Automatic Speech Recognition - An Approach for Designing Inclusive Games

Moyen Mohammad Mustaquim

Uppsala University
Department of Informatics and Media
Ekonomikum (plan 3), Kyrkogårdsg 10 Box 513, 751 20 Uppsala, Sweden

moyen.mustaquim@im.uu.se

ABSTRACT

Computer games are now part of our modern culture. However, certain categories of people are excluded from this form of entertainment and social interaction because they are unable to interface with the games. The reason for this can be deficits in motor control, vision or hearing. By using automatic speech recognition systems (ASRS), voice driven commands can be used to control the game, thus opening the possibility for people with disabilities to be included in game communities.

This paper aims to find a standard way of using voice commands for games which use a speech recognition system in the backend that can be universally applied for designing inclusive games. However, present speech recognition systems do not support emotions, attitudes, tones, etc., which is a drawback because such expressions can be vital for gaming.

Taking multiple types of existing genres of games into account and analyzing their voice command requirements, a general ASRS module is proposed that can work as a common platform for designing inclusive games. A fuzzy logic controller is then proposed to enhance the system. The standard voice driven module can be based on algorithms or a fuzzy controller, which can be used to design software plug-ins or be included in a microchip. It then can be integrated into the game engines, thereby creating the possibility for voice driven universal access for controlling games.

Keywords

design for all; inclusive design; universal design; accessibility; inclusive games

1. INTRODUCTION

1.1 Computer Games in Society Today

Games are the vehicle with which society will change the computer, but how will the games themselves be changed by society? We can expect two processes to affect the games: the mass market and the flowering of heterogeneity. In some ways, these processes work against each other [1]. There is accordance that in most industrialized countries demographical, structural and societal trends are moving towards an increase in the elderly population and single households[2]. Scientists agree that in addition to new forms of living together (generations, communities), information technology (IT) is a promising way to cope with such effects to enable the elderly to stay longer in their preferred environment [2]. Consequently, IT must be developed to support user’s needs and emerging requirements[2]. Whatever the circumstances and degree of acceptance by a user of innovative technology, a clear benefit must be offered, whether in a physical, medical or emotional respect[2].

As participation in gaming increases around the world and across a larger part of the population, it is obvious that games are not only an entertainment media for children or young people, but games can be used to serve several purposes for all types of people in society. Games can be seen as an important learning enabler, particularly for groups with severe learning difficulties or those suffering from cognitive disabilities [3].

Play is definitely one of the most important psycho-pedagogical concepts; consequently, games used in an appropriate setting can be used as excellent educational tools [4]. As they grow older, today’s technologically inclined people do not want to be excluded from future services, e.g. games, simply because of design issues associated with the game. Additionally, anyone can suffer temporary or permanent disabilities regardless of age, and it would be unfortunate to exclude them from using games. Moreover, both human death and birth rates are generally decreasing, thereby resulting in an age distribution with a large number of older people (whether disabled or not) in society in the near future. Hence, game industries should be concerned with such issues, so that they can include these groups of people and offer them properly designed services.

Automatic speech recognition systems are not a new concept in modern computing. However, the possibility of using automatic speech recognition systems to control games is one option to offer inclusiveness in game design. People with limited motor control ability can benefit from the opportunity of using voice driven commands for the game. This paper proposes that such speech recognition systems is one of the objectives in achieving a universal design in computer gaming.

1.2 Why Inclusive Games Are Needed

The philosophy underlying inclusive designs specifically extends the definition of users to include people who are excluded because of rapidly changing technology, particularly the elderly and aging, and prioritizes the role and value of extreme user groups in innovations, as well as new products and service developments. It also prioritizes the context of use, both physical and psychological, and the complexity of interactions between products, services and interfaces in specific contexts of use, such as independent living [5].

Two major demographic trends underscore the importance of considering technology adoption by older adults: the aging of the population and the rapid dissemination of technology within most societal contexts. Over the past decade, developments in computer and information technologies have taken place at an unprecedented rate, while technology has become an integral component of work, education, healthcare, communication and entertainment. Games have created a different viewpoint in the past decade in terms of not only being an amusement entity, but having the possibility to serve other functions as well. While we are witnessing explosive developments in technology, the population is aging. In 2003, people aged 65+ years in the United States numbered about 35 million, representing approximately 13% of the population [6]. By 2030, this number is expected to increase to about 71 million, representing 20% of the population (Fig. 1). Moreover, there will be a dramatic increase in those aged 85+ years, from approximately 4 million in 2000 to nearly 21 million by 2050 [6].Similar trends are also occurring worldwide. By 2030, the percentage of people aged 65+ in Europe will be about 24%, and about 12% in Asia and Latin America [6].

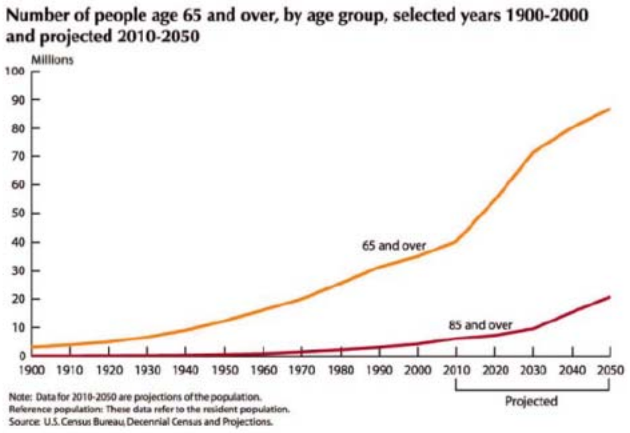

Figure 1 - Projected growth of people 65+ in the United States. Data source: federal interagency forum on aging-related statistics, 2005

Similar trends exist in the European Union (EU), in which 48% of the households had access to the Internet in 2005, and 23% had a broadband connection. However, a gap still remains between users and non-users according to age. Specifically, the proportion of computer or Internet users among those 16–24 years of age is three times higher than among persons 55–74 [7]. Recent data also indicates that older adults use other forms of technology such as ATMs and DVDs less than younger adults [6].

