Paper from Proceeding of the National Conference "Science in Media 2012" Organized by YMCA University of Science and Technology, Faridabad, Haryana (India) December 3rd -4th 2012 Published by
IJMRS's International Journal of Engineering Sciences, ISSN (Online): 2277-9698
www.ijmrs.com

Development of a humanoid robot using gesture and voice technique along with heat detection for better human-machine interaction

Amit Kumar, YMCA University of Science & Technology, Faridabad, Haryana, India

Abstract **-In recent years, gesture control and voice recognition technique have become new developmental trends for many human-based electronic products. These techniques let people control these products more naturally, intuitively and conveniently. In this paper, a fast gesture recognition scheme with voice recognition system and body heat detection (IR detectors using IR cameras), is proposed. This technique serves as basis for an interface between humans and machines, thus forming a human-machine interface for a Human-Machine-Interaction (HMI) system. This paper is designed keeping in mind small-sized intelligent humanoid robot as a foundation. It introduces the realization of humanoid-robot basic functions and performing human-computer interaction on above mentioned intelligence functions. Firstly, the machine uses 'spoken dialog system' (SDS) to recognize the person and estimate his mental state. Secondly, with the help of gesture recognition technique (also known as EMG-electromyography) and infrared heat detectors, machine confirms the state of condition (like happiness, sadness, anger, extreme anger, etc) of humans, thus maintaining a proper interaction between humans and machines. This makes them to perform the required task without any instruction given by humans.**

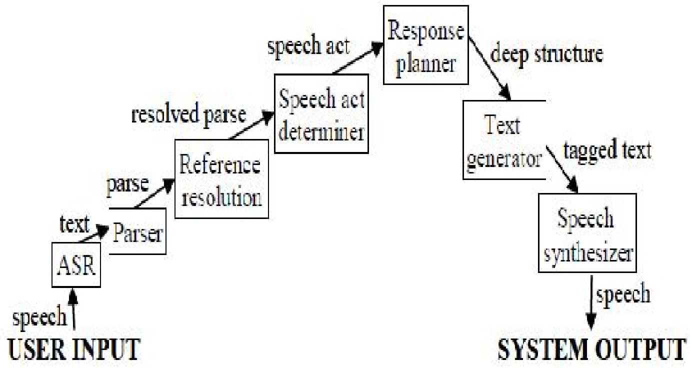
I. INTRODUCTION

For over centuries, humans have, due to its vast demands, interacted with many things, both living and non living. Humans interact with other humans and animals, share their emotions successfully in natural ways like face or voice recognition. Now, in the technical era, humans are stepping into a new world but their interactions are still led by these natural ways. Beside the use of electronic devices like keyboard and mouse, humans will, in near future, use voice recognition and face gesture recognition techniques on machines leading to a better Human-Machine Interaction (HMI).

This paper focuses on the design of a small-sized intelligent humanoid robot equipped with sensors, Spoken Dialog System (SDS), learning algorithm, face electromyography (EMG) signal detectors and body heat detectors, leading to the development of a robot capable of recognizing emotional conditions. Humanoid robots occupy less volume, are light-weight and have flexible motion along with a human appearance, action and humanization of interaction function. This helps humans to be more familiar with machine and better interface with them. This paper is divided into various sections, with each section discussing basic functions like voice recognition, face gesture recognition and body heat detection. Thereafter, a proper combination of all these techniques which states various emotional (like happy, sad, anger, threatened, tired, etc) conditions are discussed for HMI.

