Chapter 2

Advances in Interfacing through Speech and Brain Signals

2.1 Introduction

Speech Signal Processing makes possible to interact with computers through speech. Many speech recognition systems have already been developed for different applications [1]. However, there is lot of scope to develop systems using Indian languages which are of different variations. Some work is done in this direction in isolated Bengali words, Hindi and Telugu [2]. The amount of work in Indian regional languages has not yet reached to a critical level to be used as 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 cab be carried out following similar pathways of research as carried out in English [3]. In English language, the meaning of the sentences and words are clear and less ambiguous. The authenticity of correctness of language is a major problem in regional languages.

On other hand the communication using the electroencephalogram (EEG) signals between the men and machines represents one of the current challenges in signal theory research. The principle element of such communication system is known as “Brain Computer Interface (BCI)”, is the interpretation of the EEG signals related to the characteristic parameters of the brain electrical activity [4]. The role of EEG signal processing is crucial in the development of real-time Brain Computer Interface. The several improvements have been made in this area but none of them have been successful enough to use them in a real system. The goal of creating more effective classification algorithms, have focused numerous investigation in the search of few techniques of feature extraction [5]. The main objective of our work is regarding clustering of EEG signals on the basis of different types of activities.
2.2 Speech Signal Processing

Speech Processing is study of speech signal and processing method of these signals. It can be divided into various categories i.e. Speaker Recognition, Speech Coding, Speech Enhancement, Speech Synthesis, Voice analysis, and Speech Recognition. **Speech Signal processing** is refers to the acquisition, manipulation, storage, transfer and output vocal utterances by a computer. The main applications are the recognition synthesizes and Compression of human speeches.

2.2.1 Human Speech Production Mechanism

The figure 2.1 shows the process of human speech production mechanism. It is composed of four processes: language processing, in which the content of an utterance is converted into phonemic symbols in the brain’s language center; generation of motor commands to the vocal organs in the brain’s motor center; articulatory movement for the production of speech by the vocal organs based on these motor commands; and the emission of air sent from the lungs in the form of
speech [6]. Generation of voice source requires adequate configuration of the air flow from the lungs and vocal folds parameters for oscillation. The sources for voiced sounds are the air flow pulses generated at larynx, while those for some consonants are airflow noise made at a narrow constriction in the vocal tract. The expiratory and inspiratory muscles together generate relatively constant presser during speech. The laryngeal muscles adjust the onset and offset, amplitude, and frequency of vocal fold vibration.

2.2.2 Human Speech Perception Mechanism

Human hearing consists of several steps. Sound is channeled into the head through the outer ear and the ear canal. The sound is focused on the ear drum then it is transmitted and amplified by three small bones on to the side of the cochlea. The cochlea is a fluid-filled coil of varying diameter. It is filled with hairs (cilia) that sense the vibrations in the fluid as shown in figure 2.2.

Cochlea: Low-pitched sounds cause vibration at the wide end of the cochlea. High-pitched sounds cause vibrarion at the narrow end of the cochlea. Just as with the resonant bottle mentioned above, high-pitched sounds do not cause vibration at the wide end because they cancel each other out. The progression from wide to narrow allows the cochlea to discriminate between a large range of sounds, each of which will find a resonance somewhere.

Nerve Signals: The cilia within the cochlea are connected to nerves. As the cilia vibrate, the nerves fire signals into the brain. The brain combines the signals (sensory fusion) from various nerves to develop an overall sensation of the sound that is being received.

Frequency Range: The frequency range of sounds perceived by humans is between 30 Hz and 20,000 Hz. Above 20K we hear nothing. Human hearing is not equally sensitive at all frequencies. At very high frequencies, a sound must be louder to be perceived. The same is true at very low frequencies. The optimal frequency is about 1000 Hz. A sound at 1000 Hz is more easily heard than another sound with the same amount of energy at any other frequency [7].
2.2.3 Basic Components of Speech Research

Research in Speech technology is multidimensional and many factors need to be considered. The major factor that has most effect on the applicability of the current technology may be classified in four categories as follows:

1. The quality of the input speech,
2. Speaker dependant versus speaker independent recognition,
3. Vocabulary size and complexity of the task, and,
4. Continuous speech or isolated speech.

The quality of the input Speech

Two types of distortion can degrade the quality of the input signal. The first type of distortion is additive or environmental noise, i.e. many sounds from sources other than the target speaker. Examples are the noise of the engine and the tires for the speech recognition in the car and street noise for info-desk queries from a public phone. The second type of noise is convolutive noise. The recording channel causes this type of noise. Examples are room acoustics likes echoes and reverberations and
the recording device. Techniques exit to copy with both type of noise [8]. To avoid such type of noise, we have used Computerized Speech Laboratory (CSL).