To make technology useful to and usable by older adults, the challenge for the research and design community is to “know thy user” to better understand the needs, preferences and abilities of older people [6]. It is fairly well established that many technology products and systems are not easily accessible to older people. There are of course a myriad number of reasons for this such as cost, lack of access to training programs, etc. Nonetheless, to a large extent this lack of accessibility is due to the fact that designers are unaware of the needs of users with varying abilities, or do not know how to accommodate their needs in the design process [8]. For example, findings from an observational study of designers involved in a design competition for older people [9] revealed that the designers tended to restrict their use of user information and user involvement in the design process [8].

As previously described, it is apparent that as games are becoming more popular and the users of technology become older, it is now time to think about what will happen in the future when the users will be more technologically inclined, but restricted from playing games because they are not properly designed for him or her. Regardless of this issue, any age group people can suffer from certain physical disabilities that exclude them from playing games because of the design issue. For this reason, the importance of designing inclusive games is going to become more and more highly valued in the future. If people with physical limitations or people with disabilities due to age can access a game because it is designed according to their needs, universal access to game design has been achieved. This paper takes an approach to designing inclusive games for achieving a universal design for gaming and for giving accessibility and interaction. Hence, the use of speech recognition systems is the topic of research here.

1.3 Future Game Design

People with disabilities are certainly one of the groups who will benefit the most from the ICT revolution. The potential of ICT is first of all based on:

• Flexibility and adaptability of Human-Computer Interaction (HCI) to better address the needs of a diversity of end users in various situations (multimedia, multimodal) [3];
• Universal application of a limited number of HCI concepts (WIMP/SILK, GUI) in different areas; these interfaces provide a stability for the end user by providing standardized, basic interaction principles which stay the same, or at least remain similar, in changing technical and application scenarios [3].

Over the last few decades, the emerging field of Assistive Technology (AT) has made considerable progress in connecting to and expanding current interfaces, thereby opening a door for people with disabilities to access all areas in which ICT is used [3]. As society ages and the care gap increases, technology is envisaged as a means of allowing older people to live independently for longer. On its own age is not a disability, but older people are more likely to experience disabilities of various kinds. These include both single severe impairments and combinations of multiple minor sensory, cognitive and physical impairments that can combine to make products and services much more difficult to use [10] [11]. Mobility decreases with advanced age, so for many elderly people this means an equivalent increase in the difficulties attached to social interaction [3]. For those not living in a supervised environment (residential homes, geriatric centers etc.), meeting friends and visiting social centers requires organization, which can be facilitated by easy to use communication methods [3].
The main groups of people addressed by these accessibility issues are [3]:

• People who cannot use an ordinary graphical interface because they are totally blind or because they have a severe visual impairment (sight rated <0.05) [12] [3];
• People who cannot use or have limited access to ordinary input devices such as keyboards, mouses, joysticks or game pads due to limited hand dexterity [3];

• People with cognitive problems who need support to better understand the course of events and to react properly (e.g. symbols, text, speech and easy to understand support) [3];
• People with hearing problems or deafness who are not able to accommodate to sound-based interaction modalities [3];
• People with problems in reacting to a strict time setting of the game because of various functional, cognitive and psychological problems [3].

The revolutionary potential of AT to connect to standard ICT/HCI for people with disabilities emphasizes that every change to the standard HCI requires a corresponding attention in the field of accessibility to [3]:
• Maintain the achieved level of access, and
• Exploit the potential of new HCI possibilities for enhanced access for different groups of people with disabilities [3].

Although Information and Computer Technology (ICT) is a highly attractive market, mainstream games have not been considered as a subject for research and science, particularly in the field of HCI, for a long time [3]. Games have developed parallel to the standard HCI market, leading to significantly different interface principles based on individual engines and hardware [3]. This situation has been changing significantly over the last few years, as games have become more and more subject to both mainstream HCI and accessibility [3].

Designing games that work for players with disabilities presents quite a challenge: In general, the accessibility of games is a more complex problem than software or web accessibility [3]. In the near future where computer games are a major recreational activity as well as a tool for attaining tasks, a mass market of computer games not too different from what we have now should focus on “properly designed” games, in which a proper design unlocks the possibility of making a game complete and self-sufficient.

1.4 Significance and Nature of the Interaction in Games

Interaction in gaming is important for several reasons. First, it injects a social or interpersonal element into the event. It transforms the challenge of the game from a technical one to an interpersonal one. A puzzle will always present the player with exactly the same challenge, but a game opponent reacts to the player’s actions and presents different challenges in each game. This difference has a major emotional significance [1].

What is important about the modes of interaction in gaming is not their mechanical quality, but their emotional significance. Thus, the degree of interaction provides a useful index of “gaminess” [1] [13]. It then becomes important to use the right interaction technique at the right level during the design of voice interaction techniques for controlling a game.

The individual user may have excellent ability in some areas and yet be poor in others. For the population as a whole, there can be a wide variability in any given attribute. The complexity of the problem increases dramatically as more attributes are considered. Generally speaking, attributes deteriorate with aging, whereas the variability increases [14].