II. VOICE BASED HMI SPEECH RECOGNISATION SYSTEM

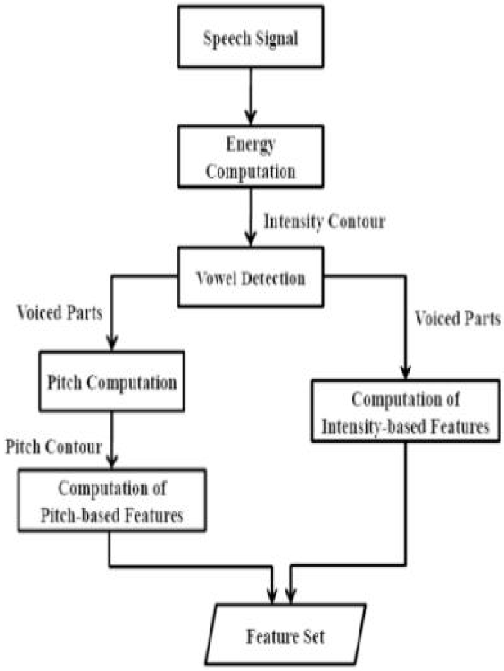
Speech recognition has proved to be a better mode of interaction between human and machine. In machines, a spoken dialog system is embedded which, with the combination of various processors, recognizes human voice. Usually the embedded speech recognition system suitable for home appliances applications consists of four parts: analog/digital and digital/analog conversion, speech recognition, voice prompts and voice playback section. This design uses a 16-bit microcontroller Sun plus SPCE061A as core chip, an A/D converter, encoding algorithm, decoding algorithm, storage and D/A converter. These are made into appropriate modules, each module having its own application interface API. Users only need to know the parameters to achieve the function, and then call the API function to realize the voice processing functions. Before the speech recognition, it should be trained. At first a standard voice signal is inputted to the memory of the chip and stored in the chip SPCE061A internal Flash. After that speech recognition can be programmed. When the person who has trained the robot input a voice signal, SPCE061A SCM collects the signal and generates an interrupt signal to MCU which is transferred to execute the interrupt subroutine to convert the voice, then store it in temporary memory, at the same time pre-deposit standard voice signal is picked up from the Flash memory for comparison. If these two sets of data are roughly the same i.e. within the permissible error range, it indicates that the input voice signal is correct, and the microcontroller can execute corresponding task; if it is quite different, then do not perform the appropriate tasks and actions.



A system architecture for a simple spoken dialogue system.

III. EMOTION IN SPEECH

Human emotions are mainly detected by the different intensity and pitch signals present in the human vocal words. From the intensity contour of the signal, vowel detection was implemented. The vowels are the voiced parts of the speech signal which are indicated by the peaks in the intensity contour. The peaks from the intensity contour are then detected. The number of peaks approximates the number of vowels present in the signal. The intensity-based features were computed from the intensity contour. The parts of the speech which contained the vowels were also passed to the pitch detection block, which produces the pitch contour of the voiced parts. Then the pitch-based features were derived from the pitch contour. The pitch of the speech frame is computed using the short-time average magnitude difference function (AMDF) (Yu-min, 2003). The average magnitude difference function of a frame of signal is computed.



Flowchart of the Feature Extraction Algorithm

The preliminary feature set used for this project are as follows:

Intensity-based features:

(1) Speaking Rate
(2) Mean Value of Individual Voiced Parts
(3) Mean Minimum of Individual Voiced Parts
(4) Mean Maximum of Individual Voiced Parts.

Pitch-related features:

(1) Minimum Pitch
(2) Maximum Pitch
(3) Median Pitch
(4) Standard Deviation of Pitch
(5) Range of Pitch
(6) Minimum Derivative of Pitch
(7) Maximum Derivative of Pitch
(8) Median Derivative of Pitch
(9) Standard Deviation of Derivative of Pitch
(10) Upslope Ratio
(11) Mean Positive Derivative of Individual Slopes

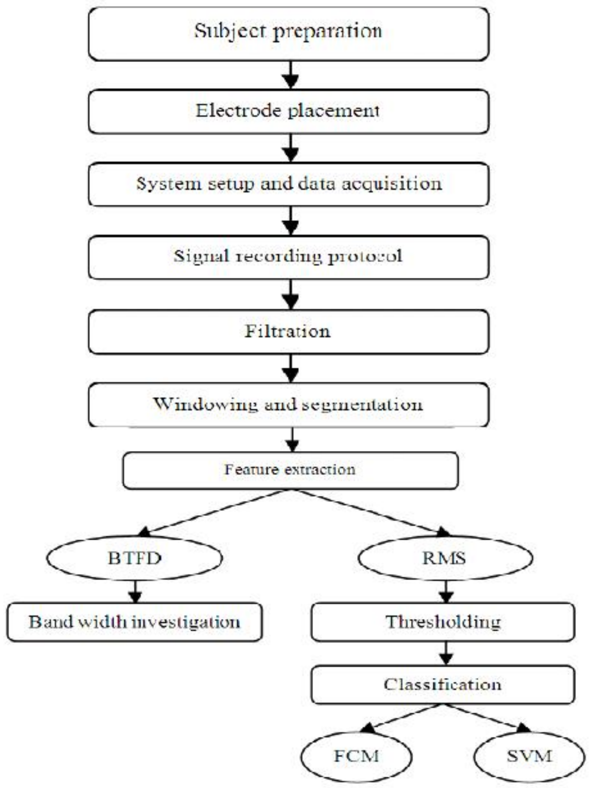
(12) Mean Negative Derivative of Individual Slopes.

