

Fuzzy-Logic Controller for Speaker-Independent Speech Recognition System in Computer Games

Moyen Mohammad Mustaquim

Department of Informatics and Media
Uppsala University
Uppsala, Sweden
moyen.mustaquim@im.uu.se

Abstract. Computer Games are now a part of modern culture. By using automatic speech recognition systems (ASRS), voice driven commands can be used to control the game, which can open up the possibility for people with disabilities and age related problems to be included in game communities and use the service offered. Conventional speech recognition systems however, do not support emotions, attitudes, tones etc. This is a problem because such expressions can be vital for gaming, especially as the majority of the users to be considered for this operation are elderly adult. For instance, if a command is uttered in a low pitch the character in the game can act differently than when it is voiced in an elevated pitch. The traditional voice controller for such operation is usually based on logic or a probability supported algorithm that can serve the purpose but has certain limitations like long process time, cost and complexity in implementation. This paper proposes a fuzzy logic based controller to detect a user's emotion from their voice command for controlling the game in a different ways. This fuzzy logic controller can be programmed even on a microchip to drive certain voice driven actions as a built-in form of hardware. To use a fuzzy logic controller, instead of an approach based on traditional algorithms, gives more accuracy, control and efficiency when handling the user's voice input commands for manipulation.

Keywords: Universal design of computer games, fuzzy logic, fuzzy controller, speech controlled computer games design.

1 Introduction

1.1 Computer Games in Society Today

Games are the vehicle with which society is and will change the computer. But how will the games themselves are changed by society? We can expect two processes to affect games: one is the mass market and another is the flowering of heterogeneity. In some ways, these processes work against each other [1].

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 anymore but games can be used to serve several purposes

for all kinds of people in the society. Today's technologically inclined people are going to be older tomorrow, and they would not like to be excluded from future services for instance games, just because of design issues associated with the game. Also, anyone can suffer temporary or permanent disabilities regardless of age and it would be unfortunate to exclude them from using games. Moreover, both the human death rates and birth rates are generally decreasing, which results in an age distribution with a large number of older people (disabled or not) in the society in the near future. So the game industries should be concerned with such issues, so that they can include these groups of people and offer them properly designed services.

1.2 Using Human Modalities in Game 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 human computer interaction (HCI) [2]. New input/output modalities and forms of multimodality are hence needed to provide easy and effective access for 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 integrating aspects of intelligent interactive behavior [2]. It is thus a wise approach to use different 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 otherwise providing input in a computer game. The controller is usually connected to a game console or a computer by means of a wire, chord and nowadays also by means of wireless communication. Typical game controllers are keyboards, mice, joysticks, game pads etc. There are also 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 the events in a video- or computer game. There is health concerns related to game controllers such as risk for injuries like repetitive strain injuries or carpal tunnel syndrome hence they need to be designed ergonomically to give best possible relaxation of 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 important to consider for such groups of people.

2.2 Speech Technology

Speech technology seems to provide new opportunities to improve 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 to see icons, to control a mouse or a keyboard [3]. Such limitations have both functional and emotional aspects that should be addressed in the design of user interfaces [4]. 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 [3].

2.3 Automatic Speech Recognition System

Automatic speech recognition (ASR) also known as speaker independent speech recognition is the automatic conversion of human speech to a sequence of words. The aim of ASR is to recognize automatically what has been said. Nowadays, ASR systems are generally based on the hidden Markov models (HMM) [12] 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 different recording conditions[3].

If a speech recognizer is part of an application, usually it is used as an input technology or technique [5]. Various other performance measures can reduce the word error rate, depending on the application. For instance, van Leeuwen et al.[6] mentions the following issues that are important to consider when assessing 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 [7].

Nowadays, speech recognition technology is used in many commercial applications. In these applications, shortcomings of the technology are reduced by tailoring the interaction to the individual user and context, for instance, by making the system speaker-dependent (e.g., a dictation system), by limiting the vocabulary (e.g., voice dialing) or by using a simple and predictable syntax (e.g., command and control, data-entry)[3]. In some applications, recognition accuracy does not need to be perfect, e.g., for spoken document retrieval a word recognition accuracy of 70% produces similar retrieval results as manually generated speech transcripts [8].

