and test an application to be developed in Rome, 2013, during the TrabHCI 2013 intensive course. The application is based on the existing Microsoft Kinect system, which will be overridden, in order to adapt

its usability according to the needs of people with upper limb motor disabilities. The application will also support the recognition of fluent speech and its conversion to text.

The STTK will not reach its maturity by the time the research process will come to an end, but the results of the research have as a main goal the building of a strong knowledge base for application's further development, which will not be the subject of discussion in the present study.

3 THE CURRENT SITUATION OF HCI

The chapter at hand will present the current situation of human-computer interaction, its importance for, and its impacts on people's daily life, while describing the current usability and accessibility issues computer users are facing, at the moment.

3.1 Different Aspects of Communication and Technological Implications

Scientists and scholars have made available several definitions of the "communication" concept. Due to the multifarious origin of the communication, the universal definition is highly unlikely to be found. However, most of the definers agreed that communication relates to "exchange of information". According to American Speech-Language-Hearing Association (ASHA, 1992) the communication is "any act by which one person gives to, or receives from another person, information…" Encyclopaedia Britannica also defines communication as a social behaviour, represented by "the act, or process" of expressing or exchanging information.

It is commonly considered that communication is a social behaviour undertaken by animals, or by humans — when it involves language, speech and/or writing (Encyclopaedia Britannica). Computers are "linked, connected, used by" humans and they even "interact with" their human users. But what if the difference between the human, human's digital avatar and the computer itself would be eradicated? Can the interaction with the computer be adjusted to a level of communication? If so, what are the benefits this would bring to computer users?

Nowadays, the idea of "smart something" represents an evolved version of the "smart thing" idea, and it is commonly known and accepted. The FidoNet jokes such as "PC's down, writing on TV set" would be less-likely considered today as a pun. The presence of "something", more than just a button, is not a privilege of only hi-tech devices, which are seen to be the "smart something", but of most of the devices used in daily life by most of people. Whether such modifications are necessary or not, stands upon individual

judgement. However, the trend of improving everything in a fast paced technology-driven environment, by adding new functionalities, becomes more and more common. The easiest way to do so is to attach extra chips to an existing "smart thing". This ultimately resulted in full home automation and to the "internet of things" concept (Duncan, 2014).

3.2 Is interaction with computer real communication

Articles on computer game addiction (Charlton, Danforth, 2007; Ferguson, Coulson 2011, Griffiths, 2008), not gambling, have described this phenomenon from different perspectives and in different circumstances.

This condition, which was defined by Ferguson and Coulson in 2011 as a problem of "game addiction", was found among 3% of all computer game players. The players, who are "lost" in computer games, are mostly the MMORPG players. The playground of MMORPG is based on a real life, with some alterations: common elements such as intelligence (real players), artificial intelligence (NPCs), animals, vegetation, and inanimate matter are present. Nevertheless, there are no available researches about how many non-player characters are considered by player as real personalities. Many gamers have never seen their friends in real life, and they do love certain NPCs more than any of their real-life friends. Furthermore, in the most complicated cases, addictive players have no positive feelings towards real persons, and they have moved their comfort zone exclusively onto the artificial world. (Shao-Kang et al., 2005).

The last case is a subject of multiple researches with sometimes really tough decisions about level of socialization of such player, for example, "conclusion: sadism or insecure masculinity?" (Sheard, Won, 2011, 11).

Some of the players avoid thinking of NPCs, as such, but they see no border between a human behind a computer and the computer itself? One of Sheard's (2011, 10) examinees said: "I kind of think of my teammates as my pets".

This only confirms Nardi's claim (1996, 105) of the same schemas used in cognitive psychology being able to be applied in HCI without any significant

modification. Therefore, the communication with the computer, by its origin, can be seen as being the same as the communication with any other animated objects.

Sticking to the concept of “pet”, nowadays, adopting or purchasing an electronic pet is not unusual. The i-Cybie or AIBO toys are made with programmable artificial intelligence that determines their complex behaviour. The Robodog allows its owner to create very complex responses to voice commands, it has its own emotional state and it can remember humans and humans’ preferences. Has commanding to a pet, training it and playing with it anything common with a sociopathic behaviour, such as considering your “teammates” as your pets (Sheard, 2011, 10)? The answer to questions as such stands upon individual judgement. However, it is not a secret that the interaction with domestic animals, like cats or dogs, is widely considered as real communication. Therefore, it is possible to assume that the interaction with computers sustains real communication, as the process of exchanging information (ASHA, 1992, Encyclopaedia Britannica), despite the type of the information.

3.3 The problems of the Available Software in Terms of Interaction and Communication with Computers

It is no secret that computers are prevalently personalised. Humanity overpassed the phase when computers were yet another technical thing, on the same level as the washing-machines, irons, stereo systems or any similar appliances. The global trend of interaction between humans and electronic devices is not anymore limited to pushing a button (Want, Schilit, 2012, 25). Interface designers and usability testers appeared in global list of IT specialists and took strong positions in any software or hardware development companies, by implementing “responsive, interactive, human-friendly” design of any devices of daily use (Google.Trends.2010, Keywords: UX Design, UI Developer, Responsive Design).

While life of the majority seems to become easier, due to the implementation of a wide range of online services and transferring business to e-commerce, people with limited access to modern technologies are left apart and require more and more help.

According to a report from the Department of Justice on Enforcement of the Americans with Disabilities (2011), 36% of people with severe disability, aged between 15 and 64, used a computer, and 29 percent used the Internet at home. On the opposite side stand the individuals with a non-severe disability or with no disability, who had considerably better computer access; 60.7 % using a computer and 50.9 % using the Internet at home.

The status of interactivity gives the developers a loophole regarding the technically complex accessibility features in software. The disclaimer ”don’t like — don’t use” helps to produce efficient, flexible, feature reach applications just by skipping the time-consuming and resource-intensive implementation of accessibility features. The number of, for example, fully accessible commercial web sites is about four times smaller than their total amount available on the Internet (Goldie, 2006, 13). Assuming that the “interaction with the computer” is possible, rather than the “communication” with it, the existing software and hardware solutions would never fully satisfy people’s communication needs. The insufficiency of communication possibilities, due to the poor implementation of accessibility, is a global issue that could be solved only by changing the attitude towards it (Larman, Basili, 2003, 50).

What if any act of usage of software would be considered as an act of communication? This would eventually lead to the necessity of providing equal communication rights to all users, by offering access to out-of-the-box full functionality. ASHA stipulates, among other communication enhancement tenets, the necessity of providing communication tools that can support, and be used by, individuals with disabilities. These tools should foster usual and computer-mediated communication, giving the disabled an opportunity to initiate, accept and respond to communication acts.

Existing best practices and approved methodologies are ready for adoption in the software development industry. The developers would obtain ready-made frameworks and assessment rules, serving for the building of next generation accessible software. (ADA, Standards for Accessible Design, 2010).

As a consequence, a wider range of people would be involved into the virtual world, this way extending communication abilities for at least 200 million people world-wide (UN enable, 2011; Brault M., 2012;). The possibility to enable, maintain and terminate interactive communication between human and computer, in the era of global computerization, could materialise, and maybe prevail upon the possibility to communicate with other humans.

3.4 Brief Look at Current Situation of AI and NUI

In real life people communicate continuously, and even involuntarily — human bodies are constantly sending signals to outer world, and receiving them from others (Knapp, Hall, Horgan, 2012, 4). This is so natural, that computer users tend to try to transfer the same behaviour into virtual environment. Guye-Vuillème et al. (1999, 1) stressed that during virtual communication with virtual characters people are trying to interpret even missing face expression, pose, hand movements, and other signs of body language. Thus, the other communicator seems to be an artificial intelligence-based entity, built with both verbal and non-verbal communication systems.

Currently, there is no real, publicly available, implementation of AI that could not be recognized by its user to be artificial (e.g.: Chatterbox Challenge). There is still a great progress to be made on the field of human language transcription onto a computer-recognizable format, and vice versa. However, it is widely known that technology world is a very fast paced environment (Biery, 2013; Inc., 2013; theonlineinvestor.com, 2014)

At the moment, NaSent recognises 90% of emotion through the analysis of a written text. The system can recognise emotions, such as ambiguities and sarcasm. (Abate, 2013) ; FaceReader™ recognises up to 90% of human emotion, based on facial expressions (Noldus Information Technology), while Google attempted to develop a system that would recognize fluent speech.

At the moment of writing, the search in Google by phrase match “AI fear” returns approximately 9 million results, “Afraid of Artificial Intelligence” — 1.4 million, but “AI development” is about 210 millions, “AI ethics” — 10,2 million.

Interactive intelligence intruded into human life, and after tens of years, it started to be accepted as being an obvious part of daily life.

Journalists (BBC news, 2006; Henderson, 2007) took the matter up to another level, by comparing AI rights to animal rights, or even to human rights. Such discussions are based on the ethics which suggest that the computer cannot anymore be considered to be "just a machine", but they can even apply for the Loebner Prize: "each year an annual cash prize and a bronze medal is awarded to the most human-like computer" (loebner.net). As the paradigm of communication regarding the interaction with the computer becomes widely accepted, serious attempts to include ethics into a concept are just a matter of time.

Despite that AI is not yet available, the current evolution state in computer technologies allows computers to recognize their users (Bronstein A., Bronstein M., Kimmel, 2007), understand users' emotional condition (Raudoniset et al., 2013; Khanna, Sasikumar, 2013), retrieve (remember and recall) information relevant to the topic relevant for the specific moment (i.e.: Mitsuku), share topic-related information with the user (i.e.: News at Seven), and infinitely repeat this cycle.

Taking into consideration all the above, the natural interface represents a set of interactive technologies, similar to the natural ones, used in inter-human communication. The natural human-computer interaction can be seen as the opportunity, offered by a technology, to use all the habitual communication abilities in their original, intuitive ways, as used in interpersonal communication.

The purpose of building a natural interface is to provide equal opportunities to people during their communication with computers, regardless their motor disabilities.

4 INITIAL REQUIREMENTS OF STTK'S DEVELOPMENT

The present chapter describes STTK's purpose and initial requirements, as considered by its developers.

4.1 STTK's Purpose

The core application is based on the natural interface framework and it is written on C#. STTK is as a test tool which was used to discover the effectiveness of the chosen approach in the daily life of people with motor disorders; the conclusions drawn serving for determining the most efficient ways of further development. The external modules and plug-ins are neither described nor mentioned in the paper at hand, as they are not relevant for the actual topic.

The natural interface framework was developed with the main aim to improve the accessibility level of the existing software, in accordance to the needs of people with upper limb disorders. This framework can be used to develop a core application implementing text-to-speech (TTS) and speech-to-text (SST) technologies, hand gesture recognition, programmable interfaces to interact with third-party software, and interfaces for creating custom plug-ins[1].

4.2 Approach

The chosen strategy to build a natural interface is based on the principle of modular structure and accessibility, regardless the severity of disorder. There is no universal solution to bring such functionality to all the on-line services and offline applications used in a daily life, but the end user should be able to gain access to a requested web resources or desktop application, and perform their routine activities.