With the help of micro-controller all these features are computed by the machine and the emotions like happiness, anger, joyful, etc are detected.

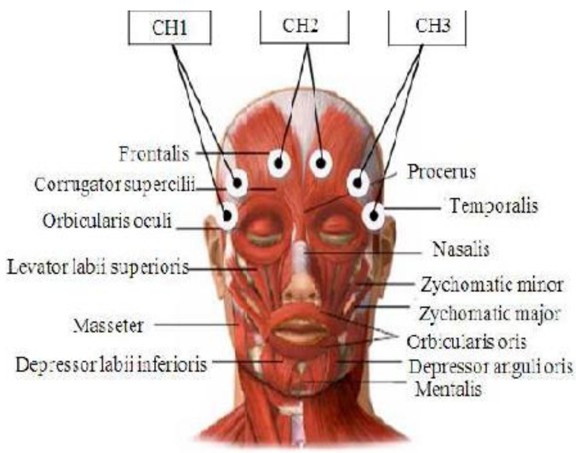
IV. FACE GESTURE RECOGNISATION

Facial expressions happen from motions or positions of facial muscles when they are stimulated by an internal or external hypo. It can be a visible manifestation of the state demonstrating cognitive activity, personality and psychopathology of a person (Donato et al., 1999). Face is the most visible and expressive of all the channels of communication such as emotions and facial gestures and it acts a communicative role in interpersonal relations. It is also one of the best sources of information in human's body since EMG, EOG, EEG signals can be extracted. Apart from them, facial expressions can play an important role wherever humans interact with machines. Automatic recognition of facial expressions and gestures may act as a major part of natural human machine interfaces.

General block diagram: The general flow chart of the proposed method has been depicted in Fig. Subject preparation and electrode placement can be seen as two first step of this project. After that, system setup and data acquisition are considered. Signal recording protocol is mentioned in next step. Then, the EMG signals which are recorded by the surface electrodes are amplified and filtered prior to process. Afterward, noise and artifacts are eliminated from the raw signal by using notch filter 48Hz-52Hz. Subsequently for processing, the acquired signal must be windowed to distinguish the active part of signal. Then, The RMS features are extracted from each window (256 msec) to feed the classifiers and BTFD features elicited for frequency band-width investigation. Prior to classification all RMSs features are passed from a threshold line to collect the active features. Finally, active RMSs are classified and clustered into eight separate facial gesture groups through FCM and SVM classifiers.



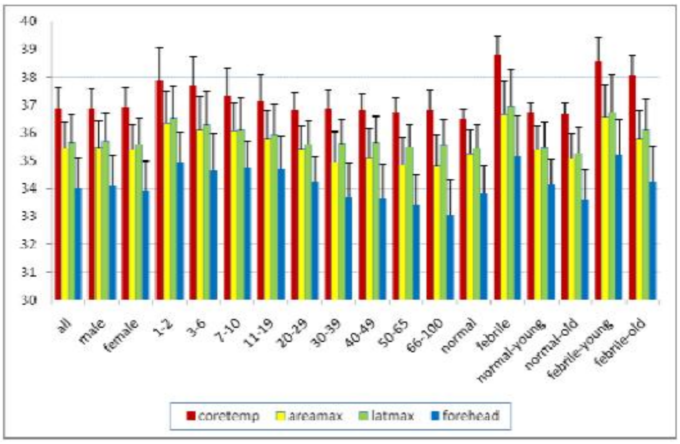
Facial gestures are tracked by markers allocated on the human face. With the movement of these points the different actions are detected by the sensors and the machine is able to recognise the emotions of human as shown in fig. and the table



Gesture no.	Gesture name	Effective channel (s)
1	Natural	1,2,3
2	Smiling (S)	2,3
3	Smiling with right side (pulling up the right corner of lip) (RS)	2
4	Smiling with left side (pulling up the left corner of lip) (LS)	3
5	Anger (Frowning with brow lowerer) (A)	1
6	Rage (clenching molar teeth) (R)	2,3
7	Gesturing 'No' (pull up the eyebrows) (N)	1
8	Opening the mouth like saying 'a' in 'apple' (O)	2,3