2.4 Requirements of a Voice Controlled System for Games

While the traditional approach of game control is using input devices, people with certain limitation of doing so would like to use other alternatives for control. People with motor system problems, limitations of moving the upper limbs, problems with hand muscles, hand joints, muscular stress, and many other neurological disorders, can be highly benefited by using voice driven command systems in gaming. Age related issues can make the hand of a person weak which is quite common and hence use of speech is reasonable for such users for game control. Also the use of speech does not require any special hardware. Ordinary computers of today are able to carry out speech transforming operations from the user to the process handler of the game, hence there is no need for any extra hardware and disabled people do not need to wear or use any special device to use the speech system for controlling games.

2.5 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 [13]. Fuzzy logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information and 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 logics of human control. It can be built into anything from small, hand-held products to large computerized process control systems. It uses an imprecise but very descriptive language to deal with input data 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 only little or no tuning.

2.6 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. The conventional control systems work either in a linear or a non-linear manner. Linear systems are simple and well explored but not too flexible. The non-linear systems on the other hand can be powerful but are often sensitive to modeling errors. A fuzzy controller is effective in situations where the control process is too complex to analyze by conventional quantitative techniques and also when the available sources of information are interpreted qualitatively, inexactly or uncertainly. A fuzzy logic controller has several advantages over a traditional controller. 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, 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 and where traditional strategies are too complex or non-linear to be controlled. Figure 1 shows the basic structure of a fuzzy logic controller.

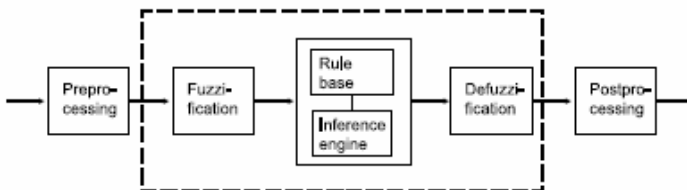


Fig. 1. Structure of a fuzzy controller

2.7 Traditional Approach of Detecting Emotion from Voice

The traditional control system to detect emotions in voice control systems for games consists of a standard automatic speech recognition system where the vocabulary array is built by fetching the commands from the most commonly used game platform. The user input is processed through the soundcard which converts the analog data to digital format and then checks with the user word model from the vocabulary register. Once the word uttered by the user matches with the word from the vocabulary list, it then sends the binary signal corresponding to that word to the process handler of the game. Users see that the action has been taken according to their given voice command, through the GUI of the game. Figure 2 shows the block diagram of such conventional system.

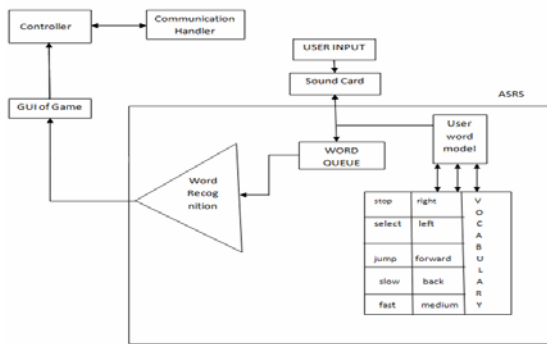


Fig. 2. Common speech recognition system for game control

3 Proposed System- The Fuzzy Controller

The proposed system replaces the traditional system’s control part as shown in Figure 2, by a fuzzy-logic controller. There exists linguistic terms which are the content of the fuzzy set describing the command actions. The rule input commands are processed in to output. The fuzzy controller is described in the following subsection. The proposed fuzzy controller’s block diagram is shown in Figure 3. The fuzzification process involves handling voice command variables (d_1, d_2, \dots, 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 velocity and the linear velocity of an object. Hence state variables (d_1, d_2, \dots, d_n), are decided with the error and change of error by the user’s voice input.