**Speaker Dependence versus Speaker Independence Recognition**

The acoustic models – the models that relate the acoustic signal with different basic sounds in a language (phones) must be able to cope with different sources of variability in speech signals. Whereas speaker dependence systems have to copy with the variability caused by the environment and with the intra-speaker variability, speaker independent system also have to manage with the inter-speaker variability.

**Vocabulary Size and Complexity of the Task**

Constraints can be imposed on both the vocabulary and the grammar. The users have to adapt the vocabulary and grammar known by the recognizer. The system can be design as per the application. For example heavily constrained vocabularies and grammar are used for small command and control applications.

**Continuous Speech or Isolated Words**

In continuous speech there is no clear indication or even no indication at all of the boundaries between the words. There is also more variation in the speaking rate in continuous speech and words are pronounced less carefully. Isolated words are thus easier to recognize, but are less natural to humans [9].

**Continuous Speech and Spontaneous Speech**

Recognizers with continuous speech capabilities are very difficult to implement because they must utilize special methods to determine utterance boundaries. Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. Speech recognition system with spontaneous speech ability should be able to handle a variety of natural Speech Features such as words being run together, “ums” and “haaa”, and even slight stutters.

**2.2.4 Speaker Recognition**

Speech Signal not only contains the massage but also auxiliary information such as the gender or identity of speaker. In case of speaker recognition, the goal is to get
information about the speaker without much importance being given to the massage. Speaker identification involves the determining identity of speaker belong to close set. It is N-ways classification. Speaker verification is a decision processing of accepting or rejection the identity claim of person based on speech. Speaker Tracking is the process of determining whether two adjacent segments of speech belongs to same speaker or not.

A key, credit card, account number, personal identification number (PIN) or password is traditional authentication methods that have been used several years. However, they have the shortcoming that the key or credit card can be stolen or lost, and the PIN number or password can be easily misused or forgotten. For the signature, fingerprints, voice, facial features of authentication methods are known as biometric person authentication [10, 11]. Depending on the algorithm used for the identification, the task can be divided in to text-dependent and text-independent identification.

### 2.2.5 Speech Coding

Speech coding is an application of data compression of speech signals. It is the operation of the public switched telephone network (PSTN), videoconferencing systems, digital cellular communications, and emerging voice over Internet protocol (VoIP) applications. The goal of speech coding is to represent speech in digital form with as few bits as possible while maintaining the intelligibility and quality required for the particular application. Interest in speech coding is motivated by the evolution to digital communications and the requirement to minimize bit rate, and hence, conserve bandwidth. There is always a tradeoff between lowering the bit rate and maintaining the delivered voice quality and intelligibility; however, depending on the application, many other constraints also must be considered, such as complexity, delay, and performance with bit errors or packet losses [12]. The most two application of speech coding are mobile telephony and voice over IT.
2.2.6 Speech Enhancement

Speech signals are usually corrupted by adverse noise, such as competing speakers, background noise, or car noise, and also they are subject to distortion caused by communication channels; examples are room reverberation, low-quality microphones, etc. Other than specialized studios or laboratories when audio signal is recorded, noise is recorded as well. In some circumstances such as cars in traffic, noise levels could exceed speech signals. Speech enhancement improves the signal quality by suppression of noise and reduction of distortion. Speech enhancement has many applications; for example, mobile communications, robust speech recognition, low-quality audio devices, and hearing aid [13].

2.2.7 Speech Synthesis

Speech Synthesis is the artificial production of human speech Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database [14]. The trend in high quality speech synthesis is toward general unit selection. A unit selection synthesizer contains a database of large amounts of speech, labeled with respect to their phonetic, prosodic and even linguistic content. Efficient search algorithms then select units of predetermined or varying size to concatenate in order to build up an utterance. When working with sufficiently large databases one might expect that anything needed would be contained in the database.

2.2.8 Voice Analysis

Voice Analysis is study of speech sound for medical analysis of the voice for instant, analysis, of the voice of patients who have had polyp removed from vocal cord through an operation. The most important indirect methods are currently inverse filtering of either microphone or oral airflow recordings and electroglottography (EGG). Work related voice analysis has been done and describe in chapter 3.
2.2.9 Speech Recognition

Speech is process of converting of an acoustic signal, captured by a microphone or telephone, to a set word. Speech recognition systems can be separated in several different classes by describing what types of utterances they have ability to recognize. The progress of speech recognition technology in last six decade can be summarized as follows.

