CHAPTER - 1

INTRODUCTION TO HUMAN COMPUTER INTERACTION
1.1 Introduction to human computer interaction

Human Computer Interaction (HCI) involves the study, planning and design of the interaction between people and computer. It is often regarded as the intersection of computer science, behavioral sciences, design and several other fields of study. The term was popularized by Card, Moran, and Newell in their seminal 1983 book, "The Psychology of Human-Computer Interaction". Because human–computer interaction studies a human and a machine in conjunction, it draws from supporting knowledge on both the machine and the human side. On the machine side, techniques in computer graphics, operating systems, programming languages, and development environments are relevant. On the human side, communication theory, graphic and industrial design disciplines, linguistics, social sciences, cognitive psychology, and human factors such as computer user satisfaction are relevant. Engineering and design methods are also relevant. Due to the multidisciplinary nature of HCI, people with different backgrounds contribute to its success. HCI is also sometimes referred to as Man–Machine Interaction (MMI) or Computer Human Interaction (CHI).

The reason, in fact, is clear: most sophisticated machines are worthless unless they can be used properly by man. This basic argument simply presents the main terms that should be considered in the design of HCI: functionality and usability [1]. Functionality of a system is defined by the set of actions or services that it provides to its users. However, the value of functionality is visible only when it becomes possible to be efficiently utilized by the user [2]. Usability of a system with a certain functionality is the range and degree by which the system can be used efficiently and adequately to accomplish certain goals for certain users [3, 70]. The actual effectiveness of a system is achieved when there is a proper balance between the functionality and usability of a system.

Having these concepts in mind and considering that the terms computer, machine and system are often used interchangeably in this context, HCI is a design that should produce a fit between the user, the machine and the required services in order to achieve a certain performance both in quality and optimality of the services [4]. Determining what makes a certain HCI design good is mostly subjective and context dependent.
In March 2007, Microsoft Research organized the ‘HCI 2020’ meeting at the El Bulli Hacienda Hotel near Seville, Spain. The event’s title expressed its key question: what will Human-Computer Interaction (HCI) be like in the year 2020? That question is important because HCI, significant as it was in the late 20th century, has a pivotal part to play in the 21st, when computers will become so pervasive that how humans interact with them will be a crucial issue for society. HCI 2020 produced many ideas, both thrilling and troubling. The great accomplishment of HCI has been, to date, that it allows investigations of matters beyond what one might call the mechanics of the interface, such things as the design of the graphical user interface, and of keyboards and of mice. Its success now allows researchers to focus on how computers can support human-to-human concerns, rather than simply human-machine interaction. HCI has helped to produce a world in which interacting with computers is easier and richer. The real HCI issues now include what might be our aspirations, our desires for self-understanding and expression, and our willingness to use imagination to create a different future [5].

1.2 Gradual growth of HCI in society:

The growth in Human-Computer Interaction (HCI) field has not only been in quality of interaction, it has also experienced different branching in its history. There have been various computer-driven revolutions in the past. It may include the computers from 1960 to 2000. A diagrammatic representation is as shown in Figure 1.1

In 1960s one mainframe was used by many users. But the situation changes in 1980s. Desktop PCs are mostly used by the users for different purpose like billing in shops, keeping records, etc. In 2000 a single user is connected with a number of computers for doing their work. Now mobility is occurred in 2000 & user can do his work from any place. But we are looking in the near future i.e. is in 2020. It may look like the 4th image as shown in Figure 1.1. We are expecting such a change in HCI.

At the start of the 21st century, HCI was an interdisciplinary field which has undergone enormous changes. In terms of a science or a discipline, these changes have occurred over a very short time. HCI now encompasses many philosophies, perspectives and types of expertise. There are multiple and overlapping groups of researchers, some
emphasizing design, others evaluating, and yet others user modeling. These experts all work within a complex space, each examining different aspects of human-computer interaction. Different techniques are used, depending on different goals.

Figure 1.1: Illustration of Changes in HCI from 1960s to 2000 and expecting 2020 being as vivid as 4th image [5]
1.3 Existing HCI technologies

HCI design should consider many aspects of human behaviors and needs to be useful. The complexity of the degree of the involvement of a human in interaction with a machine is sometimes invisible compared to the simplicity of the interaction method itself. The existing interfaces differ in the degree of complexity both because of degree of functionality/usability and the financial and economical aspect of the machine in market. For instance, an electrical kettle need not to be sophisticated in interface since its only functionality is to heat the water and it would not be cost-effective to have an interface more than a thermostatic on and off switch. On the other hand, a simple website that may be limited in functionality should be complex enough in usability to attract and keep customers. Therefore, in design of HCI, the degree of activity that involves a user with a machine should be thoroughly thought [5].