V. INFRARED BODY HEAT DETECTOR

Infrared radiation exists in the electromagnetic spectrum at a wavelength that is longer than visible light. It cannot be seen but it can be detected. Objects that generate heat also generate infrared radiation and those objects include animals and the human body whose radiation is strongest at a wavelength of 9.4um. Infrared in this range will not pass through many types of material that pass visible light such as ordinary window glass and plastic. However it will pass through, with some attenuation, material that is opaque to visible light such as germanium and silicon. An unprocessed silicon wafer makes a good IR window in a weatherproof enclosure for outdoor use. It also provides additional filtering for light in the visible range. 9.4um infrared will also pass through polyethylene which is usually used to make Fresnel lenses to focus the infrared onto sensor elements.

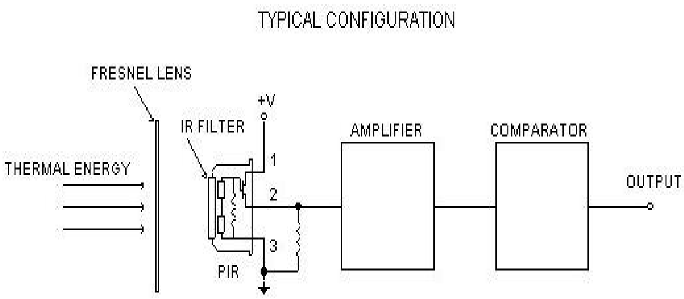


Mean and standard deviation of core and IRT temperatures in different subgroups

VI. PYROELECTRIC SENSORS

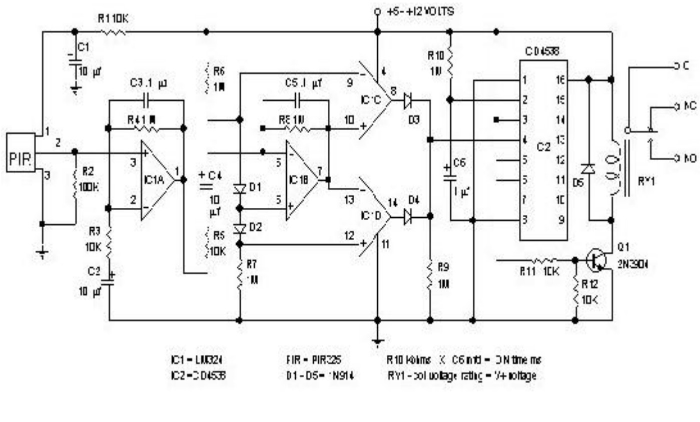
The pyroelectric sensor is made of a crystalline material that generates a surface electric charge when exposed to heat in the form of infrared radiation. When the amount of radiation striking the crystal changes, the amount of charge also changes and can then be measured with a sensitive FET device built into the sensor. The sensor elements are sensitive to radiation over a wide range so a filter window is added to the TO5 package to limit detectable radiation to the 8 to 14mm range which is most sensitive to human body radiation.

Typically, the FET source terminal pin 2 connects through a pull down resistor of about 100 K to ground and feeds into a two stage amplifier having signal conditioning circuits. The amplifier is typically bandwidth limited to below 10Hz to reject high frequency noise and is followed by a window comparator that responds to both the positive and negative transitions of the sensor output signal. A well filtered power source of from 3 to 15 volts should be connected to the FET drain terminal pin 1.



The PIR325 sensor has two sensing elements connected in a voltage bucking configuration. This arrangement cancels signals caused by vibration, temperature changes and sunlight. A body passing in front of the sensor will activate first one and then the other element whereas other sources will affect both elements simultaneously and be cancelled. The radiation source must pass across the sensor in a horizontal direction when sensor pins 1 and 2 are on a horizontal plane so that the elements are sequentially exposed to the IR source. A focusing device is usually used in front of the sensor.