In 1952, at Bell Laboratories, Davis, Biddulph, and Balashek built a system for isolated digit recognition for a single speaker [15], using the formant frequencies measured during vowel regions of each digit. In an independent effort at Radio Corporation of America (RCA) Laboratories in 1956, Olson and Belar tried to recognize 10 distinct syllables of a single speaker, as embodied in 10 monosyllabic words [16]. In 1959, at University College in England, Fry and Denes tried to build a phoneme recognizer to recognize four vowels and nine consonants [17]. In the 1960s, since computers were still not fast enough, several specialpurpose hardwares were built. Suzuki and Nakata at the Radio Research Lab in Japan built a hardware vowel recognizer [18]. Sakai and Doshita at Kyoto University built a hardware phoneme recognizer in 1962, using a hardware speech segmenter and a zero-crossing analysis of different regions of the input utterance [19].

In the 1970s, speech recognition research achieved a number of significant areas. First, the area of isolated word or discrete utterance recognition became a viable and usable technology based on fundamental studies in Russia and Japan. Velichko and Zagoruyko in Russia advanced the use of pattern-recognition ideas in speech recognition [20]. Sakoe and Chiba advanced their techniques of using dynamic programming; and Itakura, when he was staying at Bell laboratories, showed how the ideas of linear predictive coding (LPC) could be extended to speech recognition systems through the use of an appropriate distance measure based on LPC spectral parameters [21].

Defense Advanced Research Project Agency (DARPA): A first demonstration of speech understanding was achieved by CMU in 1973. Their Hearsay I system was able to use semantic information to significantly reduce the number of alternatives
considered by the recognizer [22]. CMU’s Harpy system [23] was shown to be able to recognize speech using a vocabulary of 1,011 words with reasonable accuracy. One particular contribution from the Harpy system was the concept of graph search, where the speech recognition language is represented as a connected network derived from lexical representations of words, with syntactical production rules and word boundary rules. The Harpy system was the first to take advantage of a finite state network (FSN) to reduce computation and efficiently determine the closest matching string.

Statistical Modeling: Speech recognition research in the 1980s was characterized by a shift in methodology from the more intuitive template-based approach towards a more rigorous statistical modeling framework.

Hidden Markov Model (HMM): The key technology developed in the 1980s is the hidden Markov model (HMM) approach [24, 25, and 26]. It is a doubly stochastic process in that it has an underlying stochastic process that is not observable (hence the term hidden), but can be observed through another stochastic process that produces a sequence of observations.

Delta (Δ) Cepstrum: Furui proposed to use the combination of instantaneous cepstral coefficients and their first and second order polynomial coefficients, now called delta (Δ) and Delta Delta (ΔΔ) cepstral coefficients, as fundamental spectral features for speech recognition [27]. He proposed this method for speaker recognition in the late 1970s, but no one attempted to apply it to speech recognition for many years. This method is now widely used in almost all speech recognition systems.

N-gram: The n-gram model, which defined the probability of occurrence of an ordered sequence of n words, was introduced, and, since then, the use of n-gram language models, and its variants, has become indispensable in large vocabulary speech recognition systems [28]. A primary focus of IBM was the development of a structure of a language model (grammar), which was represented by statistical syntactical rules describing how likely, in a probabilistic sense, was a sequence of language symbols (e.g., phonemes or words) that could appear in the speech signal.

Artificial Neural Network: In the 1980s, the idea of applying neural networks to speech recognition was reintroduced. Neural networks were first introduced in the 1950s, but they did not prove useful because of practical problems. A deeper understanding of the strengths and limitations of the technology was achieved, as well
as an understanding of the relationship of this technology to classical pattern classification methods [29] [30] [31].

In the 1990s, Minimum Classification Error (MCE) criterion was proposed along with a corresponding Generalized Probabilistic Descent (GPD) training algorithm to minimize an objective function which acts to approximate the error rate closely [32]. Another example was the Maximum Mutual Information (MMI) criterion. In MMI training, the mutual information between the acoustic observation and its correct lexical symbol averaged over a training set is maximized. Although this criterion is not based on a direct minimization of the classification error rate and is quite different from the MCE based approach, it is well founded in information theory and possesses good theoretical properties. Both the MMI and MCE can lead to speech recognition performance superior to the maximum likelihood based approach.

**Robust speech recognition:** Various techniques were investigated to increase the robustness of speech recognition systems against the mismatch between training and testing conditions, caused by background noises, voice individuality, microphones, transmission channels, room reverberation, etc. Major techniques include the maximum likelihood linear regression (MLLR) [33], the model decomposition [34], parallel model composition (PMC) [35], and the structural maximum a posteriori (SMAP) method [36].

