

## **A Review on Hand And Speech Based Interaction With Mobile Phone**

Ms. S. B. Maind, Prof. A. V. Dehankar

*Department of Computer Technology*

*Priyadarshini College of Engineering Nagpur*

*Nagpur, India*

[sonali.maind@gmail.com](mailto:sonali.maind@gmail.com), [archana\\_dehankar@rediffmail.com](mailto:archana_dehankar@rediffmail.com)

**ABSTRACT:** In the recent years new methods of HumanComputer Interaction (HCI) are being developed. Out of this some of them are based on interaction with machines through hand gesture, head, facial expressions, speech, touch and there are still various current topic of research. But just relying on one of them reduces the accuracy of the whole HCI and is also limiting the options available to users. Although gesture-based interaction technology is used in many areas such as robot control, navigation system, medical research, it has not fully embedded into our daily life. In today's society, one of the most popular electronic products is mobile phone. Using mobile phones is a top priority for anyone living in the world, from young to old. Therefore, we want to combine gesture-based interaction technology and mobile phones. The objective of this paper is to use two of the important modes of interaction: hand and speech to control some mobile application

**KEYWORDS:***HCI (Human Computer Interaction), Artificial Neural Network, Microsoft Speech SDK.*

### **I. INTRODUCTION**

In present-day society, with the development of computer science, the hardware and software are very advanced. But the interaction between human and machine is developing slowly and even become an big problem to the inherent technological development. Therefore, a large number of researchers start to divert their attention from computer screen or programming language to human-computer interaction (HCI). Gestures are being used in HCI since many years. Earlier, hardware based gesture recognition was more prevalent. User had to wear gloves, helmet and other heavy apparatus. Sensor, actuator and accelerometer were used for gesture recognition. But the whole process was difficult in real time environment. In today's society, one of the most popular electronic products is mobile phone. Using mobile phones is a top priority for anyone living in the world, from young to old. Therefore there is a need of combining gesture-based interaction technology and mobile phones. The objective of this paper is to develop a multimodal human mobile interaction method where hand and speech can be used to control some mobile application. To achieve this objective various phases are require. In this for obtaining the result from speech recognition some tools are used. And also for hand gesture recognition some methods are used. After getting the input from speech and hand gesture method, system will combine this both input to control some mobile application.

### **II. WHY MULTIPLE MODALITIES**

As we know the interaction of humans with their environment is naturally multimodal. For example, we speak about, point at, and look at objects all at the same time. Now thinking about HCI, we usually use only one interface device at a time. But this singly modal interaction allows us to convey our message to the computer is not that much satisfactory. Relying on just one of the modes of interaction reduces the accuracy of the whole HCI 7 also limit the options available to users.

There are some disadvantages of using single modality:

1. Strict restrictions needed for accurate & robust interaction with the individual modes.
2. Unnatural & difficult to handle.
3. Lacks robustness & accuracy.

To avoid the above drawbacks of using single modality, multiple modalities are being used. Advantages of using multiple modalities are as follows:

1. It increases naturalness of the interface for Human computer interaction.
2. Use of two or more interaction modalities may loosen the strict restrictions needed for accurate & robust interaction with the individual modes.
3. It also enables easy access to physically or cognitively handicapped people.
- 4.

### III. HAND GESTURE RECOGNITION BY USING ARTIFICIAL NEURAL NETWORK

An Artificial neural network is an information processing system that has certain performance characteristics in common with biological neural networks. It consists of an interconnected group of artificial neurons and processes information using connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

For hand gesture recognition two types of algorithms are used i.e. Backpropagation algorithm & Feedforward algorithm.

#### 1. Backpropagation algorithm :

Back Propagation is the training or learning algorithm used for training Artificial neural networks. It is most useful for feed-forward networks. The term is an abbreviation for “backward propagation of errors”. As the algorithm’s name implies, the errors propagate backwards from the output nodes to the inner nodes. It is used to calculate the gradient of the error of the network with respect to the network’s modifiable weights. Back propagation networks are basically multilayer. In order for the hidden layer to serve any useful function, multilayer networks must have non-linear activation functions for the multiple layers.

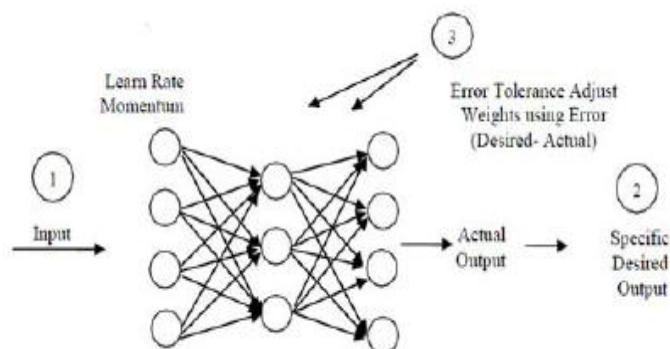


Fig.1: Backpropagation Network

#### 2. Feedforward algorithm:

The Feedforward neural network is the simplest type of artificial neural network. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. Fig below shows a representation of a simple feed-forward neural network with four inputs, one hidden layer and four outputs.