1.5 Using Human Modality in Inclusive Design

The advent of the Information Society, along with the emergence of novel technological paradigms such as mobile or ubiquitous computing, ambient intelligence and the "disappearing computer" raises new and stimulating scientific challenges in research on multimodal HCI [15]. New input/output modalities and forms of multimodality are therefore needed to provide easy and effective access for control to all users, including the very young and the elderly, as well as people with various types of disabilities in all contexts of use, especially in mobile interaction with embedded systems, wearable computers, and augmented everyday objects that integrate aspects of intelligent interactive behavior [15]. A loss or deterioration of hearing occurs when problems arise with the perception of such sound elements as frequency, pitch, timbre and loudness of the surroundings [16]. Some deaf people, though not all, use full natural language to communicate among themselves, which is known as sign language[16]. Sign language for deaf people is based on hand movements, face, eye and lip mimicry and body movement [16]. It uses a visual-sign system with defined positions, locations, orientations and movements of hands and fingers, as well as facial expressions [16]. Sign language also has its own linguistic structure, which is independent of any vocal language in the same geographical area [16]. Word order (which is different from written language) and grammatical structure are a product of the separate development of a physical language within the deaf community [16]. This method of communication has a strong impact on the culture and language of the deaf community [16] and the individuals within that community, thus enabling a multimodal approach for controlling games designed for such communities to use sign language in a more interactive and fun way that is very promising. It is therefore of paramount importance to use human modalities for controlling games.

2. BACKGROUND

2.1 Game Controller

A game controller is a device used for controlling the playable character or objects, or for otherwise providing input into a computer game. The controller is usually connected to a game console or a computer by means of a wire, cord and the more modern wireless communication of today. Typical game controllers are keyboards, mice, joysticks, game pads and so forth, in addition to special purpose devices such as steering wheels or foot pedals. The principal function of a game controller is to control the movement or actions of a playable body-object or otherwise influence events in a video or computer game. There are health concerns related to game controllers such as a risk for injuries such as repetitive strain injuries or carpal tunnel syndrome; hence, they need to be designed ergonomically to give the best possible relaxation of the hands and mind. However, certain groups of people have limited accessibility or no accessibility at all to such game controllers because of their physical limitations. Alternative ideas for game control are therefore of great importance to consider for such groups of people.

2.2 Speech Technology

Speech technology seems to provide new opportunities for improving the accessibility of electronic services and software applications, including games, by offering compensation for the limitations of specific user groups. These limitations can be quite diverse and originate from specific sensory, physical or cognitive disabilities such as difficulties in seeing icons or controlling a mouse or keyboard [14]. Such limitations have both functional and emotional aspects that should be addressed in the design of user interfaces [17]. Speech technology can be an "enabler" for understanding both the content and "tone" in user expressions, and for producing the right information with the right tone [14].

2.3 Automatic Speech Recognition System

Automatic speech recognition (ASR) is the automatic conversion of human speech to a sequence of words. The aim of ASR is to automatically recognize what has been said. Nowadays, ASR systems are generally based on the hidden Markov models (HMM) [22] for modeling the acoustics of speech, and use either statistic language models (n-grams) or rule-based grammars to model the language component. Furthermore, there are many techniques that normalize the acoustic features of the signal and adjust the acoustic models to a specific speaker or to different recording conditions[14].

If a speech recognizer is part of an application, it is usually used as an input technology or technique [18]. Various other performance measures can augment the word error rate, depending on the application. For instance, van Leeuwen et al. [19] mention the following issues that are important to consider when estimating the performance of an ASR application: the type and quality of feedback, error correction strategy, dealing with out-of-domain words, speed and response time, the user's situational awareness in the dialog structure, dialog / task success rate and subjective impression of the overall performance of the system [20].

2.4 Fuzzy Logic

Fuzzy logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PCs or workstation-based data acquisition and control systems. They can be implemented in hardware, software or a combination of both [17]. Fuzzy logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy or missing input information, as fuzzy logic's approach to control problems mimics how a person would make decisions, only much faster. Fuzzy logic was first conceived as a better method for sorting and handling data, but has later proven to be an excellent choice for many control system applications since it mimics the logic of human control. It can be built into anything from small, handheld products to large computerized process control systems. It uses an imprecise though highly descriptive language to deal with input data, which makes it more like a human operator. It is very robust and forgiving of non-perfect operator performance and data input, and often works when first implemented with little or no tuning.

2.5 Fuzzy Logic Controllers

Fuzzy logic control may be viewed as a branch of intelligent control. It can be seen as an emulator of the human decision making process in that it is approximate rather than exact. Conventional control systems work either in a linear or a non-linear manner. Linear systems are simple and well explored, though not very flexible. In contrast, non-linear systems can be powerful, but are often sensitive to modeling errors. A fuzzy controller is effective for situations in which the control process is too complex to analyze by conventional quantitative techniques, as well as when the available sources of information are interpreted qualitatively, inexactly or uncertainly. A fuzzy logic controller has several advantages over a traditional controller, as it is more flexible and easy to understand. A fuzzy controller uses human linguistic terms for control. To design a fuzzy controller for parallel or distributed control, either multiple fuzzy rules or complex non-linear systems can be used. However, there are more parameters to tune in the fuzzy controller, and the stability of the fuzzy controller is sometimes difficult to analyze mathematically. Regardless of its complexity, a fuzzy logic controller can be used where it is difficult or impossible to model mathematically, or where traditional strategies are too complex or non-linear to be controlled. Figure 2 shows the basic structure of a fuzzy logic controller.

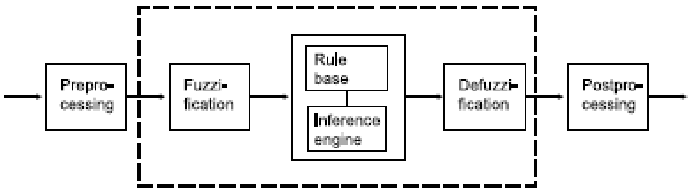

Figure 2 - Structure of a fuzzy controller

A typical design for a fuzzy controller involves some sequential process starting from the determination of state and control variables, and then an inference method and a fuzzification method are selected. The next step is to normalize the state variable space, while also determining the shapes of the fuzzy sets. Once this is accomplished, the fuzzy rule base is constructed and a defuzzification strategy is selected to convert the fuzzy sets of data to the crisp data set. There are further steps involved of course such as testing, fine tuning of the controller and construction of a lookup table, which can be used in order to save the inference and defuzzification time. The proposed fuzzy controller described in the proposed system section is based on the above description.