**Spontaneous speech recognition:** Although read speech and similar types of speech, e.g. news broadcasts reading a text, can be recognized with accuracy higher than 95% using state-of-the-art speech recognition technology, recognition accuracy drastically decreases for spontaneous speech. Broadening the application of speech recognition depends crucially on raising recognition performance for spontaneous speech. In order to increase recognition performance for spontaneous speech, several projects have been conducted. In Japan, a 5-year national project “Spontaneous Speech: Corpus and Processing Technology” was conducted. A world-largest spontaneous speech corpus, “Corpus of Spontaneous Japanese (CSJ)” consisting of approximately 7 millions of words, corresponding to 700 hours of speech, was built, and various new techniques were investigated. These new techniques include flexible acoustic modeling, sentence boundary detection, pronunciation modeling, acoustic as well as language model adaptation, and automatic speech summarization [37].
In order to have intelligent or human-like interactions in dialogue applications, it is important to attach to each recognized event a number that indicates how confidently the ASR system can accept the recognized events. The confidence measure serves as a reference guide for a dialogue system to provide an appropriate response to its users. To detect semantically significant parts and reject irrelevant portions in spontaneous utterances, a detection-based approach has recently been investigated [38]. This combined recognition and verification strategy works well especially for ill-formed utterances. In order to build acoustic models more sophisticated than conventional HMMs, the dynamic Bayesian network has recently been investigated [39].

**Multimodal speech recognition:** Humans use multimodal communication when they speak to each other. Studies in speech intelligibility have shown that having both visual and audio information increases the rate of successful transfer of information, especially when the message is complex or when communication takes place in a noisy environment. The use of the visual face information, particularly lip information, in speech recognition has been investigated, and results show that using both types of information gives better recognition performances than using only the audio or only the visual information, particularly in noisy environment.

**2.3 Brain Signal Processing**

A brain computer interface (BCI) or a brain machine interface is direct communication pathway between a human brain and external device. In one-way BCIs, computer either accepts commands from the brain or send signal to it but not both. BCI could be interface medium of the future. Instead of using peripheral input and output devices, the goal of BCI research is to enable direct interaction between humans and computers by directly receiving and transmitting signals to and from the brain.

**2.3.1 Invasive Vs. Noninvasive BCI Research**

Invasive systems interact with the brain directly, i.e., with electrodes that penetrate the brain or lay on the surface of the brain, while noninvasive systems interact with the
brain indirectly by transmissions through the skull, e.g., electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and magnetic sensor systems [40].

World Technology Evaluation Center (WTEC) has reported progress in BCI research in various parts of the world. This has included computer based animal and human testing of invasive and noninvasive systems, research and experimental protocols, experimental aimed at improving signal and pattern recognition, and hardware and software development. Whole invasive BCI work is done at institutes and University only in USA as shown in the Table 2.1. Whole non-invasive BCI work is done at various institutes and universities of European countries i.e. Austria Belgium, Denmark, England, France, Germany, Italy, Scotland and Switzerland as shown in the Table 2.2. There are few institutes or universities from Asian countries i.e. China and Japan and India as shown in the Table 2.3 [41]. There is no work related from Indian Institutes/Universities. Our work at Dr. Babasaheb Ambedkar Marathwada University is the first one to start research in this area in India.

Table 2.1: BCI work at American Institutes / Universities

<table>
<thead>
<tr>
<th>Country</th>
<th>Institutes / Universities</th>
<th>BCI related Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>Neil Squire Foundation</td>
<td>Asynchronous BCI and Brain Interface Research</td>
</tr>
<tr>
<td>USA</td>
<td>Washington University</td>
<td>Electrocorticographic (ECoG) Control of Brain-Computer Interfaces</td>
</tr>
<tr>
<td>USA</td>
<td>University of Michigan</td>
<td>Implantable Microscale Neural Interface Devices for BCI Systems</td>
</tr>
<tr>
<td>USA</td>
<td>University of Utah</td>
<td>Applications of Penetrating Microelectrodes in Nervous System Disorders</td>
</tr>
<tr>
<td>USA</td>
<td>Wadsworth Center</td>
<td>Understanding Biological Responses to Inserted Neural Prosthetic Devices: Building a Foundation to Promote Improved Tissue Integration and Device Performance</td>
</tr>
<tr>
<td>USA</td>
<td>Stanford University</td>
<td>Decoding Movement Plans for Use in Neural Prosthetic Devices</td>
</tr>
<tr>
<td>USA</td>
<td>Ad-Tech Medical Instrument Corporation</td>
<td>The Path from Research &amp; Development to FDA Approval to Commercialization</td>
</tr>
</tbody>
</table>
Table 2.2: BCI work at European Institutes/ Universities