1. The application requires a speech-recognition system to support voice commands and voice input. At the moment, the most effective speech-to-

[1] A piece of software which enhances another software application and usually cannot be run independently.

text engine is provided by Google and it is accessible remotely via HTTP protocol. Because the technology is available as an on-line service, the resulting application would require Internet connection.

2. The speech recognition system requires loud and clear speech. The tests between built-in Microsoft Kinect microphone and external headset showed significantly higher recognition level with usage of the latter. The built-it microphone is sensitive to ambient noise and the quality of the record is decreasing, as the distance increases.
3. The level of the ambient noise is defined as separate factor affecting speech recognition quality.
4. The chosen hand gesture recognition system requires the Microsoft Kinect sensor. There are different versions of the sensors, according to their precision level and available functionality. The sensor should be able to operate when the user sits and has only one hand available, which movement may be affected by tremor, therefore amplitude of hand movement being limited. To fulfil those requirements, the desktop version of Kinect sensor is preferred. The resulting software must include settings for sensitivity level adjustment and pointer speed multiplication.
5. The distance from the user to a screen is not strictly limited, so all elements of the final application must have an adjustable size and support special colour schemes (high contrast, colour blindness).
6. Because of its modular origin, the core should allow the insertion of external plug-ins for additional functionality implementation; therefore the application must have available API for plug-in development.

The internationalization should be implemented for all textual elements. However, the implementation of such a feature is not relevant for the present research’s scope.

5 THE STUDY CASE

This chapter presents the study case for the current research, describing tester's disability health state, which has an impact on the feedback development stages stated in the subchapter 2.4.

5.1 Acknowledgement

Due to ethical reasons, the name of the tester will not be disclosed, but only the initial of their first name, "J".

5.2 Tester's background

J. voluntarily became STTK's tester due to their interested in an assistive technology solution which would improve the quality of their experience in the utilisation of computer on a daily basis.

At the age of 50, J. had suffered of a complex regional pain syndrome (CRPS) for over 3 years. The trauma affects the upper part of their spine and it is a consequence of an accident at work. The CPRS, J. suffers of, causes severe pain in the upper limb as well as it reduces the mobility in shoulders and arm. In order to ease the pain, a spinal cord stimulator was implanted. However, the incorrect implantation resulted in J's high intolerance to electromagnetic fields. This intolerance is manifested as a strong pain in neck, shoulders and arms during J's proximity to any EM field source (i.e.: halogen light sources, TV and computer screens, microwave oven and others).

J's arms are partly functional and disallow them to use the keyboard and mouse. Basic interaction with the computer is possible by using in one arm for handling a wireless mouse; however, the effective control over mouse is not fully possible due to the limited mobility in their shoulder joint. The use of a keyboard is the most complicated task, because it requires a very precise finger movements, and J. is only able to use effectively enough one hand; therefore J. must toggle between a mouse and a keyboard. It is understandable that the speed and comfort level of one-hand typing are poor.

Taking into consideration the encumbrances in the use of computers, brought to J. by their condition, an application which would enable them to control the computer by voice commands or hand gestures performed from a certain distance seemed to be an appropriate solution.

5.3 STTK Application

At the moment of writing, the application has a working prototype built for further investigation and development of the effectiveness. It concerns modern communication technologies, with the aim of improving the communication possibilities of people with motor disorders.

The development of the STTK started at TrabHCI 2013 intensive course in Rome, Italy and was continued by the author of the study, under the supervision of Lahti University of Applied Sciences, after the first feedback session with the tester. There were two original concepts developed during TrabHCI, by a team of international students, during a two-week code jam: the override of the existing Kinect hand gesture system, and the implementation of fluent speech recognition and its conversion to text. Overriding the current Kinect hand gesture system had the aim of adapting its functionalities to the needs of people with upper limb motor disorders.

The outcomes of TrabHCI 2013 were two prototypes: MailMe and ChatOn. MailMe is a prototype which is represented by the implementation of a new accessibility pattern of hand gestures using Microsoft Kinect system, whereas ChatOn is a prototype of the implementation of the front-end for speech-to-text recognition module.

The logical continuation of the above mentioned prototype, having the common goal of representing a natural computer interface, was the STTK. The STTK can be seen as a synergy of applications which supports both hand gestures and voice input, when controlled by upper limb disordered users.

6 TESTING AND DEVELOPMENT OF STTK

The present chapter will describe the evolution of the ChatOn and MailMe into STTK. STTK's development is based on tester's feedback, which will also be introduced to the reader through this chapter. The information gathered during the feedback sessions served as a base for conclusion drawing.

6.1 Acknowledgements

The initial version of the application, as well as its later version, will be referred to, as to STTK.

In order to gather appropriate, reliable information for STTK's development there were two testing phases, resulting into two feedback sessions. Each testing phase lasted one week, during which the tester had the time to freely try out STTK's features and their functionalities, as well as whether they suit their daily life needs, in terms of computer use.

6.2 STTK after TrabHCI intensive course

After the TrabHCI intensive course, STTK had no real GUI, but its interface which allowed the access to Google Speech API (console application).
The "Circle-Like" hand gesture model, as presented in STTK's final version (Figure 9), was only partly functional.

The class view of the application, at that specific moment is shown in Figure 4

Due to the lack of GUI and its low functionality, this version of STTK could not be sent for testing. However, the application was further developed, by the author of the present study, under the supervision of Lahti University
of Applied Sciences.

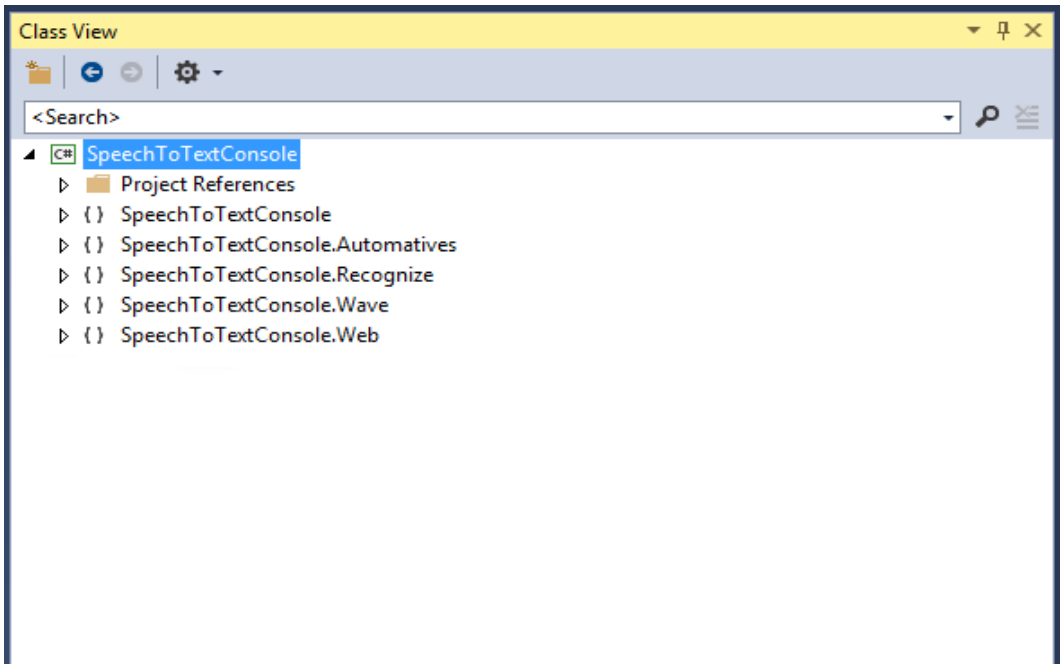

Figure 4. STTK's Class View, after the TrabHCI intensive course

6.3 STTK at First Phase of Testing

At the first testing phase, STTK had undergone significant changes. In order to have the application available for testing, the main development priority was the GUI (Figure 5), which in the previous stage did not exist. The initial GUI was developed and adapted to the Circle-Like hand gesture model (Figure 9), however the whole GUI was not adopted for the usage at long distance between the screen and user. The application was only available in English, and the output text size was extremely small, which made it hard to be read from a longer distance.

In terms of functionality, the application recognised text, but did not format it, and the voice recognition language could not be changed by the user, because of the time constraints which lead to a “hard coding” type of approach. Furthermore, the voice recognition module was functional, but it was running in the same thread as the application window — this eventually caused the interface to be irresponsive during the voice capturing and voice recognition tasks.

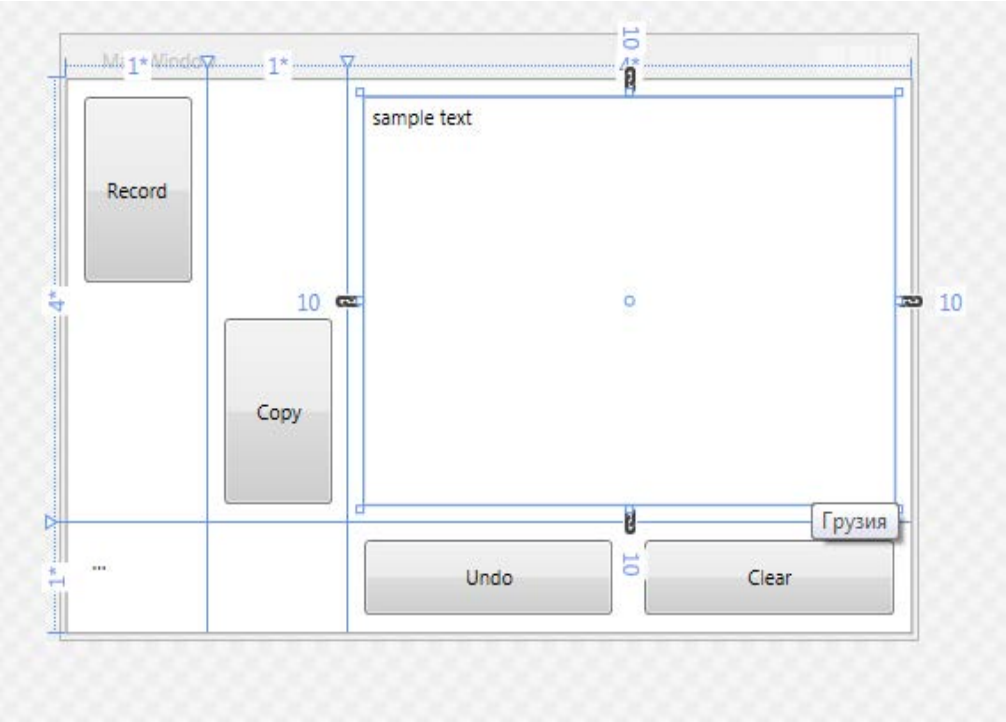

Figure 5. STTK's GUI at the First Testing Phase

Although, as shown in Figure 5, the Undo/Redo button existed, its functionality was not implemented. However, the application was able to inherit the input model from the original “Circle-Like” hand gesture model (Figure 9), but the application captured only the hand gestures and did not allow the Kinect-independent utilization of a mouse.

While one may argue that such a version of the application should never be sent for testing, the author of the study chose a development approach which would first consider users’ needs and discover their needs during application’s development (Software Development with Users).