The user activity has three different levels: physical [7], cognitive [8], and affective [9]. The physical aspect determines the mechanics of interaction between human and computer while the cognitive aspect deals with ways that users can understand the -system and interact with it. The affective aspect is a more recent issue and it tries not only to make the interaction a pleasurable experience for the user but also to affect the user in a way that makes user continue to use the machine by changing attitudes and emotions toward the user. The focus of this chapter is mostly on the advances in physical aspect of interaction and to show how different methods of interaction can be combined (Multi-Modal Interaction) and how each method can be improved in performance (Intelligent Interaction) to provide a better and easier interface for the user.

1.4 Overview on HCI

Technology is changing, people are changing, and society is changing. All this is happening at a rapid rate. What can the HCI community do to interfere and help? How can it build on what it has achieved? HCI needs to extend its methods and approaches so as to focus more clearly on human values. This will require a more sensitive view about the role, function and consequences of design, just as it will force HCI to be more inventive. HCI will need to form new partnership with other disciplines, too, and for this to take place HCI
practitioners will need to be sympathetic to the tools and techniques of other trades. Finally, HCI will need to re-examine and reflect on its basic terms and concepts. Outdated notions of the ‘user’, the ‘computer’ and ‘interaction’ are hardly sufficient to encompass all that HCI will need to attend to.

The advances made in last decade in HCI have almost made it impossible to realize which concept is fiction and which is and can be real. The thrust in research and the constant twists in marketing cause the new technology to become available to everyone in no time. However, not all existing technologies are accessible and/or affordable by public.

1.5 HCI systems architecture

Most important factor of a HCI design is its configuration. In fact, any given interface is generally defined by the number and diversity of inputs and outputs it provides. Architecture of a HCI system shows what these inputs and outputs are and how they work together. Following sections explain different configurations and designs upon which an interface is based.

A. Unimodal HCI systems

B. Multimodal HCI system

1.5.1 Unimodal HCI systems

As mentioned earlier, an interface mainly relies on number and diversity of its inputs and outputs which are communication channels that enable users to interact with computer via this interface. Each of the different independent single channels is called a modality. A system that is based on only one modality is called unimodal. Based on the nature of different modalities, they can be divided into three categories:

1. Visual-based:

The visual based human computer interaction is probably the most widespread area in HCI research. Considering the extent of applications and variety of open problems and
Chapter 1. Introduction to Human Computer Interaction

approaches researchers tried to tackle different aspects of human responses which can be recognized as a visual signal. Some of the main research areas in this section are as follow:

• Facial Expression Analysis
• Body Movement Tracking (Large-scale)
• Gesture Recognition
• Gaze Detection (Eyes Movement Tracking)

Figure 1.2: The Reactable: a multitouch interface for playing music

Figure 1.3: Gaze detection pointing system for people with disabilities
2. Audio-based:

The audio based interaction between a computer and a human is another important area of HCI systems. This area deals with information acquired by different audio signals. While the nature of audio signals may not be as variable as visual signals but the information gathered from audio signals can be more trustable, helpful, and in some cases unique providers of information. Research areas in this section can be divided to the following parts:

- Speech recognition
- Speaker recognition
- Auditory emotion analysis
- Human-made noise/sign detections (gasp, sigh, laugh, cry, etc.)
- Musical interaction

Figure 1.4: Speaker recognition
3. Sensor-based:

This section is a combination of variety of areas with a wide range of applications. The commonality of these different areas is that at least one physical sensor is used between user and machine to provide the interaction. These sensors as shown below can be very primitive or very sophisticated.

1. Pen-based interaction
2. Mouse & Keyboard
3. Joysticks
4. Motion tracking sensors and digitizers
5. Pressure sensors
6. Taste/smell sensors

Some of these sensors have been around for a while and some of them are very new technologies. Pen-Based sensors are specifically of interest in mobile devices and are related to pen gesture and handwriting recognition areas.

![Figure 1.5: Microsoft's 'Surface'](image)

Microsoft’s surface is an interactive tabletop allowing two-handed interaction with digital objects such as photos, music files, games and maps. These kinds of interactive surfaces encourage collaborative, creative engagement [5]. The next sub-sections describe each category and provide examples and references to each modality.
1.5.2 Multimodal HCI systems

The term multimodal refers to combination of multiple modalities. In MMHCI systems, these modalities mostly refer to the ways that the system responds to the inputs, i.e. communication channels.