The proposed system replaces the traditional system’s control part as shown in Figure 2, by a fuzzy-logic controller. There exists linguistic terms which are the content of the fuzzy set describing the command actions. The rule input commands are processed in to output. The fuzzy controller is described in the following subsection. The proposed fuzzy controller’s block diagram is shown in Figure 3. The fuzzification process involves handling voice command variables (d_1, d_2, \dots, 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 velocity and the linear velocity of an object. Hence state variables (d_1, d_2, \dots, 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. The controlled variables ω and v are decided, as said before to determine the angular and linear velocity of an object which is important to consider for the different movement based on the user-given voice command. The fuzzification process converts the crisp input values to fuzzy set values.

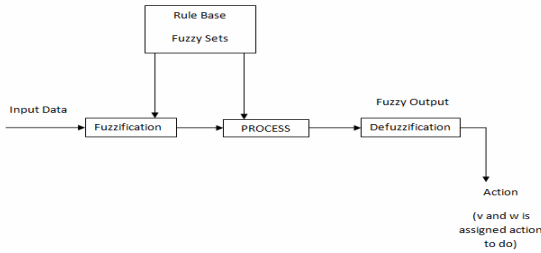


Fig. 3. Proposed fuzzy logic controller system

An inference method was chosen to design this fuzzy controller which is the Mamdani inference method[11]. Also the fuzzy singleton [14] fuzzification method is selected as it measures the state variables without uncertainty. The inference process defines the connective implication and rule combination operations. The controller uses minimum connectives 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, reaching the same conclusion.

Output variables (ω and v) are also modeled as discrete fuzzy set as it increases speed in inference and defuzzification process. Figure 4 shows the shape of fuzzy input sets where $d_1, d_2, d_3, \dots, d_n$ are considered to be the voice sensor data received from the user. The angle θ is used as a guidance operator.

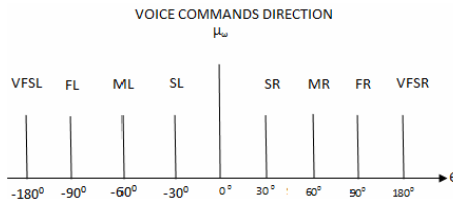


Fig. 4. Shape of fuzzy input sets deciding direction of object

That is, the way d is sensed from the user input, triggers the value of \square and puts a decision value on the output variable describing at which angle the object should move. The value of \square is discretized and normalized within the range of $[-180, 180]$.

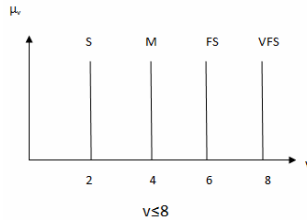


Fig. 5. Shape of fuzzy output sets for linear velocity

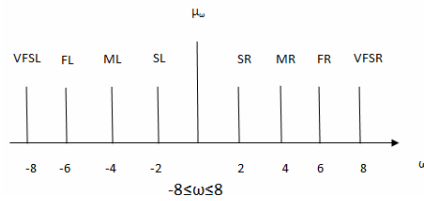


Fig. 6. Shape of fuzzy output sets for angular velocity

Figure 5 and figure 6 shows the shape of fuzzy output sets where limit of v and ω are discretized and normalized within the range of $[0, 8]$ for v and $[-8, 8]$ for ω .

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 are linguistic terms which is 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) [13] method is used for defuzzification strategy, as in this problem it results less processing time with no degradation of defuzzified value. Also COG method simplifies the computation of inference mechanism.

4 Proposed System - The Fuzzy Controller

4.1 Linguistic Terms

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

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

4.2 Sum-Product Interference

Degree of each rule is determined by the following rule:

$\beta_l = \min(\mu_{A_i}(d_i), \dots, \mu_{A_k}(d_k))$. A_i are the input fuzzy sets that is the defined linguistic terms and d_j ($j=1 \dots k$) are the input variables.

Output fuzzy set B_l' is constructed using 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.