6.3.1 Tester’s Feedback on Initial Version of STTK and Steps Taken for the Improvement of the Application

J., the tester of STTK received the above described version of the application and had it in user for a week. As one of J’s arms is less affected by their disability, they rather preferred the use of a mouse, than the use on Kinect, the on-screen pointer control needed to be reworked so that it would provide simultaneous access for the attached Kinect and mouse.

The voice input and recognition caused the UI to be irresponsive, at times. The issue was solved on the programming level: split all voice recognition related modules into separate threads; and added finite state machine to control the voice recognition process.

The UI was uncomfortable, when used on a distant screen. The location, colours, and sizes of all the available control elements of the application, as well as all text string were altered, so that they would clearly be visible from long distance between the user and the screen.

The punctuation during the voice input did not fit J's expectations and it required significant improvements. Because of task's high complexity and importance, the following modifications were agreed on: first letter capitalization, a period and a whitespace after each sentence, digits to be represented by digits, not by their textual transcription, and punctuation interface while dictating.

6.4 STTK at Second Phase of Testing

After the first feedback session and analysis of the next steps to be taken, the STTK's GUI (Figure 6) was adapted for dual input (mouse and Kinect) and for larger screens.

The application benefited of native colors and the internationalization (allowing the user to change the language in use) was also implemented.

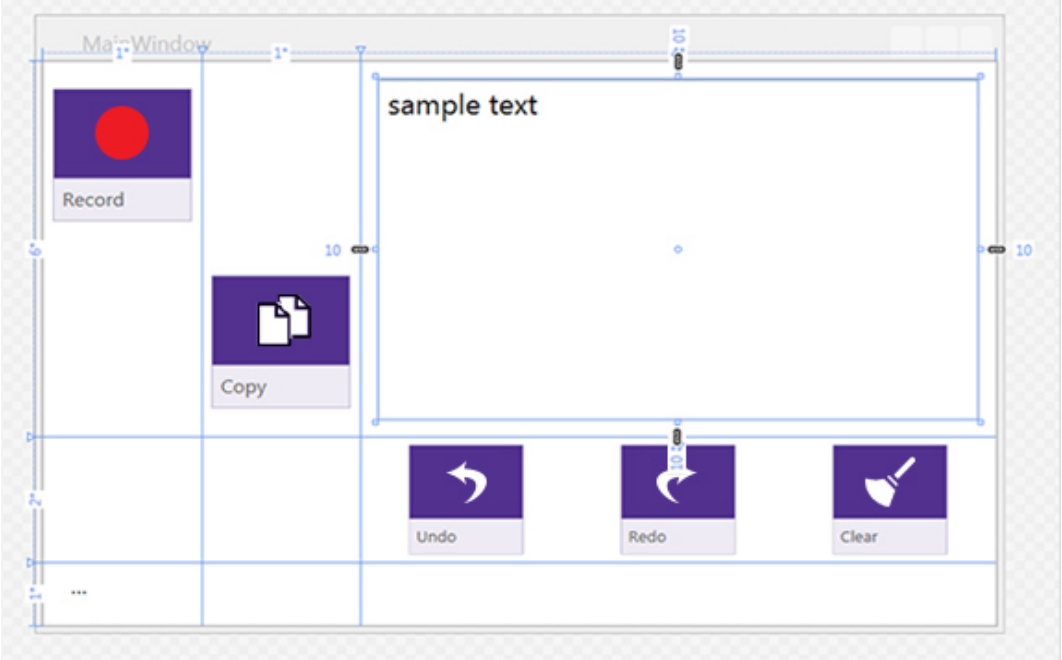

Figure 6. STTK's GUI at the Second Testing Phase

In order to achieve the modifications agreed on, the code was reorganized into a modular structure (Figure 7), prevented locks and starvation of processes during long-run operations.

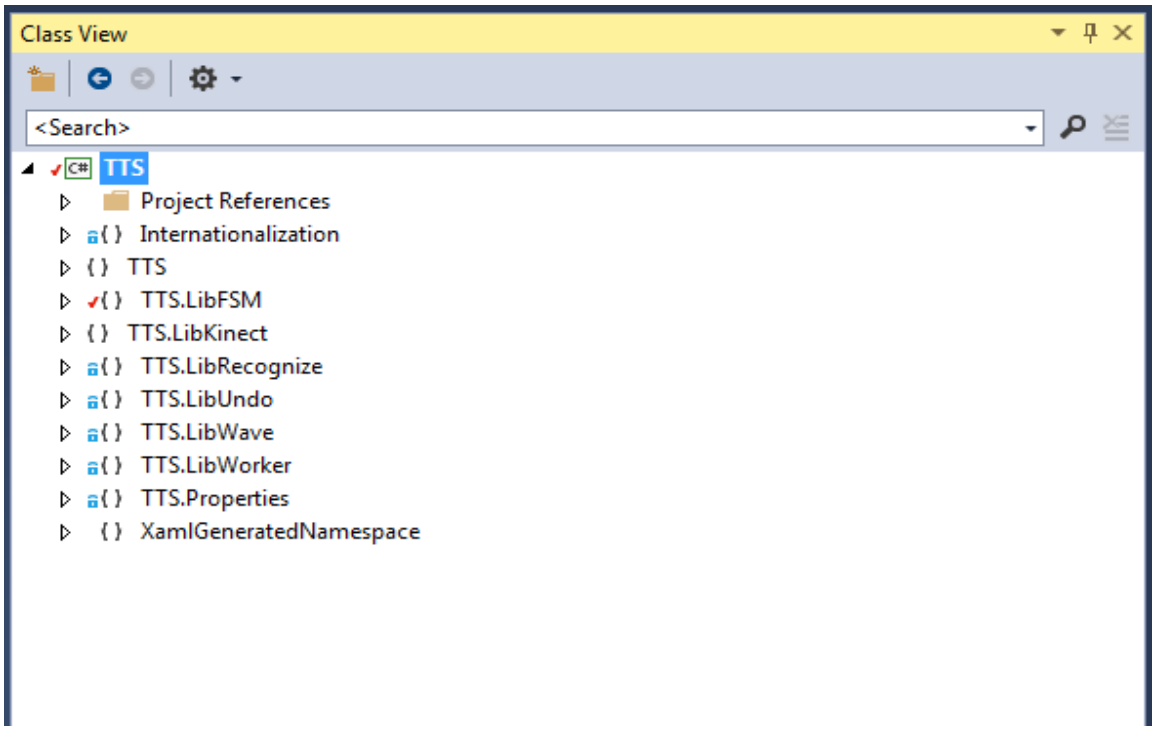

Figure 7. STTK's Class View, at the Second Testing Phase

The Finite State Machine was also implemented, in order to control the whole voice recognition process. STTK supported asynchronous execution of module, and the basic text formatting and formatting rules saving, as a last phase of text recognition, were possible.

6.4.1 Tester's Feedback on the Second Version of STTK

In order to receive feedback on the second version of STTK, representing an improved version of the above described application, J. tested it, at their own convenience, for a week.

J. was pleased with the voice input system, which significantly improved their computer-use experience while writing short emails, chatting in social networks, etc. In their point of view, the GUI did not anymore cause so much discomfort, while using the application from a greater distance, but it could even be used effectively. Furthermore, J. experienced a reduction of the total amount time in keyboard use. The combination of mouse use and voice input was the most positive change in J's interaction with the computer.

STTK was also reported to be responsive while being used simultaneously with another applications and to not cause the stress of constantly switching between them.

However, at its intermediary development stage, but final stage for the study at hand, STTK's punctuation is very basic; therefore complex sentences are difficult to dictate. Furthermore the application does not recognise first-, or surnames, names of places, or trade names.

Although switching between the main and the assistive programs did not cause difficulties, having an application that would start recording, when asked by voice command, and perform the speech recognition, without taking a focus on itself, would increase the user's experience's quality.

7 APPLICATION DESIGN. USER INTERFACE

This chapter presents the UI of the STTK at its intermediary development phase, which is considered to be the final stage, as far as this study is concerned.

7.1 Acknowledgement

As the communication has verbal and non-verbal components, both should be implemented during the development of the natural interface. The non-verbal component includes body language and face emotions. The verbal component is based on speech.

At the moment of writing, the body movement could be successfully captured and recognized by computers due to the hand gesture recognition systems provided by Microsoft Kinect sensor. The emotion recognition systems are still in development, and currently there are no public tools available for an efficient implementation.

There are two solutions for verbal communication available: fluent speech recognition, which is available as online service provided by Google; and speech recognition, which based on pre-made dictionary as part of Microsoft .Net framework.

The hand gesture represents the way to remotely interact with the active elements of the application, by moving a hand, or both, in different directions. The hand gestures, this study focuses on, are the following:

- The “selection gesture” — the gesture of executing any programmed activities by the use of the application (i.e.: choosing an active element, confirming the choice, performing a virtual click on elements, etc.)
- The “rest gesture” — the voluntary or involuntary action of resting the hand. Commonly “the rest gesture” is the vertical hand movement downwards, for example, onto any supportive surfaces, hip or lap.

The speech-to-text system is a programmed way of capturing fluent speech, verbal commands, and converting them into machine-recognizable format.

7.2 Hand gesture control

There are two main approaches which are used in the implementation of the hand gesture control of application: The unified set of supported gestures provided by Microsoft Kinect Toolkit, where the action of confirmation is attached to the virtual click, and it is implemented as a frontal movement of a working hand towards the chosen element; and the custom “Circle-Like” hand gesture system, which was promoted during TrabHCI 2013 brain storm, with the aim of eradicating the high complexity of Kinect default gestures for people with motor disorders.

7.2.1 Kinect default selection gesture system

The Kinect default selection, graphically presented in Figure 8, is bound to the hand movement towards the Kinect sensor.