2.6 Requirements of Voice Controlled Systems for Games

While the traditional controlling approach of game is uses input devices, people with certain limitations for doing so would like to be able to use other control alternatives. The approach presented in this paper is to use speech for controlling the game. People with motor system problems, limitations in moving the upper limbs, problems with hand muscles, hand joints, muscular stress and many other neurological disorders can highly benefit by using a voice driven command system in gaming.

It is also important to utilize the user's emotions to control the character of game in a different manner. An emotion can interrupt ongoing interactions, and tends to change the course of action, [21] e.g. if the user's goal of browsing a certain website for information is continually impeded by the slow loading of the site, and the relevance of this site to their goal is low, they will change their course of action and search for the information on another website [21]. If they can only find the information on this site, then the relevance towards the goal is high, the frustration tolerance is higher and the likelihood of leaving the site will be lower [21]. Similarly, the response a user receives from the environment of the game based on the voice input he or she implies can be highly dependent on how low or loudly the user is uttering the voice command. Emotions are created every time there is a perception of important changes in the environment or in the physical body [21]. Basically, an emotion is a psychological state or process that functions in the management of maintaining the balance of information processes in the brain and relevant goals [21]. Every time an event is evaluated as being relevant to a goal, an emotion is elicited [21]. Positive emotions occur when the goal is advanced, whereas negative emotions occur when the goal is impeded [21]. Emotions also have an impact on cognitive processing during every interaction a user performs with an interface [21]. A person's voice can also be weak or loud regardless of their age, so considering the emotion in voice control is optimal. Age related issues can make a person's hand weak; hence, the use of speech is reasonable for such users and does not require any special hardware. The ordinary computers of today are able to carry speech transforming operations from user to the process handler of the game. As a result, there is no need for any extra hardware and disabled people do not need to wear or use any special device to use speech systems for controlling games. For these reasons, the use of human speech modality is a constructive move towards achieving a universal design of games.

3. PROPOSED SYSTEM

3.1 Description of the Proposed System

Figure 3 shows a block diagram of the proposed system, which consists of a standard automatic speech recognition system in which the vocabulary array is built by fetching commands from the most commonly used game platform. The user input is processed through a soundcard that converts the analog data to a digital format and then checks with the user word model from the vocabulary register. Once the word uttered by the user matches with a word from the vocabulary list, it then sends the binary signals corresponding to that word to the game's process handler. Users see that the action has been taken according to their given voice command, through the GUI of the game. A command voiced in a low pitch should result in something different compared to one said in a higher pitch, thereby allowing the user's emotion to control the game.

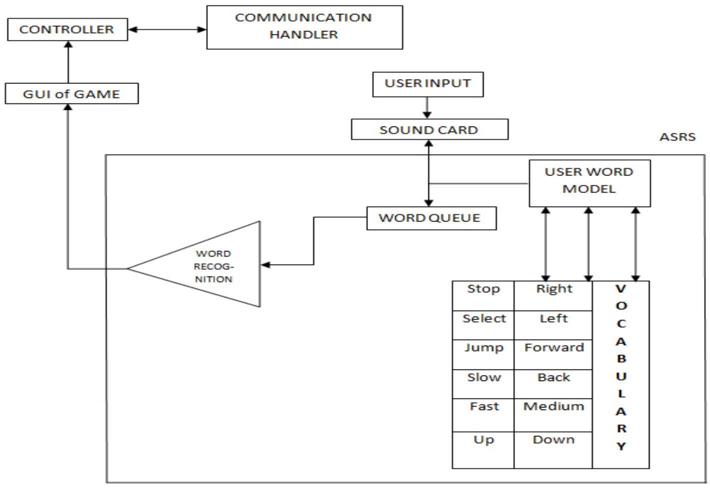

Figure 3 - Proposed automatic speech recognition system model for controlling the game

3.2 The Traditional Controller

The idea of using a controller between the word recognition system and the communication handler of the game engine is that the user's command can be interpreted in a different manner, i.e. the way users send the command for executing a certain operation. The controller proposed here can be everything from a simple logic-based controller to a sophisticated fuzzy controller. A probability- based model such as an HMM [22] can also be used to construct an algorithm, which is shown in the results section. A logic- or probability-based controller that follow certain algorithms derived by the method of this research can be programmed on a single microchip (a PIC microcontroller for example), which can be used for universal voice access prospects for controlling the game.

4. METHOD

4.1 Selection of a Set of Commands

To initiate this research, computer games have been classified while bearing in mind that elderly and disabled people of any age can be the user of the game, whether now or in the future. The classifications for computer games chosen for this research were: action, first-person shooter, strategy, board, action-adventure and sports. At least three games from each of the group of classified games were played and observed. The most frequently used commands from each game were filtered and listed. After analyzing the three different values of the filtered commands, the best possible commands were picked for each category of the game. The same technique is used to filter and fetch the best commands for all classified groups of the game. Two groups of users were involved in playing and observing the games, and were told to write down the command they are using the most with the input devices. They were asked which command they would like to use if the game was to be controlled by voice commands. Their written commands and personal choices in using voice commands were closely matched before creating the set of commands from various categories.

4.2 Obtaining User Voice Command Samples

The next step was to record and monitor the frequency of each command uttered by the users, and six users were selected for these test: three males and three females. Their voices were recorded uttering the commands and tested in two different environments. One was a fairly silent environment that was assumed to be the normal setting when playing a computer game, while another had a little noisier atmosphere. The idea was to see how the level of frequency of command varied with respect to a different environment, and each user was told to give their voice command three ways in two different environments: 1) the way they wanted, 2) with some emotion in it, and 3) with no emotion at all. Therefore, one particular command has been recorded from one user for analysis, total 6 times in two different environments and one particular command had 36 different data; 18 from three male users and 18 from three female users for further analysis. For this test, the audacity audio editor and recorder (available in: http://audacity.sourceforge.net/) for windows was used which is a free multilingual audio editor.