<table>
<thead>
<tr>
<th>Country</th>
<th>Institutes</th>
<th>BCI related Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Graz University of Technology</td>
<td>The Graz group is working on several potential control features including the P300, slow cortical potential, steady state potential, event-related desynchronizations and synchronization. The group has an active research program mainly concerned with the development of noninvasive methods for EEG- Based communication and control. [42][43]</td>
</tr>
<tr>
<td>Austria</td>
<td>Guger Technologies OEG (g.tec)</td>
<td>Austria’s g.tec is one of the major developers of system for noninvasive and invasive brain computer interface technologies. The company has excellent products including amplifiers for recording of up to 64 channel of EEG/ECOG/ECG/EMG/EOG and related signals. The company has strongly software programs, which extensively with matrix laboratory (MATLAB) and Simulink drivers.</td>
</tr>
<tr>
<td>Belgium</td>
<td>European Union—Research Directorate General</td>
<td>This group is working in non-invasive brain interaction with robots- Mental Augmentation through determination of intended Action is researching a non-intensive brain computer interface.</td>
</tr>
<tr>
<td>Denmark</td>
<td>Aalborg University</td>
<td>The university includes research in stem cells, motor control and rehabilitation, sensory systems and technology, and medical and health informatics. Specifically, the neural prosthetic research is focused on FES, BCI, electrode development, biomechanics, and rehabilitation. Other areas of interest include human brain mapping, pain and biomechanics research, EEG analysis, human performance, motor control, health information systems, surgery simulation, image analysis, and virtual reality.</td>
</tr>
<tr>
<td>England</td>
<td>University of Oxford</td>
<td>Oxford groups primary goal is to develop highly accurate real time brain interface that can be used without training by any one and required only a very few small number of electrode. This group focuses on robust classification of discrete state from noisy EEG signals and analysis the data using a Bayesian statistical framework.</td>
</tr>
<tr>
<td>France</td>
<td>CEA (Atomic Energy)</td>
<td>This group is already substantial portion of an</td>
</tr>
<tr>
<td>Country</td>
<td>Organization</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>France</td>
<td>Physiology of Perception and Action Laboratory (CNRS/College de France)</td>
<td>The group is strong in the area of brain recognition of spatial orientation, neuron-robotics, knowledge engineering, virtual reality, and sensory interfaces.</td>
</tr>
<tr>
<td>Germany</td>
<td>Max Planck Institute for Biochemistry</td>
<td>The major goal of Max Planck institute is to study the basic physiological processes in the interface of brain tissue and semiconductor chips. The lab focus more on developing neuron-silicon interface than on BCI applications [44].</td>
</tr>
<tr>
<td>Germany</td>
<td>Natural and Medical Sciences Institute and Retina Implant (NMI)</td>
<td>The mission of NMI is to produce innovative research in medical science. The Retina Implant develops and implants sub-retinal CMOS array that converts incoming light into electrical patterns to stimulate surviving bipolar cells [45].</td>
</tr>
<tr>
<td>Germany</td>
<td>Berlin Brain-Computer Interface (BBCI)</td>
<td>Aims of BBCI is to reduce the time for patient training in BCI application and to enable faster neuroscience experimentations protocol by using signals from over learned movements. Statistical and machine learning methodologies include supervised and unsupervised learning approach such as nonlinear classification, regression, and prediction, with support vector machines (SVM) and kernel Fisher discriminant (KFD) analysis and support vector data description (SVDD), clustering[46].</td>
</tr>
<tr>
<td>Germany</td>
<td>Multi Channel Systems (MCS)</td>
<td>MCS develops automated multi-channel measurement devices that provide fast secondary functional screening for basic research and pharmaceutical applications. The company is developing data acquisition software provides additional filters and simple analytical tools with more advanced function available as free MATLAB toolbox [47].</td>
</tr>
<tr>
<td>Germany</td>
<td>University of Freiburg</td>
<td>The university is conducting the research in human epicortical field potentials, for this purpose array thin, flat disks electrodes surgically implanted on the brain surface for the purpose of identifying seizer generating</td>
</tr>
<tr>
<td>Country</td>
<td>Institutes</td>
<td>BCI Work</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Germany</td>
<td>University of Tübingen</td>
<td>In 2000 the group developed BCI2000, since this time it is used sensorimotor and P300 based BCI systems in addition to the slow cortical potential based BCI [49].</td>
</tr>
<tr>
<td>Italy</td>
<td>Polo Sant'Anna Valdera</td>
<td>The Lab is developing artificial limb especially the upper extremities such as human hand, which reached a very high level of achievement. Using the bio-mechatronic approach to duplicate the natural hand, the group has produced Cyber hand an elegant robotic surrogate that uses an under actuate design with multiple degrees of freedom and incorporates tactile bio-mimetic sensor feedback.</td>
</tr>
<tr>
<td>Italy</td>
<td>The Santa Lucia Foundation</td>
<td>The major aim is to develop computer-controlled environments for disable patients who can utilize EEG recording for assisted control. The group focuses on increasing the special resolution of EEG recordings by using MRI images of the head compartment and brain. This approach is being used to reveal the locations of cortical activity relating to executing of task [50].</td>
</tr>
<tr>
<td>Scotland</td>
<td>University of Edinburgh</td>
<td>The goal of University is successful silicon-neuron interaction for cellular recordings and also design of neuromorphic learning system.</td>