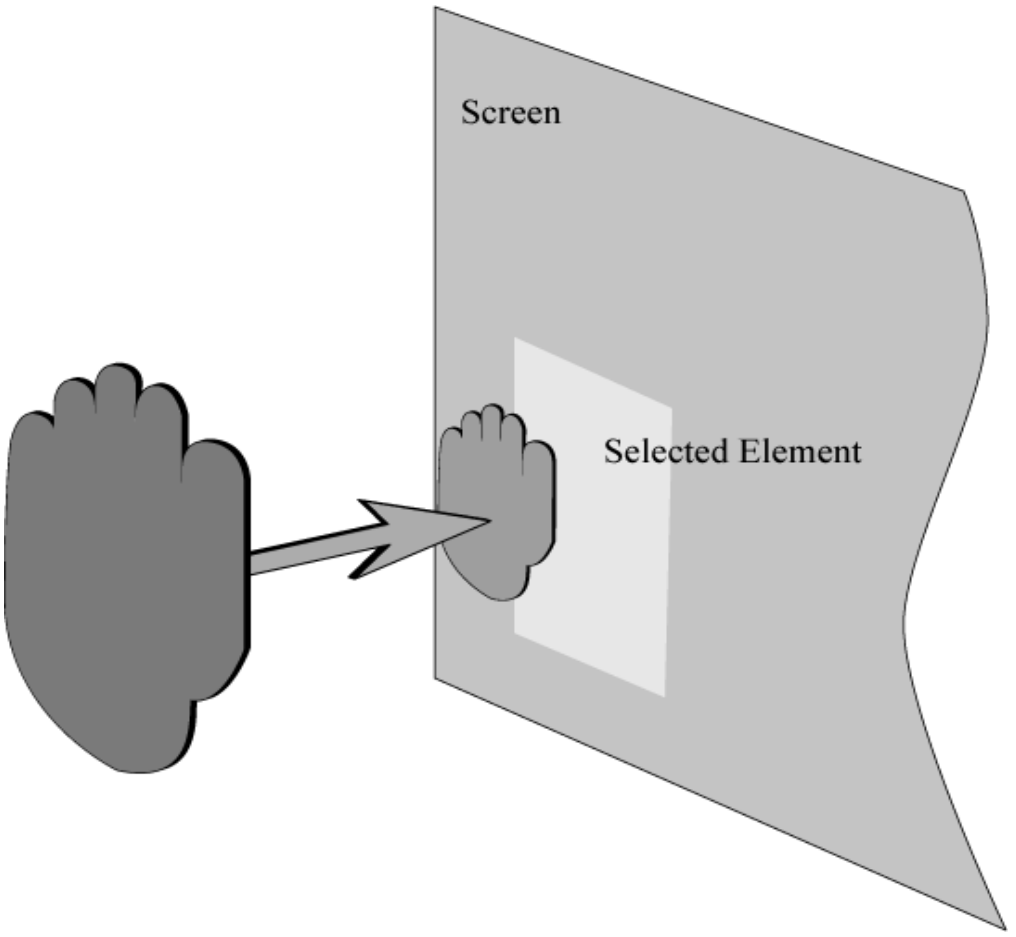

Figure 8. Kinect’s Default Selection Gesture

There are two main advantages provided by Kinect's implementation:
The unification with, and similarity to all the other Kinect-based applications;
and the ready-made dll's[2] available to be used the application.

On the other hand, the complexity of the hand movement depends on the severity of the tremor. While Kinect requires a movement towards camera, not towards a projection of an object on the screen, the high complexity of such movements for people with motor disabilities and atonic dystonia can surely be seen as a disadvantage.

7.2.2 "Circle-Like" selection gesture system

The "Circle-Like" hand gesture selection method, presented in Figure 9, was developed as the solution for the conditional inaccessibility of Kinect's original selection gesture.

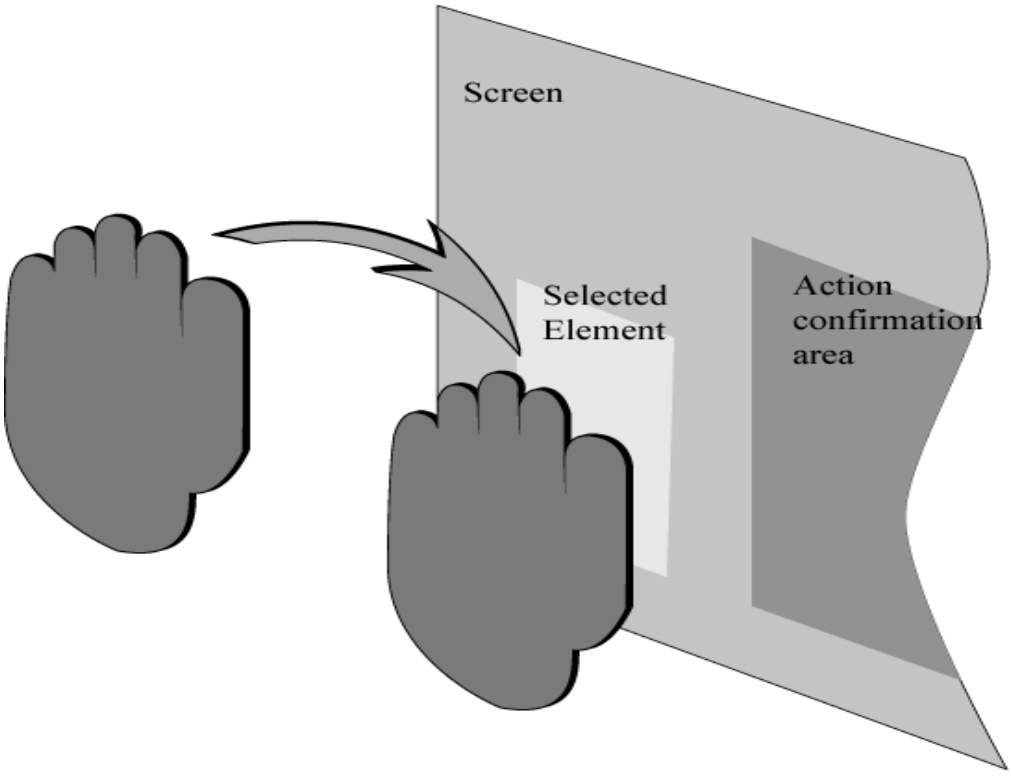

Figure 9. The "Circle-Like" Selection Gesture

[2] Dynamic-link library, is Microsoft's implementation of the shared library concept in the Microsoft Windows operating systems.

This action includes the same steps for the selection and execution of an action, but the movement curve is modified in order to exclude the frontal component and to preserve “rest action”.

During a “Circle-Like” gesture, the selection of any on-screen objects occurs once the virtual pointer is positioned over an element. The selected object can be released by executing the “Rest Gesture” (vertical downwards movement), by selecting another object, or by placing the selected object to action confirmation area.

The user interface of the “Circle-Like” hand gesture is created based on three main principles: the arrangement of active elements is in one row; the confirmation area is required and placed aside of active elements; and the cancellation action is a hand movement towards the bottom of the screen (same as the “Rest Gesture”).

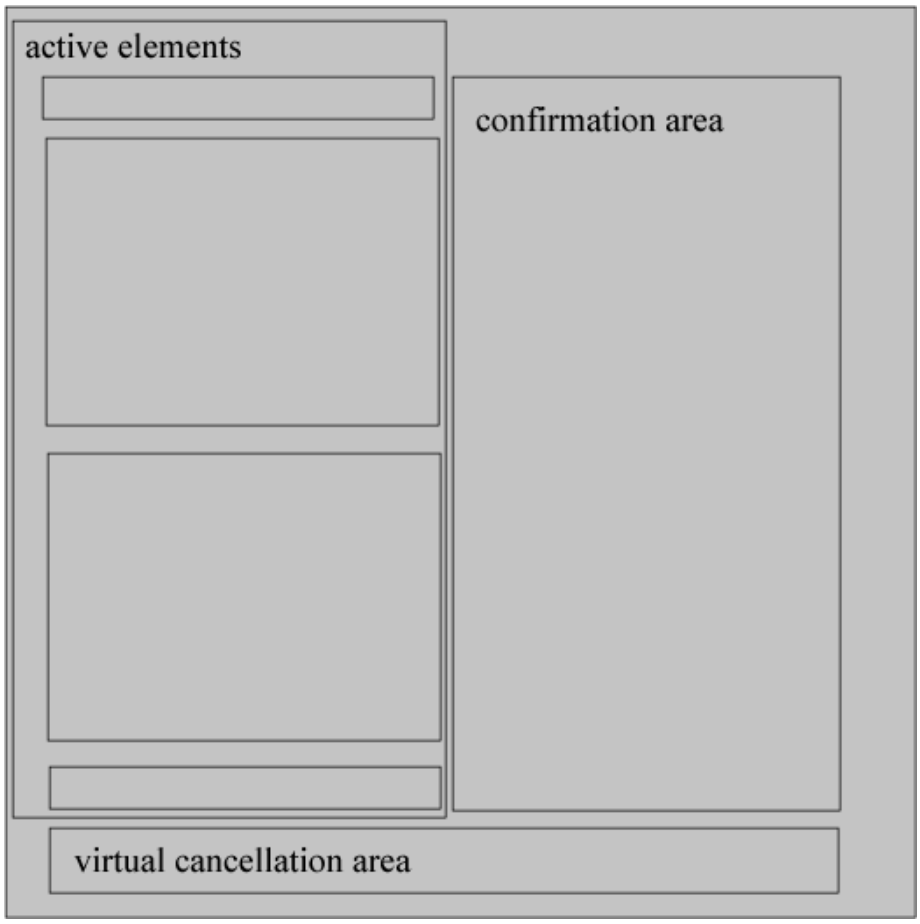

Figure 10. The UI for the “Circle-Like” Gesture – Active Elements are in a Column

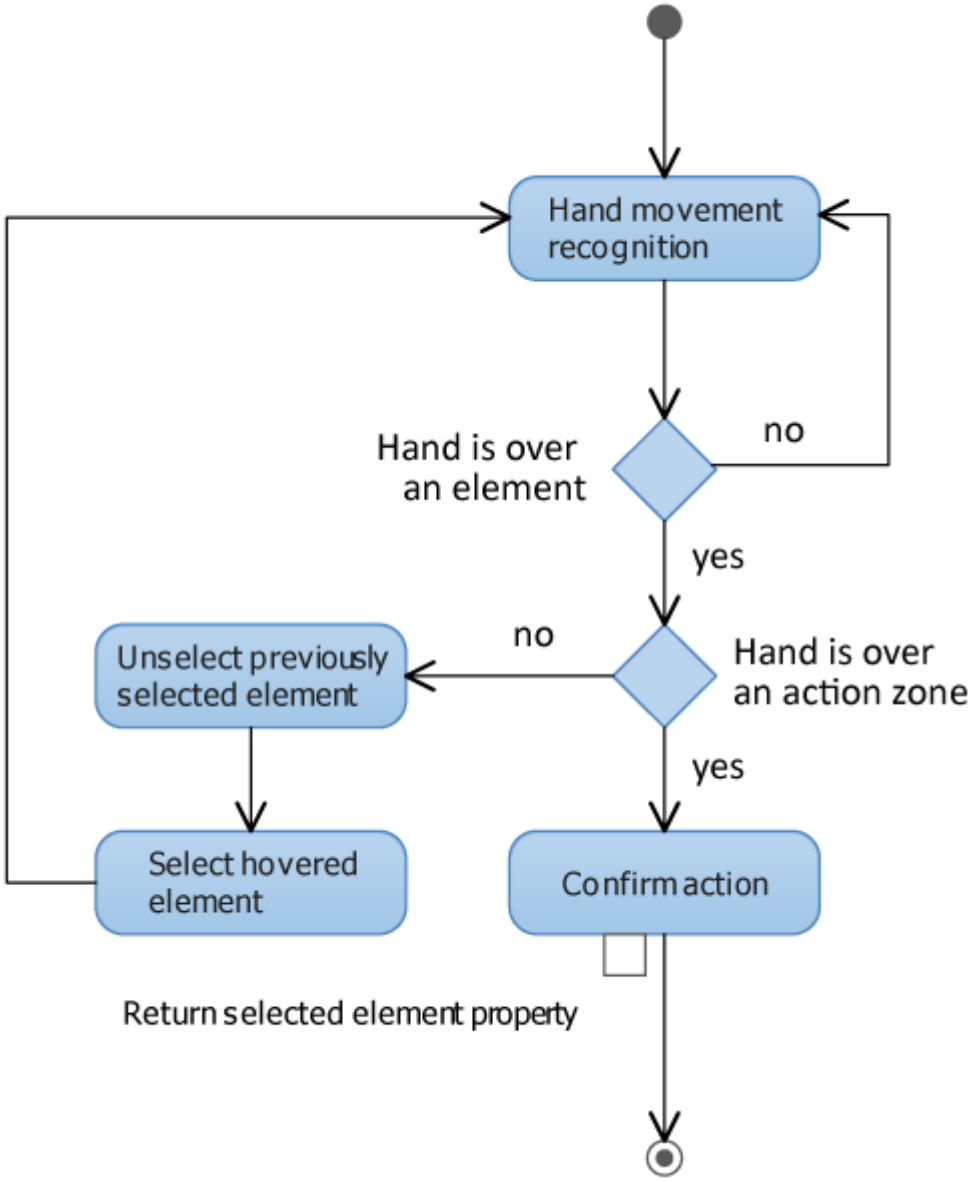

Figure 11. The “Circle-Like” Gesture Activity Diagram

The “Circle-Like” gesture offers the possibility to perform even to people with complex upper limb movement constraints. This gesture requires significantly less movement precision than Kinect’s default selection gesture system, namely the independence from the frontal component makes it easy to use also by people with muscle atonic syndromes, while the “Rest gesture” is preserved.