4.3 Analysis of Voice Data

The frequency domain filtering uses a Fourier transform filtering technique implemented in MATLAB. In this filtering technique noise is basically considered as unwanted high frequency content in the sample data and applying a low pass filter to the noisy data allows low-frequency components to remain unchanged while high-frequencies are smoothed or attenuated. After feature extraction using frequency domain filtering method the best data for one particular command recorded by three male and three female users in two different settings was selected which was a satisfactory average value of frequency and intensity for each command and repeating the same procedure frequency and intensity values for all 15 commands were picked up.

5. RESULTS

5.1 Set of Commands

Once the individual category's selected commands were found, a set of commands from 'all type of games' were created. If a command from one category was member of the main set, then it was picked up. This is how the best possible 15 commands were picked up from all these categories of games, considering that they are mostly used and quite obligatory at controlling any of the categorized games. Table 1 shows a sample data collection from three different users while involving in selecting best commands from different types of games.

Table **1-** Sample data from three different users

Game Type	**User 1**	**User2**	**User3**
Action	Fire	Right	Left
First person shooter	Select	Fire	Back
Strategy	Stop	Select	Forward
Board	Select	Back	Forward
Action-adventure	Jump	Fire	Select
Sport	Stop	Slow	Forward

5.2 User Voice Samples

The next step was to analyze the recorded data. For doing this, each individual command from six different users were plotted in graphs to analyze the spectrum. The frequency (Hz) and intensity (dB) values were picked up from the graph and frequency domain filtering method was used to eliminate the noisy dataset. Figure 4 shows the spectrum analysis of the command ‘Select’ by one male and one female user.

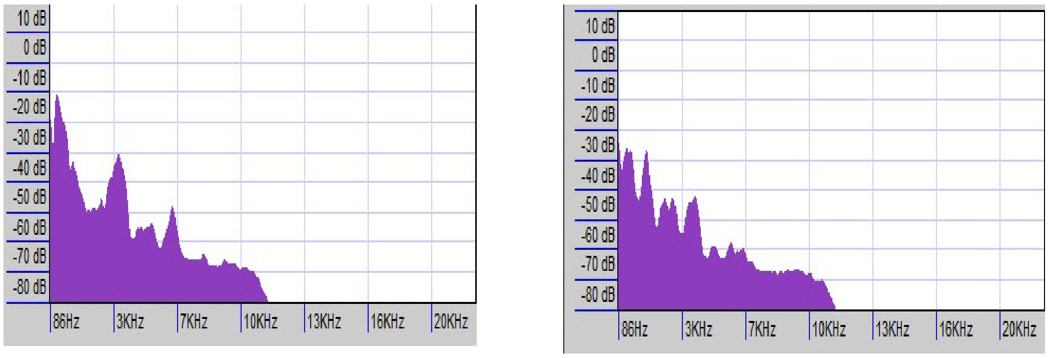

Figure 4- Left spectrum is obtained by a male player and the right spectrum is obtained by a female player for saying same command ‘Select’

5.3 Analysis of Voice Data

Figure 5 shows how the frequency domain filtering was used to remove noise from the data obtained from the spectrum. The left most curve shows the filter function that was constructed. Then, to filter the data in the frequency domain, the Fourier transform of the data is multiplied by the frequency response of a filter and then an inverse Fourier transform was applied to return the data to the spatial domain. The third curve shows that the same filter function can also be used as a high pass filter that is, allowing only the high frequency or noise components through. MATLAB was used to generate these curves.

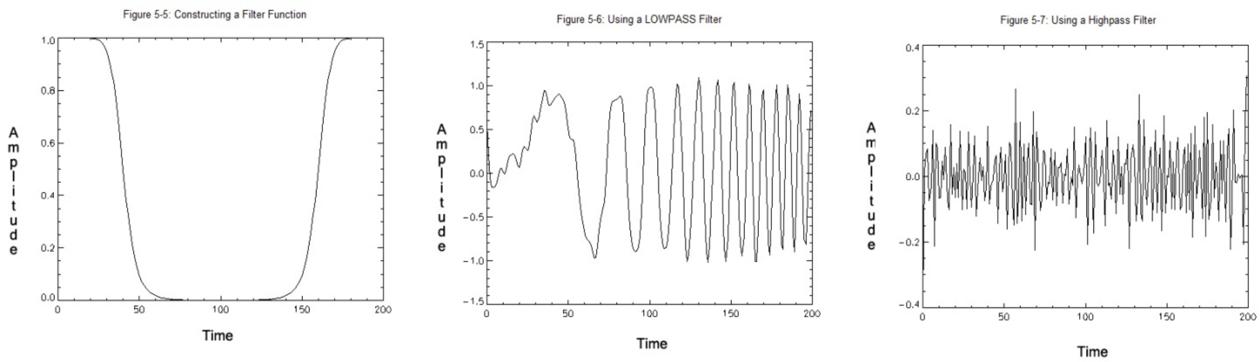

Figure 5- Frequency domain filtering method for eliminating noise data obtained from voice spectrum
The curve in the middle shows the effect of using a low pass filter on the noisy dataset, while the right-hand curve shows the use of a high pass filter for the same function, which allows only the high frequency or noise components through. The high pass and low pass filtering were simultaneously used for male and female voice data according to the number of high or low noisy levels in the data.