The definition of these channels is inherited from human types of communication which are basically his senses: sight, hearing, touch, smell, and taste. The possibilities for interaction with a machine include but are not limited to these types.

A multimodal interface acts as a facilitator of human-computer interaction via two or more modes of input that go beyond the traditional keyboard and mouse. The exact number of supported input modes, their types and the way in which they work together may vary widely from one multimodal system to another.

Multimodal interfaces incorporate different combinations of speech, gesture, gaze, facial expressions and other non-conventional modes of input. One of the most commonly supported combinations of input methods is that of gesture and speech [6]. Figure 1.6 shows the multimodal HCI system.
1.6 Pros and cons compared to other biometrics

There are several types of biometrics methods of authentication and identification. These include fingerprints, face recognition, speaker verification, iris recognition, hand and finger geometry and signature verification [12].

**Fingerprints** of a person, including identical twins, are unique. It has therefore been used for a long time in law enforcement as a tool of biometric identification. Fingerprint systems can be used in both authentication and identification mode [12]. The biggest security issue with these systems is the vulnerability to artificial prints, "gummy fingers", which are cheap to produce and work well on both optical and capacitive sensors [14].

**Face recognition** uses facial images to identify subjects. There are several ways including the use of a normal camera using the visible spectrum or by the use of infrared cameras to capture the facial heat emission patterns. The biggest issues are detection of masks or photographs and handling the impact of lighting on the subjects face [12].

**Iris recognition** uses the unique features of a person's iris to identify him. Today's systems are accurate enough to work even in presence of eyeglasses or contact lenses. It works well for both authentication and identification purposes. Both face recognition and iris recognition have the advantage of not demanding any physical contact for recognition unlike fingerprint authentication systems [12].

**Hand and finger geometry identification** systems are similar to fingerprint systems but they measure physical characteristics like length, width, thickness and surface area instead [12].

**Signature verification** systems use a person’s written signature for authentication. Speed, pressure and angle of the signature are produced to determine the validity of identity. Authentication systems like this are mostly used in e-business applications [12].

**Speaker recognition** uses the unique anatomy of an individual’s throat and mouth to identify them by their voice [10]. The biggest advantage compared to all other biometric systems of identification and authentication is that it is the only biometric system that works over a telephone. Compared to other methods it's cheap because the major cost only comes
from the software being used [11]. On the other hand there are a lot of problems to be handled in speaker recognition. One of the biggest problems is the variability produced by the talkers themselves and the variability produced by the transmitting and recording channels [15].

1.7 HCI through speech recognition system

Speech recognition applications are becoming more and more useful nowadays. Various interactive speech aware applications are available in the market. But they are usually meant for and executed on the traditional general-purpose computers. With growth in the needs for embedded computing and the demand for emerging embedded platforms, it is required that the speech recognition systems (SRS) be available on them too. Speech recognition basically means talking to a computer, having it to recognize what we are saying and lastly, doing this in real time. This process fundamentally functions as a pipeline that converts PCM (Pulse Code Modulation) digital audio from a sound card into recognized speech. The elements of the pipeline are given below.

![Figure 1.7: Block diagram of a speech recognizer [13]](image)

1. In the first step feature analysis is done i.e. the input to speech recognizer is in the form of a stream of amplitudes, sampled at about 16,000 times per second. But audio in this form is not useful for the recognizer. Hence, Fast-Fourier transformations are used to produce graphs of frequency components describing the
sound heard for $1/100\text{s}$ of a second. Any sound is then identified by matching it to its closest entry in the database of such graphs, producing a number, called the “feature number” that describes the sound.

2. In the second step unit matching system provides likelihoods of a match of all sequences of speech recognition units to the input speech. These units may be phones, diphones, syllables or derivative units such as acoustic units.

3. Third step is lexical decoding which constraints the unit matching system to follow only those search paths sequences whose speech units are present in a word dictionary.

4. In fourth step apply a "grammar" so the speech recognizer knows what phonemes to expect. This further places constraints on the search sequence of unit matching system. A grammar could be anything from a context-free grammar to full-blown English.