Along with the above mentioned advantages, the “Circle-Like” gesture comes with two main disadvantages. Namely, the UI requires one relatively large

area for action confirmation, as well as it requires to separately be implemented as a non-supported by default Kinect Software Development Kit.

The variation of the user interface, adopted for the “Circle-Like” hand gesture, could be the screen layout, as shown in Figure 12, where the action area is placed above a list of active elements.

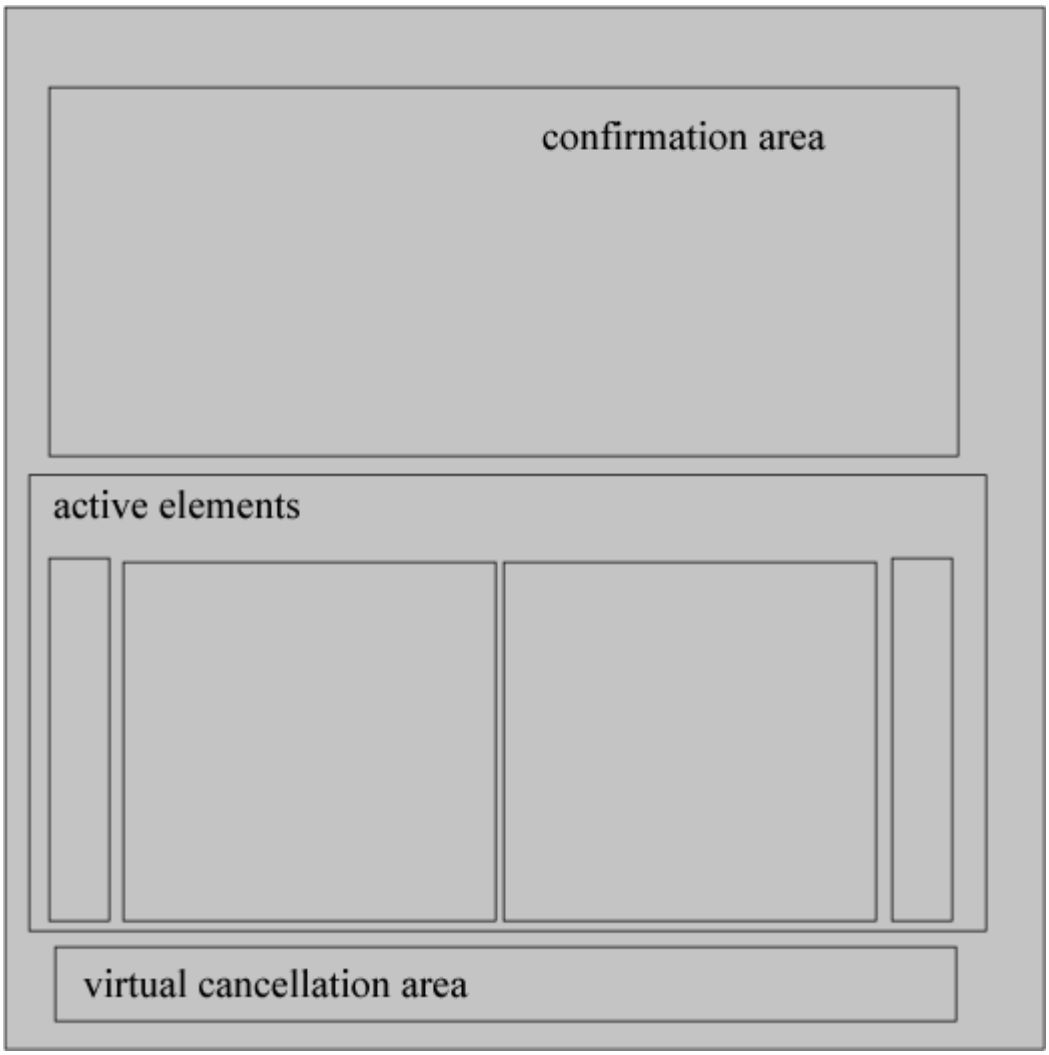

Figure 12. UI for the “Circle”-Like Gesture – Active Elements are in a Row

This style of user interface should be available in the application settings window and should offer the possibility of selecting it, according to the available type of hand movement.

The Windows Presentation Foundation (WPF) technology allows sufficient flexibility of elements' placement, by dynamic rearrangement of nodes in document's XML markup.

7.2.3 "Hold-and-Drag" Problem

The default gesture set, provided by Microsoft and used with Kinect sensor, allows the "Hold-and-Drag" action to operate with the objects on a screen. The action sequence is the following: to hover an element by a hand pointer, to clench the fist, to move an object, to open the fist for releasing an object. Despite the correlation with the natural gestures, there are two major problematic issues that would need to be tackled in this action sequence: the necessity of steadily holding the hand over an element while clenching the fist, and the necessity to clenching the fist.

In case of various dysfunctions of the upper limb, one or both of the above mentioned steps could either be too complicated to perform, or completely inaccessible. The large number of people with distal arthritis has difficulties in picking up any objects by the use of hand, or this action produces discomfort, or even pain. On the other hand, there are multiple diseases accompanied by atonic or hypertonic muscle conditions, which cause the clenching of an element to be impossible. All the conditions complicated by tremor make the need of steadily holding a hand unfulfillable.

The screen layout adopted for "Circle-Like" hand gesture tackled the "Hold-and-Drag" problem by reducing the required accuracy of hand movements. The selection of elements could be done and undone after the natural movement of a hand across the virtual screen, excluding frontal dimension (similar to moving mouse pointer over screen). Positioning the elements in a single row, as shown in Figure 12, reduces the chance of selecting wrong items, while placing the confirmation area aside of the element row prevents the unexpected execution of a selected action.

7.3 Voice Control and Voice Input

There are two major voice input systems that could simultaneously be implemented in an application, and used, either independently or as two complementary systems: .Net Speech Recognition System and the Google Speech Recognition System.

The .Net Speech-To-Text system has one major limitation, which can be also used as an advantage. The .Net 3.0 and higher System.Speech.Recognition class provides speech recognition tools, used in a pre-defined language environment and by using pre-defined finite grammar dictionary.

It benefits of high recognition quality of phrases from grammar file and it can work locally, without internet connection. However, the .Net 3.0 and higher System.Speech.Recognition class requires a separate grammar file, which would contain all phrases that should be recognized, while the size of the grammar file is limited by the performance of the local system. Furthermore it has no functionality for performing the recognition of fluent speech.

At the moment of writing, the speech recognition system offered by Google was at an undefined stage and it only had a public implementation for Chrome browsers (Shires, Wennborg, 2012).

The unofficial speech recognition system used in Google Translate works by analysing the flac-encoded[3] audio file. The output is given in JSON[4] format, including the recognized phrase and its accuracy level.

The utilisation of the unofficial Speech API is not regulated so far, as commented by Glen Shires (2013), member of Google Research, Speech Recognition team — “The speech-API URL (http://www.google.com/speech-api/v1/recognize?) is not an official API, and it's totally unsupported, and it may disappear at any time.”

---

[3] FLAC stands for Free Lossless Audio Codec, an lossless audio format means that audio is compressed in FLAC without any loss in quality.
[4] JSON, or JavaScript Object Notation, is a text-based open standard designed for human-readable data interchange.

However, the current progress in speech recognition could soon make available the public speech API for non-browsers, and the functionality of both, browser-based and unofficial APIs, allows an evaluation of the quality and functionality of speech recognition.

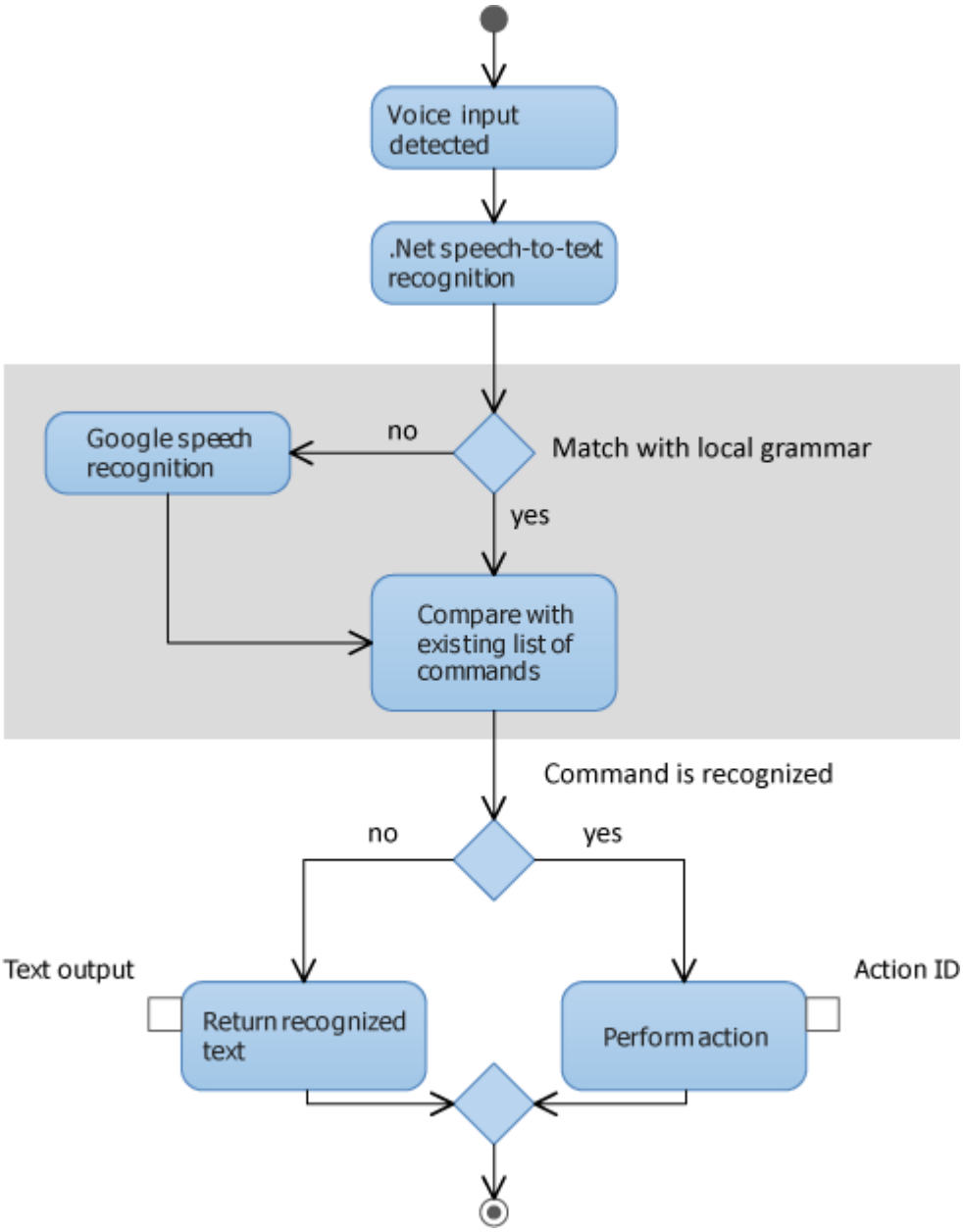

Figure 13. The Dual Speech Recognition Module Activity Diagram

The main advantages of a combined implementation of browser-based and unofficial APIs regard the possibilities of recognizing fluent speech without any implementation of extra dictionaries. Furthermore the centralized remote recognition system allows the improvements of recognition, without updating

the local software, while the recognition system is not using any additional local resources; therefore the effectiveness of speech recognition would not affect the local system's performance.