5.4 Algorithm

The frequency values well filtered and obtained from different commands were further analyzed to create the average, which is the mean value. The Hz and dB levels were ranged and based on the filtered data and then used to construct a controller connected between the automatic speech recognition system and the game communication handler. The ASRS system detects and matches the stored vocabularies, which in this case are the commands, while the controller checks the emotion of the uttered command (frequency and intensity) to instruct the game communication handler according to the received pitch from a user. It is possible to construct a simple algorithm using the straightforward logic-based loop in which the purpose will be served, but the run time might be quite long. A hidden Markov model (HMM) [22] is the broad-spectrum basis for modern forms of speech recognition systems, and is a statistical model where the states are not directly visible to the user. This model can use the results of this research. The idea of using the frequency values obtained from the different users commands in HMM that combine many other algorithms for temporal pattern recognition in the game (speech in this case) can be viable. An example algorithm from this research for deciding a command between “Fire” and “Jump” based on a statistical HMM may look as follows:

```
States = (“Jump”, “Fire”)
Observations = (“Hz”, “dB”)
start_probability = {“Fire”: 0.6, “Jump”: 0.4}
 transition_probability = {
 “Fire” : {“Fire”: 0.7, “Jump”: 0.3},
 “Jump”: {“Fire”: 0.4, “Jump”: 0.6},
   }
 emission_probability = {
   “Fire” : {“Hz”: 0.7, “dB”: 0.3},
   “Jump” : {“Hz”: 0.6, “dB”: 0.4},
   }
```

From the above piece of code, it can be understood that if a user tends to use a higher tone to utter a command, it is possibly “Fire”, with a higher probability assigned in the function start_ probability. The probability of the command “Fire” to be “Fire” when ‘Fire’ is really uttered is high when it is uttered in a high pitch; otherwise, it is a “Jump command and this logic is assigned in the function transition_probability. The last function, which is called emission_probability, uses the frequency (Hz) and intensity (dB) value’s average (in this example) or limits and decides the uttered command received from the user and sends it to the game communication handler for the execution of the command on the game environment and shows it on the game’s GUI.

5.5 The Fuzzy Controller

The proposed fuzzy controller’s block diagram is shown in Figure 6. The fuzzification process involves handling voice command variables ($d_1,d_2 \ldots d_n$), output variables (ω and v) and transforming numerical variables into fuzzy sets. Two output variables ω and v are used, respectively, for determining the angular and linear velocity of an object. Hence, state variables ($d_1,d_2 \ldots d_n$) are decided with the error and change of error by the user’s voice input. Here, error is determined by the error from the processed output minus the error from the last processed output. As stated before, the controlled variables ω and v are decided to determine the angular and linear velocity of an object, which is important to consider for the various movements based on the user-given voice command. The fuzzification process converts the crisp input values to fuzzy set values.

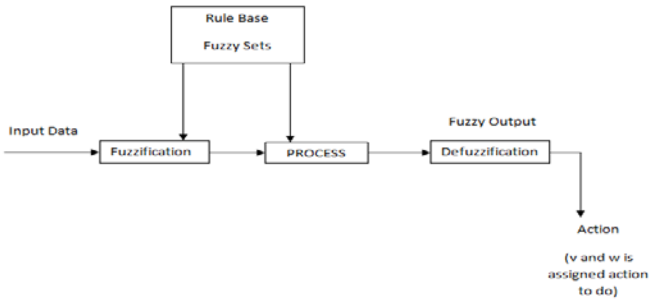

Figure 6 - Proposed fuzzy logic controller system

The Mamdani inference method was chosen to design this fuzzy controller [23]. The fuzzy singleton [24] fuzzification method is selected, as it measures the state variables without uncertainty. The inference process defines the connective implication and rules the combination operations. The controller uses minimum connections and a Singleton sum product inference mechanism. This choice relies on two reasons. First, the product preserves the shape of the output fuzzy set and second, with the sum, the result is influenced by difficult rules, thus reaching the same conclusion.

Output variables (ω and v) are also modeled as a discrete fuzzy set, as it increases speed in the inference and defuzzification process. Figure 7 shows the shape of fuzzy input sets, where d_1, d_2, d_3……d_n are considered to be the voice sensor data received from the user, and the angle θ is used as a guidance operator.

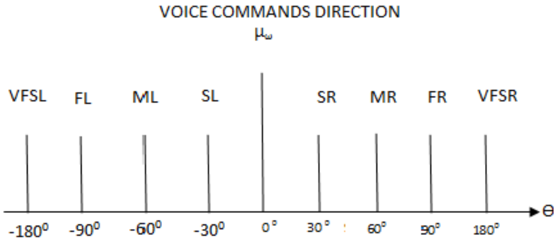

Figure 7 - Shape of fuzzy input sets deciding direction of object

That is, the way d is sensed from the user input triggers the value of θ and puts a decision value on the output variable, describing which angle the object should move at. The value of θ is discritized and normalized within the range of [-180,180]. Figure 8 and Figure 9 show the shape of fuzzy output sets where the limit of v and ω are discretized and normalized within the range of [0, 8] for v and [-8, 8] for ω.

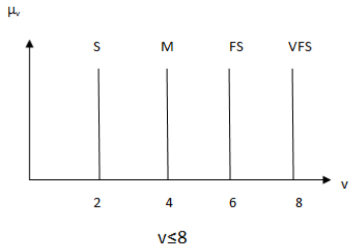

Figure 8 - Shape of fuzzy output sets for linear velocity

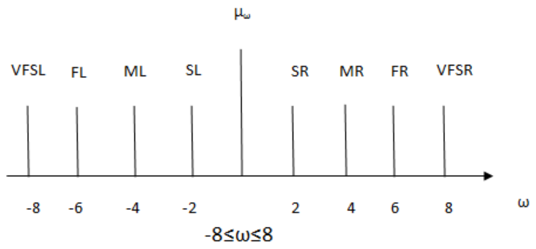

Figure 9 - Shape of fuzzy output sets for angular velocity

The numerical values obtained from the two output sets are sent for a combination of two sets, and then sent to the communication handler of the game to execute the desired action. The content of the fuzzy set is linguistic terms which are the rule-based fuzzy set used to measure the pitch of the voiced command to act differently from the character of interest.