</tr>
<tr>
<td>France</td>
<td>Swiss Federal Institute of Technology</td>
<td>The institute has developed an asynchronous protocol EEG analysis and given it to machine learning techniques and artificial-intelligence robotics. In this, the institute has used principles of ‘adaptive sheared autonomy’ to enhance the BCI’s functionality [51].</td>
</tr>
<tr>
<td>Country</td>
<td>Institution Name</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>China</td>
<td>Shanghai Institute of Brain Functional Genomics</td>
<td>The major goal of this group is record large scale neural ensemble recording in the brains of freely rat and mice. The group has identified hippocampal neurons associated with the animal identification of a nest and that shares many of the properties of place cells of the hippocampus [52].</td>
</tr>
<tr>
<td>China</td>
<td>Tsinghua University, Department of Electrical Engineering</td>
<td>The group’s interests are combining the high temporal resolution of EEG methods with the high special resolution methods of fMRI and other imaging methods to achieve bioengineered systems for BCI, diagnosis of neurological diseases, and neurobiological understanding of cognition and perception.</td>
</tr>
<tr>
<td>China</td>
<td>Tsinghua University Institute of Microelectronics</td>
<td>The CMOS chip design study involves development microelectrode array technology to invasive electrode approaches. The electrode chip has onboard electronics that have been developed through a CMOS process and currently contains a 16-channel.</td>
</tr>
<tr>
<td>China</td>
<td>Shanghai Jiao Tong University</td>
<td>The group is developing novel approaches to real time assessment of wakefulness, using measurement of facial expressions and/or multi-site EEG. The group is busy in mathematical modeling and algorism development of EEG based BCIs. The group is also investigating application of independent analysis components (ICA) with multi-channel nonlinear adaptive filtering to decompose the EEG into separate sources.</td>
</tr>
<tr>
<td>China</td>
<td>Wuhan University</td>
<td>The group is pursuing the development of silicon-based micro electrode array for stimulating targeted neural populations in the central nerves system and recording of single-unit activity in using iridium</td>
</tr>
<tr>
<td>Country</td>
<td>Institution</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>China</td>
<td>Shanghai Institute of Brain Functional Genomics</td>
<td>The major goal of this group is to record large scale neural ensemble recording in the brains of freely rat and mice. The group has identified hippocampal neurons associated with the animal identification of a nest and that shares many of the properties of place cells of the hippocampus [52].</td>
</tr>
<tr>
<td>China</td>
<td>Tsinghua University, Department of Electrical Engineering</td>
<td>The group’s interests are combining the high temporal resolution of EEG methods with the high special resolution methods of fMRI and other imaging methods to achieved bioengineered systems for BCI, diagnosis of neurological diseases, and neurobiological understanding of cognition and perception.</td>
</tr>
<tr>
<td>China</td>
<td>Tsinghua University Institute of Microelectronics</td>
<td>The CMOS chip design study involves development microelectrode array technology to invasive electrode approaches. The electrode chip has onboard electronics that have been developed through a CMOS process and currently contains a 16-channel.</td>
</tr>
<tr>
<td>China</td>
<td>Shanghai Jiao Tong University</td>
<td>The group is developing novel approaches to real time assessment of wakefulness, using measurement of facial expressions and/or multi-site EEG. The group is busy in mathematical modeling and algorism development of EEG based BCIs. The group is also investigating application of independent analysis components (ICA) with multi-channel nonlinear adaptive filtering to decompose the EEG into separate sources.</td>
</tr>
<tr>
<td>China</td>
<td>Wuhan University</td>
<td>The group is pursuing the development of silicon-based micro electrode array for stimulating targeted neural populations in the central nerves system and recording of single-unit activity in using iridium</td>
</tr>
<tr>
<td>Country</td>
<td>Organization</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Japan</td>
<td>RIKEN Brain Science Institute</td>
<td>This group has the greatest collection of EEG hardware. The group is recording EEG signals from as many as 256 passive or active gel type electrodes for the potential early diagnosis of Alzheimer’s disease [54].</td>
</tr>
<tr>
<td>Japan</td>
<td>Advanced Telecommunications Research Institute (ATR)</td>
<td>The major effort are of the ATR Central Nervous System (CNS) relevant to BCI’s noninvasive neural decoding project. The aim of this research effort is to use a combination of noninvasive recording and imaging method. Such as EEG, EMG, fMRI, and NIRS, to decode and classify brain representation of external events emotional states and movement plans.</td>
</tr>
<tr>
<td>Japan</td>
<td>NTT Communication Science Laboratories</td>
<td>A major focus involving NTT BCIs is Parasitic Humanoid (PH) robot for modeling nonverbal human behavior. Parasitic Humanoid project is involved the development of a Variety of sensors and actuators that are useful for providing feedback for BCI application [55].</td>
</tr>
<tr>
<td>Japan</td>
<td>Waseda University</td>
<td>The work is focused on fundamental studies of detecting and learning patterns in time series data (data from sensors). The group is developing mathematical methods that will allow models of nonlinear, non-stationary processes that provide predictive power that can classify patterns in the data [56] [57].</td>
</tr>
<tr>
<td>India</td>
<td>Dr. Babasaheb Ambedkar Marathwada University, Aurangabad</td>
<td>Analysis of EEG signals induced by different mental thought.</td>
</tr>
</tbody>
</table>
2.3.2 Sensor Technology