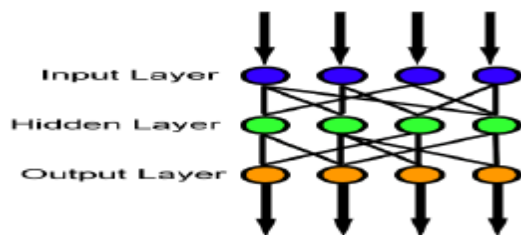


Fig.2: Feedforward Network

The hand gesture recognition process consists of two phases:

1. Training or learning phase.
  2. Testing or classification phase.
- Training phase involve following steps:

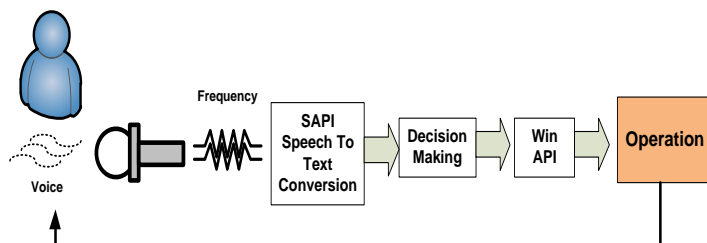
1. Image capturing form training set.
2. Segmentation.
3. Noise Reduction.
4. Edge detection
5. Feature vector computing
6. Saving feature image as a training pattern.
7. Learning using neural networks with back-propagation algorithm.

Testing phase involve following steps:

1. Image capturing from testing set
2. Segmentation.
3. Noise reduction
4. Edge detection
5. Feature vector computing
6. Recognized by neural networks.

#### IV. SPEECH RECOGNITION

For speech recognition purpose we are using Microsoft Speech SDK.



**Microsoft Speech SDK:** These SDK are developed by Microsoft for building speech based application. It has predefined set of speech frequencies which will calculate the frequency graph and returns input signals in the form of text or in the form of speech.

For hand recognition image will capture by using web camera and process each frame as a input and calculate the desired value as per the system required once the system get the desired value it will send signal to processing unit. After getting the hand and speech input, the fusion of input will be send to the processing unit where it will control some mobile applications as per defined.

SAPI(Speech Application Programming Interface):

It is used to develop speech related functionality software. It is not a part of language, but can access through language provided by Microsoft.

It has some functionality: Text to Speech, Speech to Text conversion.

## V. FUSION

Combining multiple modalities or the input from human to perform single or multiple task is called as fusion technique.

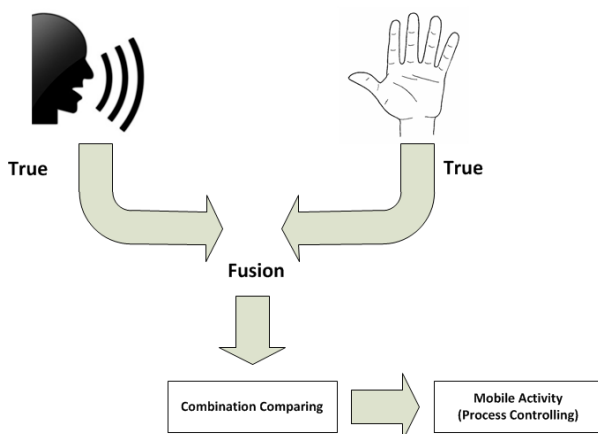
Basically there are three types of fusion: 1. Data fusion, 2. Feature fusion, 3. Decision feature.

Data fusion involves integration of raw observation and can occur only in the case when the observations are of same type. They are not use for multimodal HCI. For example, it is occur when one or more cameras are used to capture visual information on one object.

Feature fusion assumes that each stream of sensory data is first analyzed for features, after which the features themselves are fused. For example, it is appropriate for closely coupled 7 synchronized modalities, possibly speech & lip movement.

Decision fusion is based on the fusion of individual mode decisions or interpretations. For example, once an arm movement is interpreted as a pointing gesture and a spoken sentence is recognized as "Make this box white". The two can be fused to interpret that a particular object needs to be painted white.

Figure below shows the decision level fusion of hand gesture and speech for controlling the mobile activity.



## VI. CONCLUSION

By using hand and speech together, there is a significant increase in the number of gestures and also now the gestures are more intuitive and user friendly. The accuracy of combined multimodal gesture recognition system is increased in comparison to the case when hand and speech gestures were used individually. This application for mobile phones can be very useful for developing countries because

there is large amount of mobile phone. We can handle number of mobile applications by using the combination of hand and speech.

#### **VII. ACKNOWLEDGMENT**

System will control some application of mobile phone by using hand and speech based interaction.

#### **REFERENCES**

1. AnupamAgrawal, Rohit Raj and ShubhaPorwal, "Vision-based multimodal human computer interaction using hand and head gesture". IEEE, 2013.
2. Siddharth S. Rautaray " Real Time Multiple Hand Gesture Recognition System for Human Computer Interaction,"2012.
3. Shweta K. Yewale et al. "Artificial Neural Network Approach For Hand Gesture Recognition", International Journal of Engineering Science and Technology (IJEST).
4. HaithamHasan · S. Abdul-Kareem, "Static Hand Gesture Recognition using Neural Network", springer-2012