The center of gravity (COG) [25] method is used for the defuzzification strategy, as there is less processing time in this problem with no degradation of the defuzzified value. In addition, the COG method simplifies the computation of the inference mechanism.

5.6 Fuzzy Linguistic Terms

Sample linguistic terms for the rule-base fuzzy set for a few of the game commands are shown below. The knowledge of these commands is picked up from the most widely used voice commands for controlling the game.

R=Right

SR= Right turn slow

MR=Right turn medium

FR= Right turn fast

VFSR= Right turn very fast

L=Left

SCT= Select

NS=Not select

ST= Stop

U=Up

MV=Move

S= Slow

M= Medium

FS= Fast

VFS= Very fast

5.7 Sum-Product Inference

The degree of each rule is determined by the following rule:

β_l=min($\mu A_i(d_i)$…… μA_i (d_k)). Ai are the input fuzzy sets that defined linguistic terms and d_j (j=1….k) are the input variables.

Output fuzzy set $B_l´$ is constructed using the following rule:

For each rule l, we use product operation $\mu\ \beta_l´(v)= \beta_l\ \mu B_l(v)$. Here, B_l are the output fuzzy sets and v is the output variable.

The combination of output fuzzy sets to single fuzzy set B´:

Sum operation: $\mu B´_l(v)= \mu B´_l(v)+\ldots+\mu B´_N(v)$

5.8 Construction of Fuzzy Rule

The following example fuzzy rules are the representation of the gathered knowledge of the user's voice command for game control.

Turn Left:

If θ is SCT and d0 is L and d1 is SCT Then v is ST and ω is L

Stop:

If θ is MV and d0 is SCT and d1 is ST Then v is ST and ω is ST

Right Turn Fast:

If θ is SL and d0 is R and d1 is FR and d2 is MV Then v is ST and ω is FR

Select and Move Forward

If θ is NS and d1 is SCT and d2 is FW and d3 is MV Then v is MV and ω is ST

5.9 Implementation- Hardware and Industry Perspective

The use of a PIC (Programmable Interface Controller) microcontroller to program the logic instructions derived in this research can be a crucial initiative. For example, if a soundcard of the computer comes up with a built-in chip with the specific instructions of voice command, then the soundcard can be called an inclusive design supported soundcard, designed for controlling games with specific voice commands. Game designers then only need to integrate their control structure of the game with that hardware. There are two advantages of implementing the results in hardware form. First, it can be highly effective at replacing relatively complex discreet logic. Secondly, if a probability-based algorithm can be implemented and programmed in the chip, it may replace the idea of even using a speech recognition system as long as we focus on certain voice input parameters. The memory type in the PIC is EPROM, hence it can alter the data in the memory and retain its value, even when the power is removed. The company (Microchip technology) that designed the PIC provides a freeware IDE package called MPLAB, which can be used to program the PIC. Using "Programmer", which is hardware designed to configure programmable non-volatile circuits such as EPROM, a PIC can easily be programmed with certain instructions to accomplish the desired task.

It is now understood that the hardware implementation of the proposed system has several constructive issues that will take place. First, the overall system will be more compact and integrated, and the command execution time will be reduced. A speech recognition system might not be required for pattern matching of words from a vocabulary array, as it can be replaced by the stylish use of algorithms to find selected commands where the algorithm comes to a decision on the instruction based on the frequency level and intensity values from the user.

This can lead to an industrial and engineering challenge. While the hardware satisfies certain voice commands in built-in format, the game industries are then going to be under pressure to integrate with the service from such hardware with their designed games. Thus, a universal design standard can be produced and games can be benchmarked in terms of their inclusivity, as well as being standardized in some structure.

6. DISCUSSION

The performance of the proposed system depends on how the algorithm is implemented. If it is a clean code-based controller, it will be interesting to observe the run time to evaluate the system's performance. Using the frequency level to detect and generate commands, the proposed model can be made more sophisticated, while issues such as the word error rate, the single word error rate and the command success rate will be verified and fine tuned.

One of the impediments faced by the elderly is not having grown up with computers or computer languages and when confronted with the technological terminology frequently used in both handbooks and interfaces, some feel overtaxed [2]. Interfaces based on the experience, language and expectations of this group, which offer to support their special requirements while taking physical deficiencies into account, can supply the necessary incentives [2]. Metrics for the evaluation of trustworthiness and the acceptance of passive technology for the elderly must be approached from the view of the elderly, which can be successfully achieved by working together with the end users [2]. Based on the analogies between user anxiety and metrics by A. Holzinger et al. [2], the following table shows the sample evaluation metrics of the voice- controlled inclusive games that can be designed by the proposed model.

Table **2 -** Evaluation metrics of the proposed voice control system for controlling games

Questions	**Resulting Metric**
Can I trust it?	How to make the control system more trustworthy?
Is it easily learnable?	How to improve the learning experience, without too many confusing details?
Can I switch between voice and regular control?	How to make the proposed system controllable?
Will the voice command system obey me?	How to give more confidence to the user of the control system?
Is this control system for me?	Explain the benefits, purpose and appropriateness to the users

The use of the Viterbi algorithm is reasonable as long as it does not increase the cost of the overall algorithm. As the system will mostly be dealing with old or disabled people, decoding the speech could sometimes be difficult when the system is presented with a new utterance and it must then compute the most likely source word. It is therefore prudent to use the Viterbi algorithm to find the best path, which can be an added advantage in this system. Using a fuzzy logic controller between the ASRS system and game controller is definitely going to make the command execution process faster. However, in the proposed fuzzy controller, some issues have to be considered in terms of performance. The measurement of uncertainty is the first issue in which the classical rules to find the max or min value have been used. Second, the importance of the measurement was considered, i.e. the uncertainty measurement rule could be altered by the value of importance of the members in the fuzzy set, which may not just give the priority to the maximum or minimum value in this case. The quality measurement rule was also considered. The quality of a rule is measured by multiplying the value obtained by a rule with the quality value assigned to a rule. The quality value assigned to a rule can be 0 or 1, where 0 describes a rule to be lower quality and 1 describes the rule to be of a good quality.