The majority of BCI science in America involves “invasive” sensor technologies through these multi-electrode recordings from arrays of microelectrodes implanted directly into the brain. This technology is becoming popular in America. European countries involve “noninvasive” sensor technologies, i.e., using multi-electrode recordings from arrays of EEG electrodes mounted onto the surface of the skull. This sensor technology has experienced a very limited growth and requires substantial improvement. Certain BCI sites in Europe are capable of providing sensor technologies that could aid in the advancement of “invasive” sensor technologies; however, this is not their current plan. In Asia, there is a clear emphasis on less expensive EEG BCI approaches. Reasons include the large population in China and the need for low-cost, noninvasive BCI technology for improved public healthcare there. Japan is also focused on noninvasive EEG-based BCI technologies.

2.3.2.1 Invasive types of Sensors

In 1980 Chapin and Woodward (1986) reported the development of 50 μm tungsten microwire arrays for multiple single-unit recordings. Basically, this type of technology is used today by many laboratories for the more routine multiple single-unit recordings and many applications of BCI in animals.

2.3.2.1.1 Wire Type Microelectrode

Most wire-type microelectrodes are constructed by sealing a metal (tungsten, gold, platinum, iridium, platinum-iridium, stainless steel) wire in an insulating material. The metal wires from the brain and the connections between the recording wires are insulated using Teflon or plastics. The microelectrode surface area is determined by cutting the exposed wire to a desired length. Typical wire electrodes range in diameter from 13–200 μm, with an exposed length of up to 1 mm [58] as shown in figure 2.3. Figure 2.3 (A) is base foundation for the micro drive, (B) indicates four 36-pin
connector arrays positioned at the base of the microdrive in parallel, (C) is a microdrive on the assembly stage, (D) is a fully assembled adjustable 128- electrode microdrive, (E) indicates that 128-channels can be formatted with either tetrodes or stereotetrodes on each bundle.

Figure 2.3: Wire Type Micro Electrode

2.3.2.1.2 Silicon Based Microelectrodes

The semiconductor properties of silicon can be altered by doping. Also, silicon is very compatible with on board circuitry. Silicon has many features that have made it widely used as the foundation for forming microelectrode arrays. Silicon-based microelectrodes constructed at the Center for Neural Communication Technology at the University of Michigan, is shown in Figure 2.4 [59] [60].
2.3.2.1.3 Ceramic-Based Microelectrodes

The insulator ceramic has been used as a substrate to reduce crosstalk between adjacent connecting lines. Ceramic is mechanically strong, allowing for development of microelectrodes that can access much deeper brain structures. Precise placement of the microelectrode in tissue without flexing or breaking can be achieved. Multi-site microelectrodes on ceramic substrates for use in animal models have been constructed [61]. The part ‘A’ from figure shows complex ceramic substrate-based microelectrode shape cut by laser machining. The part ‘B’ is less complex microelectrode shape formed by a computer-controlled diamond saw and ‘C’ is a magnification of the microelectrode’s much smoother edge as shown in figure 2.5.