However, such an implementation would require internet connection, meaning that it could be used locally. Furthermore, another disadvantage is represented the limited and not specified functionality of such a combination.

Considering possible the availability of both abovementioned systems, an implementation of a recognition system which would use the advantages of each method, as shown in Figure 13, is doable.

In addition to all the listed advantages of the both speech-to-text recognition systems, this combination streamlines the implementation of a "security mode" which would prevent the user input from being transferred to third-parties. The security mode would serve the user's privacy and security when using the voice control to input personal information, passwords, and other private information.

7.3.1 Punctuation Issue

When using voice input, punctuation marks should also be inserted. Currently, the only solution is the pronunciation of all the punctuation marks as words, and the manual correction of the input. Various technologies provide adaptive voice input, which "learns" the style of pronunciation, and it is supposed to recognize and insert punctuation marks more efficiently. Unfortunately, the use of such adaptation can cause troubles to its users: "Well, I went through adaptation and now it works MUCH worse for me. Is there a way to undo it?" (J., STTK's tester).

Furthermore, the semantics involved by words associated with punctuation marks, such as "period", which in many languages means "That's it!", "End of story!" "This is how it is", prevent the adaptive voice input from properly serving its users' purpose.

So far, no absolute solution for this problem was developed. However, the dual input may provide an intermediate solution, by the manual insertion of punctuation marks, during the speech. A possible use case is graphically shown in Figure 14.

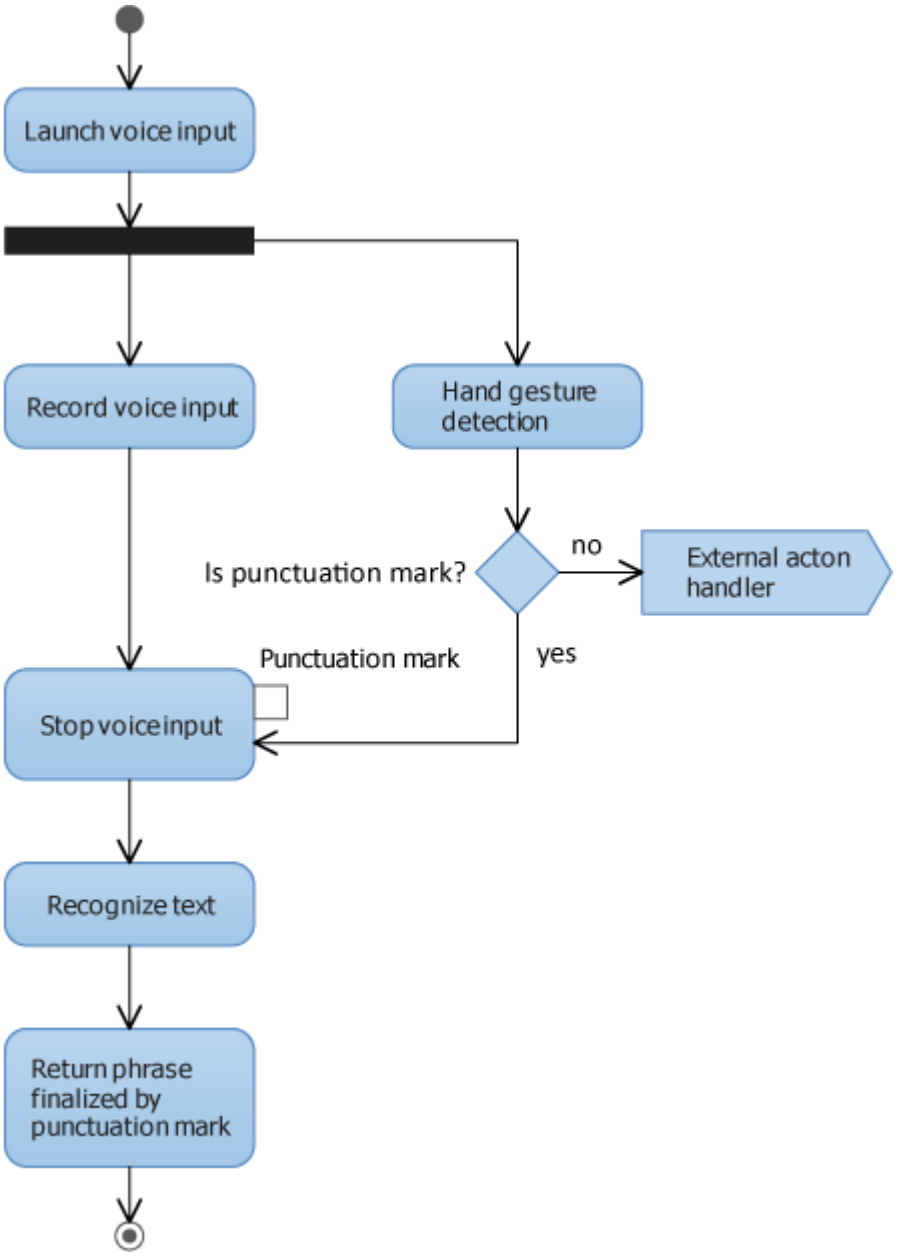

Figure 14. Voice Input Combined with Hand Gesture Activity Diagram

The UI, for combined input, could be created according to a one row layout, where the available controls are represented by the most commonly used punctuation marks, as shown in Figure 15.

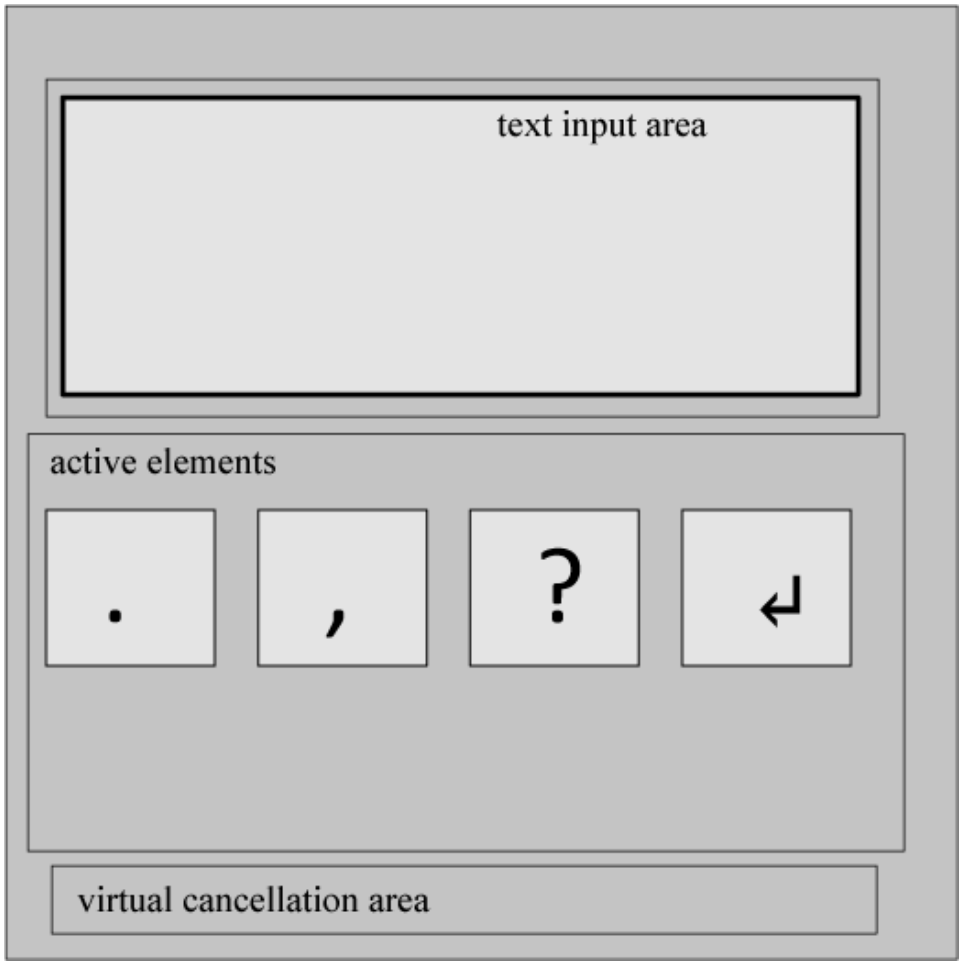

Figure 15. UI for Combined Voice and Hand Input

The use of a dual input system, as shown in Figure 15, preserves user's speech style while providing them with the opportunity of manually input punctuation marks, as in written language. This supports, and correlates with, the natural behaviour, where punctuation marks are not pronounced, but they are always used in writing (use of hand). A further advantage of such an implementation is that the interruption of voice input is connected to a meaningful action (i.e.: the insertion of a period). Therefore correction of misused or misinterpreted punctuation marks, after speech input, would not be necessary.

From the technical point of view, the same break in voice input, continued by hand gestures, allows the application to temporary split the audio file; thus the recognition speed is increased, by sending smaller audio files more frequently. Furthermore, the fragmentation of text input enables a versatile undo/redo functionality implementation and improves the editing possibilities.

8 APPLICATION DESIGN. CORE ARCHITECTURE

This chapter presents the technical description of STTK's core architecture as constructed by its developers.

8.1 STTK's Core Activity

The main application should provide sufficient functionality to: manage the GUI and modify it according to the current task, extend its functionality depending on user's needs, and provide sufficient core functionality for the implementation of the desired input methods.