Combination of output fuzzy sets to single fuzzy set B' :

Sum operation: $\mu_{B'}(v) = \mu_{B_1'}(v) + \dots + \mu_{B_N'}(v)$

4.3 Construction of Fuzzy Rules

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

Turn Left:

If θ is SCT and d_0 is L and d_1 is SCT Then v is ST and ω is L

Stop:

If θ is MV and d_0 is SCT and d_1 is ST Then v is ST and ω is ST

Right Turn Fast:

If θ is SL and d_0 is R and d_1 is FR and d_2 is MV Then v is ST and ω is FR

Select and Move Forward

If θ is NS and d_1 is SCT and d_2 is FW and d_3 is MV Then v is MV and ω is ST

5 Discussion

Although vocal interaction has technological limitations, these might be overcome when used as a component of a multimodal user interface. Such interfaces seem to have higher levels of user preference, among other things, because people experience a greater degree of flexibility [9]. The performance of the proposed fuzzy controller depends on how accurately it is implemented and fine tuned. The proposed fuzzy controller can be made more sophisticated while issues like word error rate, single word error rate and command success rate will be verified and fine tuned while combined with any algorithm such as HMM [10]. Use of Viterbi algorithm [12] is reasonable as long as it does not increase the cost of the overall controller. As the system will mostly be dealing with old or disabled people, decoding of the speech

might be difficult sometimes, especially when the system is presented with a new utterance and it must then compute the most likely source word. Use of the Viterbi algorithm is thus prudent to find the best path which can be an added advantage in this system.

However in the proposed fuzzy controller, some issues have been considered in terms of performance. The measurement of uncertainty is the first issue where the classical rules to find the max or min value has been use. Second, the importance measurement was considered, that is the uncertainty measurement rule might be altered by the value of importance of the members in 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.

6 Conclusion

The use of speech to control games is only one step towards inclusiveness and thereby universal design and accessibility in games. The result from this paper, to use a fuzzy controller for designing voice controlled game has multiple possibilities in game design and PC industry. While any existing algorithm can be altered and new algorithm can be created for detecting user's emotions for executing one command in different way in game playing; a fuzzy-logic based controller programmed in chip and integrated with sound card, substantiate the lasting universal design of certain standard of voice command for playing game. This enables and opens enormous opportunity both for hardware manufacturers and game industries. Voice support for game, based on this paper's idea can show the way towards finding an industry standard of universal design in computer gaming in an optimal way.

References

1. Crawford, C.: *The Art of Computer Game Design* (1982)
2. Carbonell, N.: Multimodality: a step towards universal access. *Universal Access in the Information Society* 2(2), 89–90 (2003)
3. Neerinx, M., Cremers, A., Kessens, J., van Leeuwen, D., Truong, K.: Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice. *Universal Access in the Information Society* 8(2), 109–122 (2009)
4. Picard, R.W.: *Affective Computing*. MIT Press, Cambridge (1997)
5. Hinckley, K.: *Input technologies and techniques*. Lawrence Erlbaum, Mahwah (2003)
6. 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* 20, 128–158 (2006)
7. 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. In: *Proceedings of Eurospeech 2001, Aalborg*, pp. 1675–1678 (September 2001)

8. Ordelman, R.J.F., de Jong, F., Huijbregts, M.A.H., van Leeuwen, D.A.: Robust audio indexing for Dutch spoken-word collections. In: Proceedings of the XVIth International Conference of the Association for History and Computing (AHC 2005), Amsterdam, The Netherlands, September 14-17 (2005)
9. Oviatt, S.: User-centered modeling and evaluation of multimodal interfaces. *IEEE* 91(9), 1457–1468 (2003)
10. Eddy, S.R.: Hidden Markov models. *Current Opinion in Structural Biology* 6(3), 361–365 (1996)
11. Ying, H.: The simplest fuzzy controllers using different inference methods are different nonlinear proportional-integral controllers with variable gains. *Automatica* 29(6), 1579–1589 (1993)
12. <http://www.iem.uni-due.de/~vinck/data-com/The%20viterbi%20algorithm.ppt>
13. <http://farfromneutral.com/kaizen/center-of-gravity-method/comment-page-1/>
14. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/jp6/article2.html