5. Finally in fifth step figure out which phonemes are spoken. This is quite dicey as different words sound differently as spoken by different persons. Also, background noises from microphone make the recognizer hear a different vector. Thus a probability analysis is done during recognition. A hypothesis is formed based on this analysis. A speech recognizer works by hypothesizing a number of different "states" at once. Each state contains a phoneme with a history of previous phonemes. The hypothesized state with the highest score is used as the final recognition result.

1.8 Problem statement

Speech recognition system plays an important role in human computer interaction due to its hand free computing nature. It is difficult to design a perfect technique for the speaker recognition system in all environments. The amount of work in Indian regional languages has not yet reached to a critical level to be used it as a real communication tool, as already done in other languages in developed countries. Thus, this work was taken to focus on Marathi language. It is important to see that whether speech recognition system for Marathi can be carried out in similar pathways of research as carried out in English.
Chapter 1. Introduction to Human Computer Interaction

1.9 Objectives

The speech is the most prominent and natural form of communication between humans. Work on speech recognition system is now a wide range of research area. Research work is going on to make computers hear, understand and speak our natural language. These tasks can be covered in three aspects.

1. Speech recognition to allow the machine to catch words phrases and sentences that we speak.
2. Natural language processing to allow the machine to understand what we speak.
3. After understanding what has been spoken, allow the system to take a particular action.

Proposed work is carried out for the first aspect. Although a complete implementation of the first part would require that the computer be able to catch words and sentences from a continuous and natural speech over broad variations in accents, ages etc. To simplify the task, we focus on speaker dependent & isolated word recognition. We chose Marathi language for work because there is a lot of scope to develop systems using Indian languages which are of different variations. Some work has been already done in this direction in Bengali, Hindi and Telugu languages using isolated words.

The objective of this research is to analyze the Marathi language for speech recognition as well as for speaker recognition system.

1.10 Summary & thesis outline

This thesis is formally organized as follows:

Chapter 1: entitled “Introduction to Human Computer Interaction”, provides an overview of the man machine interface. Here the discussion is done on the development of HCI & different disciplines where HCI is used. Rather than this we discussed about the HCI architecture in which we focused on unimodal (visual, audio & sensor based system) & multimodal HCI system. This chapter gives the brief introduction about different biometric techniques. It also focuses on speech recognition biometric technique and research
objectives. This chapter includes formal description of the material to be covered in each of
the chapters.

Chapter 2: entitled the “Literature Reviews”, provides the valuable literatures by renowned
resources which are very much helpful for this research. In introduction different terms are
discussed regarding to speech recognition system. The processes of human speech
generation along with various features and techniques that have been identified and
developed for the purpose of speaker recognition are presented in the following sections.
Different feature extraction technique like Pitch counter, Formant, LDA, LPC, Cepstral
analysis, Spectral analysis etc. are discussed in detail. Also it includes different approach
which are used in speech recognition systems, namely acoustic phonetic approach, pattern
recognition approach, template based approach, dynamic time warping, knowledge based
approach, etc.. This chapter also focuses on database generation for this research work.
Chapter 2 gives the primary information about the international phonetic alphabets &
comparative study of Marathi akshara.

Chapter 3: entitled with the “Research Technique -I (Linear predictive coding for speech
recognition)”. In next step classification algorithm is plays important role for classifying the
data in a particular manner. After the feature extraction technique & classifier discussion
actual experimental work is done to obtain the results for two types of database. One is for
vowel classification & second is for speaker recognition. After using the proposed
methodology 98.7% recognition rate is obtained for vowel classification & 98.5%
recognition rate is obtained for speaker recognition. Though the LPC model is oldest one
and it is used from many years. Then as compared to other feature extraction techniques it is
works well.

Chapter 4: entitled as “Research Technique -II (Linear predictive coding & dynamic time
warping technique)”. This chapter focuses on fourier transforms FFT & DCT. Coefficients
obtained after DCT are provided to DTW to obtain a distance between two speech patterns
DTW is discussed & used for the spoken Marathi Isolated word. Steps took for the research
technique 2 is discussed. & finally observations are obtained from the experiment. In the
last we conclude that the differences between the same words are less as compared to
differences between the two different words.
Chapter 5: entitled as “Results & conclusions” depict the results obtained after experimental work. Results given in three forms i.e. vowel classification, speaker recognition & isolated word recognition. After the result conclusions and future work is stated in this chapter.

Chapter 6: in this chapter valuable references and appendices are included. Appendix-A It has some screen shots of the experiment conducted. Appendix-B consists of the details of published research papers related to this work.