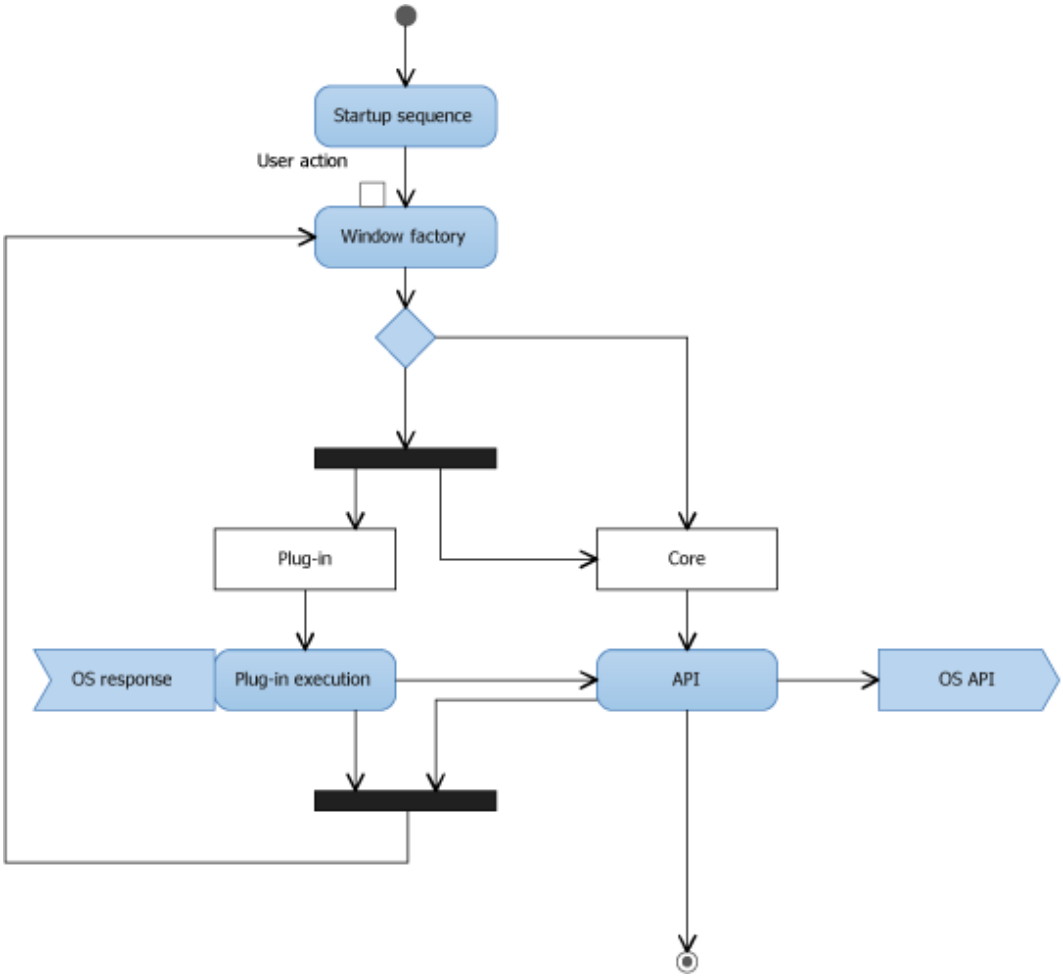

Figure 16. Application's Core Activity Diagram

According to the required functionalities the structure of the application, as shown in Figure 17, would fulfil all the requirements. The following subsystems could be specified on the application level: the application management system,

responsible for application's lifecycle, and the input-output system, that integrates the hand gesture recognition system and the speech-to-text systems.

By applying the advantages of the dual voice input system, described in Subchapter 7.3, the application requires separate modules for the implementation of both, the voice command interface (e.g. .Net SpeechRecognition class) and fluent speech recognition interface (e.g. Google Speech API). The core modules are present in Figure 17.

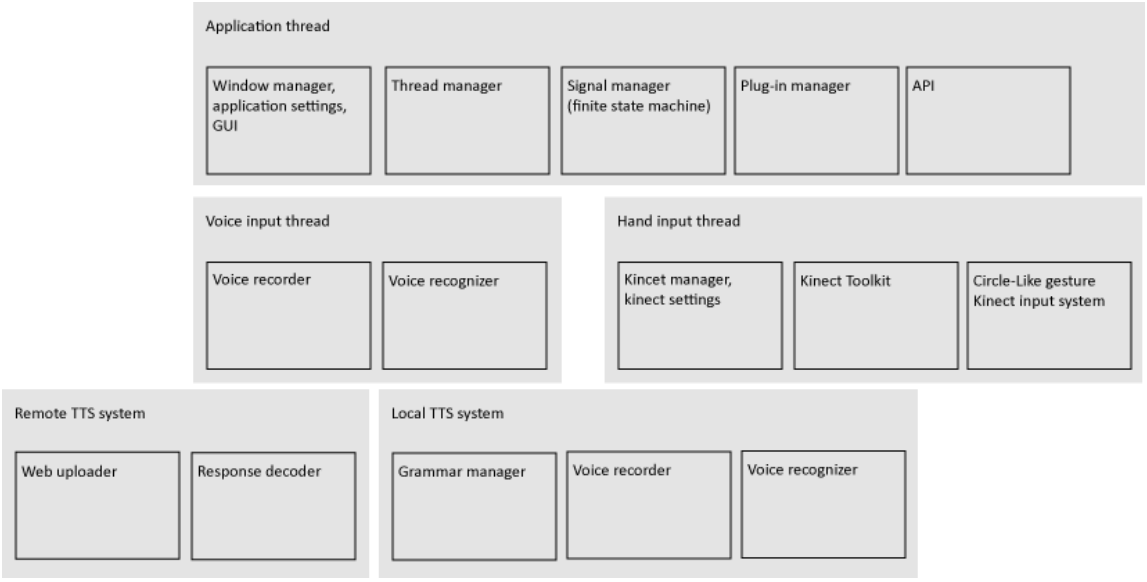

Figure 17. Application's Core Modules Structure

8.2 Architecture of Modules

The natural behaviour of the application is reached by the implementation of multi-threading architecture. There are three tasks are to be performed by the core application: providing availability and functionality of the application (window management, thread management, control over execution sequences), providing the natural input functionality (speech-to-text module, hand gesture module) and providing extendibility (API, plug-in management).

8.2.1 Application Thread and Finite State Machine

As the user input management should not affect the functionality of the main window and the execution of all modules, in a main thread may cause starvation

of the application, a thread management system was implemented in application's prototype. A second reason for the thread management system implementation is that any time-consuming operations, connectivity problems, or failures in file management, may cause actions in main window to be temporary unavailable.

Eventually, during their execution, some tasks should be sequentially executed and controlled by both the user and application. In order to avoid any troubles caused by such situations, the developers have implemented the Finite State Machine, which contains a tree of actions and respective initialization and termination signals.

8.2.2 Voice Module

The voice module is responsible for handling audio data (voice input) and the recognized phrase. Its second function concerns the delegation of recorded audio stream to a specific speech-to-text recognition system, controlled, on the application level, by the main workers and signals of the finite state machine.

Both Speech-To-Text recognition modules have common recording and output classes, and are distinguished only by intermediate actions: for Google Speech API, web uploader and JSON decoder; for .Net SpeechRecognizer class, operations with build-in commands' dictionary.

For the audio stream recording a customized FLAC library, written in C, is used.

8.2.3 Hand gesture module

STTK's hand gesture module includes: the common Kinect Toolkit to handle the Kinect sensor, the default Kinect hand gesture set (Figure 8) and the custom "Circle-Like" gesture set (Figure 9).

8.2.4 Plug-in System and Scripting the Ordinary Tasks

The AutoIt scripting language is suggested to be the preferable scripting language for preparing and implementing the integration scripts with third-party software. The purpose of such scripting is to extend the functionality of the available, but inaccessible or limitedly accessible applications, required in the regular lives of people with motor disorders.

The implemented Custom XML markup allows the modification of main application’s GUI, according to plug-in’s requirements.

9 CONCLUSIONS

The study at hand aimed to discover the needs of people with upper limb disorders, who due to their physical impairment cannot use computers in their daily activities, such as: sending e-mails, paying bills, etc. The initial research question “How to design the NUI for people with upper limb disorders implementing speech recognition and hand gesture recognition?” found its answers after STTK’s tester, a person suffering from an upper limb disorder, tried the application and provided feedback to the developer. STTK, the application introduced by this research, has undergone two sets of changes based on the feedback provided by J., application’s tester.

9.1 Achievements in NUI development

The innovative approach of implementing an UI, supporting gesture-based dual input (mouse and Kinect), in combination with speech recognition, has proven to significantly boost users’ experience in their daily computer use. Furthermore, the improvements brought to the UI, in terms of its usability from a great distance (controls’ size justification, text formatting and formatting rules saving, native colours), were a must for the tester of the application, who, due to their disorder, being close to a computer causes pain. On the other hand, as a motor disabled computer user, who relies on speech input and eventual gesture recognition for accomplishing computer-mediated tasks, it is also very important to have an application which stays responsive at all times, despite the eventual need of simultaneously running many applications, and switching between them.

While things as the earlier mentioned might sound like common sense, it is to be noticed that a NUI cannot be developed without sequential and thorough testing and feedback from users who actually need such an UI. The need for appropriate punctuation and character recognition through voice input (i.e.: digits to be represented as digits, not as their textual transcription), is only one example of detail needs which cannot be discovered, unless testing is involved.

Special attention should be also paid on the usability in terms of application’s reaction when invoked. For instance, the STTK should have started recording

and voice command processing, without taking focus on itself, but performing the tasks appropriately.

9.2 Suggested innovations

The combination of hand gesture and voice input has proven to be a very powerful tool in giving the physically impaired a chance to use the computer, for any computer-related tasks. It reduces eventual effort and it fosters comfort in use. Hand tremor-caused difficulties in selecting an object on the screen can be tackled with Kinects' default hand gesture recognition system's slight modifications; the "Circle-Like" hand gesture model was, in this case, the perfect workaround.

The task to build natural interface for human-computer communication belongs to both social and technological domains.

The effort to build NUI should be started keeping in mind that, nowadays, computers have a significant influence on everyone's quality of life. For people, limited by physical conditions, the computer should not be an impassable barrier between real and virtual life; but on the contrary, computers must help these individuals to enter the global communication space as supportive companions, they must simplify daily activities, provide additional communication possibilities, and help them reach a variety of online services in the most comfortable way. The task of creating an unobtrusive access to information resources is essential to human wellbeing and human rights in equal access to information.

On the technological level, there are multiple solutions, either already available, or which will be available, in near future, for becoming a base of NCI. The already made progress in the HCI, makes the implementation of universal human-computer communication solutions possible and applicable both for disabled and non-disabled users.

The NCI may be implemented without the customization, very expensive, or rare devices. The motion capture sensors available on market, commonly used

headsets, and regular computer systems are ready to support NCI, as well as to provide any necessary integration with third-party software or web services.

9.3 Suggestions for Further Studies

Further studies could be undertaken in order to find the most efficient way of implementing a more complex punctuation system, as well as to implement the recognition of forenames, surnames, or names of places. The implementation of such functionalities could help the physically impaired in the HCI, their computer use experience being boosted in several ways.