7. CONCLUSION

The use of speech to control games is only one step towards achieving inclusiveness, and thereby universal design and accessibility in games. Other human modalities combined with speech will give more optimal performance, overcoming any drawbacks of only using speech. The result from this paper has multiple possibilities in game design and the computer industry. While any existing algorithm can be altered and new algorithms can be created for detecting user emotions for executing one command in different ways in game playing, which is important for elderly adults or disabled people, an electronic chip can be programmed and integrated with a sound card (for instance a PIC), substantiating the lasting universal design of certain standards of voice commands for playing games. This opens up an enormous opportunity for both hardware manufacturers and the game industry. Voice supports for the game based on this paper's idea can therefore show the way towards finding an industry standard for the universal design of computer games.

8. REFERENCES

1. Crawford, C., *The Art of Computer Game Design.* 1982.
2. Holzinger, A., et al., *Investigating Usability Metrics for the Design and Development of Applications for the Elderly*, in *Computers Helping People with Special Needs*, K. Miesenberger et al., Editors. 2008, Springer Berlin/Heidelberg. pp. 98-105.
3. Miesenberger, K., et al., *More Than Just a Game: Accessibility in Computer Games*, in *HCI and Usability for Education and Work*, A. Holzinger, Editor. 2008, Springer Berlin / Heidelberg. pp. 247-260.
4. Law, E., et al., *A Mixed-Method Approach on Digital Educational Games for K12: Gender, Attitudes and Performance*, in *HCI and Usability for e-Inclusion*, A. Holzinger and K. Miesenberger, Editors). 2009, Springer Berlin/Heidelberg. pp. 42-54.
5. Langdon, P., J. Clarkson, and P. Robinson, *Designing accessible technology.* Universal Access in the Information Society, 2007. **6**(2): pp. 117-118.
6. Czaja, S. and C. Lee, *The impact of aging on access to technology.* Universal Access in the Information Society, 2007. **5**(4): pp. 341-349.
7. Eurostat, *Eurostat, Internet usage in the EU25 in 2005.* 2006.
8. Clarkson, J. and R. Coleman, Keates, S., Lebbon, C. (eds.), *Inclusive Design: Design for the Whole Population.* 2003: pp. 88-102.
9. Goodman, J., Langdon, P.M., Clarkson, P.J., *Equipping designers for inclusive design. Gerontechnology.* 2006(4): pp. 229-233.
10. Goodman-Deane, J., S. Keith, and G. Whitney, *HCI and the older population.* Universal Access in the Information Society, 2009. **8**(1): pp. 1-3.
11. Hawthorn, D., *Possible implications of aging for interface designers. Interact. Comput.* 2000. **12**(5): pp. 507-528.
12. Olivier, D.A.a.A.B.a.S.L.a.D., *Adapting mainstream multimedia games for severely visually impaired children.* GAMEON, 2004 (November).

13. Holzinger, A., *Universal access to technology-enhanced learning.* Universal Access in the Information Society, 2008. **7**(4): pp. 195-197.
14. Neerincx, M., et al., *Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice.* Universal Access in the Information Society, 2009. **8**(2): pp. 109-122.
15. Carbonell, N., *Multimodality: a step towards universal access.* Universal Access in the Information Society, 2003. **2**(2): pp. 89-90.
16. Debevc, M., P. Kosec, and A. Holzinger, *Improving multimodal web accessibility for deaf people: sign language interpreter module.* Multimedia Tools and Applications, 2010: pp. 1-19.
17. Picard, R.W., *Affective Computing.* MIT Press, Cambridge 1997.
18. Hinckley, K., *Input technologies and techniques.* Lawrence Erlbaum, 2003.
19. van Leeuwen, D.A., Martin, A.F., Przybocki, M.A., Bouten, J.S., *NIST and TNO-NFI evaluations of automatic speaker recognition, Computer.* Speech Language, 2006. **20**: pp. 128–158.
20. Van Wijngaarden, S.J., Smeele, P.M.T., Steeneken, H.J.M., *A new method for testing communication efficiency and user acceptability of speech communication channels.* Proceedings of Eurospeech 2001, Aalborg, September 2001: pp. 1675–1678.
21. Stickel, C., et al., *Emotion Detection: Application of the Valence Arousal Space for Rapid Biological Usability Testing to Enhance Universal Access*, in *Universal Access in Human-Computer Interaction. Addressing Diversity*, C. Stephanidis, Editor. 2009, Springer Berlin / Heidelberg. pp. 615-624.
22. Eddy, S.R., *Hidden Markov models.* Current Opinion in Structural Biology, June 1996. **6**(3): pp. 361-365.
23. Ying, H., *The simplest fuzzy controllers using different inference methods are different nonlinear proportional-integral controllers with variable gains.* Automatica, November 1993. **29** (6): pp. 1579-1589
24. Mizumoto, M. *Product-sum-gravity method=fuzzy singleton-type reasoning method=simplified fuzzy reasoning method*. in *Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on*. 1996.
25. Van Broekhoven, E. and B. De Baets, *Fast and accurate center of gravity defuzzification of fuzzy system outputs defined on trapezoidal fuzzy partitions.* Fuzzy Sets and Systems, 2006. **157**(7): pp. 904-918.