Figure 2.4: Silicon-Based Microelectrode arrays

Figure 2.5: Ceramic Substrate-Based Microelectrode
2.3.2.1.4 ECoG Strip Electrodes

These electrodes are used primarily by comprehensive epilepsy centers and major medical centers that provide brain mapping in their neurological programs. These electrodes are made of implant silicone or polyurethane with micro-conductors attached to stainless steel or platinum contacts (usually 7 or 10 mm disks) that populate the dielectric area. Figure 2.6 shows numerous Ad-Tech ECoG strip electrodes ranging in size from 4 to 64 recording sites. Proprietary connectors/cables attach these electrodes to commercial monitoring equipment [62].

![Figure 2.6: ECoG Electrodes](image)

2.3.2.2 Noninvasive EEG Sensors for BCI

All BCI studies using noninvasive sensors involve the use of Ag or Au disk electrodes with conducting paste that are affixed to the skull using some type of head cap configuration to facilitate the application of the EEG electrodes. In particular, the electrode cap design requires extra time for attachment of electrodes but achieves excellent signal-to-noise characteristics. This highly versatile design can be employed with other products and amplifiers, as well as other suppliers of such instrumentation [63]. The process of fitting individuals with EEG electrodes with head caps, however, is time consuming, requires testing of individual electrodes for their impedance, and results in a system that is not comfortable or practical for routine BCI use as shown in figure 2.7.
2.3.3 BCI Modeling and Signal Processing

Time-domain features, frequency domain and the spike features have been accurately extracted; the neuronal spike firings become the decoding model input. It is possible to translate the decoding problem into a system identification framework, where a parametric linear or nonlinear system is trained directly from the collected data to achieve outputs close to the hand positions as shown in Figure 2.8. Model building has been extensively studied in control theory and signal processing, so there is a wealth of methods that can be utilized [64].

Figure 2.8: BCI System Identification Frameworks
2.3.4 Hardware Implementation

Paralysis refers to complete loss of motor function, whereas “paresis” refers to relatively minor loss. Severe paralysis is a major problem, not just because patients lose their ability to have a normal life, but also because of the tremendous cost of patient maintenance. Patients with tetraplegic spinal cord injuries, for example, lose virtually all voluntary motor function below the neck, but also lose somatosensation, i.e., their senses of touch, pain, temperature, and limb position. Severe amyotrophic lateral sclerosis (ALS) is even worse, in that patients can lose all motor function throughout their bodies. For these types of patients, the direct nerves machine interfacing can be called Biomimetic Hand Prostheses. Thus the Cyberhand [65] is inspired by understanding of the kinesiology of hand and finger movements, in figure 2.9. Since it will be impossible to build a robot that includes all of the features that exist in the real human, the challenge is to design modifications that combine similar functionality with less complexity.

Figure 2.9: Cyberhand system
2.3.5 Noninvasive Communication System

A BCI system consists of sensors that record neural activity, signal processing that extracts features, and a translation algorithm that creates device commands to operate an external device [66]. The loop is completed with feedback from the external device to the BCI system user. The P300 Evoked Potential This potential occurs with latency around 300 msec in response to target stimuli that occur infrequently and that subjects are instructed to respond to in some manner, shown in figure 2.10. Donchin and colleagues [67] [68] first reported the use of the P300 for BCI communication. Their paradigm involved a 6×6 matrix of grey symbols on a dark background. Rows and columns of the matrix were randomly intensified. A P300 was produced when the attended row or column flashed.

![P300 Spike Graph]

Figure 2.10: Event Related Potential (ERP) P300 Spike

The attended symbol was selected by averaging responses for rows and columns. Accurate performance was obtained in users with and without disabilities. However, users attended only to the letter “P” in these studies; although demonstrating proof of principle, these initial studies did not actually involve communication.
2.3.6 Cognitive and Emotional Neuroprostheses

Emotions are high-level cognitive states that encode subjective feelings to a situation or environment. As such, emotions can carry large amounts of information in a compact form. For example, anger or frustration at continued BCI errors could lead to user rejection of a system. The emotion-oriented computing is based on psychobiological investigations of emotion and is designed to interface with human users on an emotional level. Robocasa [69], collaboration between the Scuola Superiore Sant’Anna in Italy and Waseda University in Japan, is creating emotional humanoid robots with expressive gestures, capable of expressing several human emotions with face and arms Figure 2.11[70]. Combining the ability to decode emotions from EEG or other brain recordings with such emotion-oriented computing or emotional robots would enable BCI users to express their emotions or enable the BCI to respond appropriately to the user’s emotional state.

Figure 2.11: Seven Emotions expressed by Humanoid Robot
2.4 Conclusion

Tremendous amount of activities have been going on in HCI all over world. In spite of this trend, India is not playing active role, whereas there is need to do lot of research for Indian languages as well as under Indian conditions. Our group at BAMU has initiated work in this direction [71]. This thesis reports the work done in speech recognition in Marathi and also BCI using noninvasive EEG technique.
References