While the implementation of forenames', surnames', or places' names', recognition could be considered to not be a priority, appropriate punctuation should be on top of the development list, when it comes to NUIs, including voice recognition and speech-to-text transcriptions. Appropriate punctuation is always necessary, in order to associate the right meaning to one's message. For instance "Everyone needs to be a role model.", is a basic statement, which does not imply any specific emphasis, while "Everyone needs to be a role model, period." shows that the speaker is convinced of the statement to be true, their own strong position towards the fact that "Everyone needs to be a role model" — "Everyone needs to be a role model. That's it! / End of Story! / There's no other way!"

Such functionalities were not developed for the STTK application, at the moment when the present study was undertaken, due to time constraints and low relevance, when the goals of the research were set.

REFERENCES

Published References

Hevner A., Chatterjee S., 2010, Design Research in Information Systems: Theory and Practice, Springer Science+Business Media, LLC, New York, New York.

Knapp M., Hall J., Horgan T., 2012, Nonverbal Communication in Human Interaction, Wadsword, Cengage Learning, Boston, USA

Lessig L., 1999, Code and Other Laws of Cyberspace, Basic Books, New York, NY

Magee J. J., Epstein S., Missimer E. S., Kwan C., Betke M., 2011, Adaptive mouse-replacement interface control functions for users with disabilities, Boston University Computer Science Technical Report No. BUCS-TR-2011-008

Nardi B. A., 1996, Context and Consciousness: Activity Theory and Human-computer Interaction, Massachusetts Institute of Technology

Saunders M., Lewis P., Thornhill A., 2007, Research methods for business students, 4th edition, Pearson education Ltd., Harlow, England

Electronic References

Abate T., 2013, Stanford Algorithm Analyzes Sentence Sentiment, Advances Machine Learning, Standford University, Stanford Engineering. Available at: http://engineering.stanford.edu/news/stanford-algorithm-analyzes-sentence-sentiment-advances-machine-learning

Baker N., Rogers J., Rubinstein E., Allaire S., Wasko M., 2009, Problems experienced by people with arthritis when using a computer. Arthritis & Rheumatism, 2009; 61 (5): 614-22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19405002

BBC News. December 21, 2006, "Robots could demand legal rights", Available at: http://news.bbc.co.uk/2/hi/6200005.stm

Beeks D., Collins R., 2001, Speech Recognition and Synthesis, Available at: http://www.davi.ws/avionics/TheAvionicsHandbook_Cap_8.pdf

Biery M., 2013, Forbes.com, Industries To Watch In 2014: The 10 Fastest-Growing Fields, Available at: http://www.forbes.com/sites/sageworks/2013/12/29/industries-to-watch-in-2014-the-10-fastest-growing-fields/

Blake J., 2010, Deconstructing the NUI, What is the natural user interface?, Available at: http://nui.joshland.org/2010/03/what-is-natural-user-interface-book.html

Brault M., 2012, Americans With Disabilities: 2010, Household Economic Studies, Current Population Reports, Available at: http://www.census.gov/prod/2012pubs/p70-131.pdf

Bronstein A, Bronstein M., Kimmel R., 2007, "Expression-invariant representation of faces", IEEE Trans. Image Processing, Vol. 16/1, pp. 188-197, January 2007, Available at: http://visl.technion.ac.il/bron/publications/BroBroKimTIP06.pdf

Charlton J, Danforth I, 2007, Distinguishing addiction and high engagement in the context of online game playing. Computers in Human Behavior 2007;23(3):1531e48., Available at: http://sophiaco.wikispaces.com/file/view/Distinguishing%20addiction%20and%20high%20engagement%20in%20the%20context%20of%20online%20game%20playing.pdf/30607767/Distinguishing%20addiction%20and%20high%20engagement%20in%20the%20context%20of%20online%20game%20playing.pdf

Duncan G., 2014, Digital Trends, You can't avoid the 'Internet of things' hype, so you might as well understand it, Available at: http://www.digitaltrends.com/home/heck-internet-things-dont-yet/#!A7uoT

Dutoit T., 1996, A Short Introduction to Text-to-Speech Synthesis, Available at: http://tcts.fpms.ac.be/synthesis/introtts_old.html

Encyclopaedia Britannica, Merriam-Webster, Inc., 2013, Available at: http://global.britannica.com/EBchecked/topic/376313/Merriam-Webster-dictionary

Ferguson, C.J.; M. Coulson, J. 2011, A meta-analysis of pathological gaming prevalence and comorbidity with mental health, academic and social problems. Journal of Psychiatric Research 45 (12): 1573–1576. Available: http://www.tamiu.edu/~cferguson/Video%20Game%20Addiction.pdf

Goldie L., 2006, Online retailers fail to take Web accessibility seriously enough, New Media Age, Available at: http://connection.ebscohost.com/c/articles/23056796/online-retailers-fail-take-web-accessibility-seriously-enough

Google.Trends., Responsive Design, Available at: http://www.google.com/trends/explore#q=Responsive%20design%20

Google.Trends., UI Developer, Available at: http://www.google.com/trends/explore#q=UI%20Developer

Google.Trends.,UX Design, Available at: http://www.google.com/trends/explore#q=UX%20Design

Griffiths M., 2008, Videogame addiction: further thoughts and observations. International, Journal of Mental Health and Addiction 2008;6(2):182-5, Available at: http://addiction2013.files.wordpress.com/2013/02/griffiths.pdf

Guidelines for Meeting the Communication Needs of Persons With Severe Disabilities, National Joint Committee for the Communication Needs of Persons With Severe Disabilities. Available at: www.asha.org

Guye-Vuillème A., Capin T., Pandzic S., Thalmann N., 1999, Nonverbal communication interface for collaborative virtual environments. Virtual Reality, Springer, Available at: http://www.cybertherapy.info/pages/non_verbal.pdf

Health and Safety Executive, Available at: www.hse.gov.uk/statistics/lfs/ulnage1w12.xls and www.hse.gov.uk/statistics/lfs/typesex1w12.xls

Henderson M., April 24, 2007, The Times Online (The Times of London), "Human rights for robots? We're getting carried away", Available at: http://www.thetimes.co.uk/tto/technology/article1966391.ece

Inc., The 2013 Inc. 5000 List, Available at: http://www.inc.com/inc5000/list/2013/x/revenue

Intelligent Information Laboratory, the News at Seven Project, Available at: http://infolab.northwestern.edu/

International Day of Persons with Disabilities, 3 December 2011, Theme for 2011: “Together for a better world for all: Including persons with disabilities in development”. United Nations, Available at: http://www.un.org/disabilities/default.asp?id=1561

Khanna P., Sasikumar M., 2013, “Rule Based System for Recognizing Emotions Using Multimodal Approach”, International Journal of Advanced Computer

Science and Applications - IJACSA, Vol. 4, No. 7, 2013, Available at: http://thesai.org/Downloads/Volume4No7/Paper_5-Rule_Based_System_for_Recognizing_Emotions_Using_Multimodal_Approach.pdf

Larman C., Basili V., 2003, “Iterative and Incremental Development: A Brief History”, Available at: https://www.it.uu.se/edu/course/homepage/acsd/vt08/SE1.pdf

Mitsuku Chatbot, An Artificial Life Form Living on the Net, Available at: http://mitsuku.com/

National Center for Health Statistics, Available at: http://www.cdc.gov

Patton M., Cocharn M., 2002, A Guide to Using Qualitative Research Methodology, Available at: http://fieldresearch.msf.org/msf/bitstream/10144/84230/1/Qualitative%20research%20methodology.pdf

Payne G., Williams M., 2005, Generalization in Qualitative Research, Sociology 2005 39: 295, BSA Publications Ltd®, SAGE Publications, London, Thousand Oaks, New Delhi Available at: http://www.uk.sagepub.com/gray/Website%20material/Journals/soc_payne.pdf

Raudonis V., Dervinis G., Vilkauskas A., Kersulyte GG. Kersulyte-Raudone G., 2013, “Evaluation of Human Emotion from Eye Motions”, International Journal of Advanced Computer Science and Applications - IJACSA , vol. 4, no. 8, 2013, Available at: http://thesai.org/Publications/ViewPaper?Volume=4&Issue=8&Code=IJACSA&SerialNo=12

Rouse M., 2011, SearchHealthIT, Definition Kinect, Available at: http://searchhealthit.techtarget.com/definition/Kinect

Shao-Kang L., Chih-Chien W., Wenchang F., 2005, Physical Interpersonal Relationships and Social Anxiety among Online Game Players. CyberPsychology

& Behavior, February 2005, 8(1): pp. 15-20, Available at: http://online.liebertpub.com/doi/pdf/10.1089/cpb.2005.8.15

Sheard A., Won Y., 2012, PWNED: Motivation of South Koreans Who Engage in Person vs. Person Gameplay in World of Warcraft. Available at: http://www.digra.org/dl/db/12168.29338.pdf

Shires G., 2013, Response to an e-mail message:“ Is the google speech recognition and text to speech engine under free license?” (Published by W3C), Available at: http://lists.w3.org/Archives/Public/public-speech-api/2013Jul/0001.html

Shires G., Wennborg H., 2012, W3C Community Group Final Report, Web Speech API Specification, Available at: https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html

Shlomoa B., Kuhb D., 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. International Journal of Epidemiology, Volume 31, Issue 2, pp. 285-293, Available at: http://ije.oxfordjournals.org/content/31/2/285.full

Szente M., Breipohl W., 2003, Review of impairments related to physical disabilities of the upper limb, EU Minerva-Socrate Project, Available at: http://www.designforall.net/files/downloads/LIMB.pdf

Taylor MG., Linch SM., 2004, Trajectories of Impairment, Social Support, and Depressive Symptoms in Later Life. The Journals of Gerontology, Series B. Psychological Sciences and Social Sciences, 2004, 59 (4), pp. 238-246. Available: http://www.ncbi.nlm.nih.gov/pubmed/15294928

The Loebner Initiative, Available at: http://www.loebner.net/

The Online Investor, 20 Largest U.S. Companies by Market Capitalization, Available at: http://www.theonlineinvestor.com/large_caps/

Trochim W., 2006, Research Methods Knowledge Base, Deduction & Induction, Available at: http://www.socialresearchmethods.net/kb/dedind.php

U.S. Department of Justice, Civil Rights Division, “2010 ADA Standards for Accessible Design”, Available at: http://www.ada.gov/regs2010/2010ADAStandards/2010ADAStandards.pdf

U.S. Department of Justice, Civil Rights Division, “Access for All: Five Years of Progress”, A Report from the Department of Justice on Enforcement of the Americans with Disabilities Act. June 30, 2011, Available at: http://www.ada.gov/5yearadarpt/i_introduction.htm

U.S. Centers for Disease Control and Prevention, 2013, Prevalence of Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation — United States, 2010–2012. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6244a1.htm

Want R., Schilit B., 2012, IEEE Computer Society, Interactive Digital Signage, Computer, vol. 45, no. 5, 21-24 Available at: http://63.84.220.100/csdl/mags/co/2012/05/mco2012050021.pdf

Yadav A., Gesture Recognition Technology, Voyager - The Journal of Computer Science and Information Technology, Vol. 3, No.1 (2006), 86-88, Available at: http://www.itmindia.edu/images/ITM/pdf/Gesture%20Recognition%20Technology.pdf