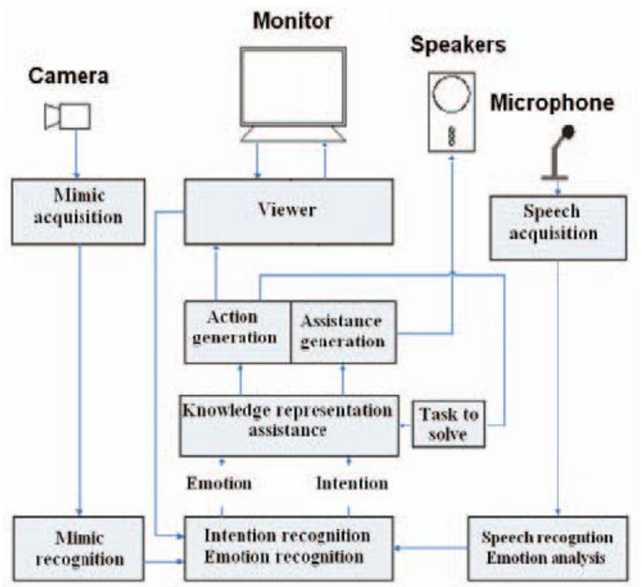
This is a typical application circuit that drives a relay. R10 and C6 adjust the amount of time that RY1 remains energized after motion is detected.



MOTION DETECTOR

VII. HUMAN MACHINE INTERFACE

Multimodal Human Machine Interfaces (MHMI) have recently become a new feature for different applications. We describe one of possible MHMI architecture for a Spoken Dialog System.



Systematic overview of the technical modules involved in the demonstrator.

Above figure shows the structure of the NIMITEK multimodal human computer interaction system. Humans employ several emotions and with these recognisation systems the emotional states can be easily classified by the machine.

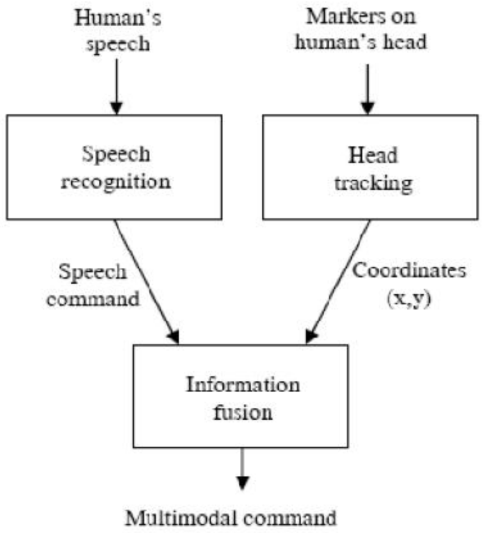


Diagram of a bimodal system

On the basis of all these sensors and techniques the judgment made by the machine is described in the table.

Face gesture	**Voice**	**Temperature**	**Human emotions**
Dull/ unstretched	Low voice	Moderate	Sad
Exterme stretched	High voice	Very high	Anger
Stretched	High (little)	High	Happy
Extreme stretched	Feeble	Very high	Joyfull
Loose/ unstretched	Low	Low	Tired
Normal	Moderate	Moderate	Relaxed
Stretched	Low	High	Terrified
Stretched	Low	Normal	Shy
Stretched	High	High	Jealous

Dull/ loose	Low	High	Depression

Some other conditions like the human state like being in the morning the temperature of human is high with stretched face muscle and high voice pitch recognize the state of high energy and relaxed condition.

Thus all these techniques help the humanoid robot to recognize the emotional condition of human and thus help in the better Human-Machine Interface (HMI).

CONCLUSION

The better human machine interface system can be established by using more than one real time sensor system to sense the real time environment and to process the data by taking all the domains in consideration which results in effective result which is more in proximity with the desired or actual conditions so that the machine can work in better and efficient way. These techniques let people control these products more naturally, intuitively and conveniently.

Here I consider, a fast gesture recognition scheme with voice recognition system and body heat detection (IR detectors using IR cameras) which serves as basis for an interface between humans and machines, thus forming a human-machine interface for a Human-Machine-Interaction (HMI) system.

REFERENCES

[1] Bogdan vlasenko and Andreas wendemuth, "heading toward to the natural way of human-machine interaction:The nimitek project".

[2] Lishtiaq rasool khan, Hiroyuki miyamoto, and Takashi morie, "face and arm-posture recognition for secure human-machine interaction,"

[3] Aching-hao lai*, "a fast gesture recognition scheme for real-time human-machine interaction systems".

[4] Vonghui xing and Wenzhuo chen, "design of speech recognition robot based on mcu,"ieee trans. Pattern anal. Mach. Intell., vol. 26, no. 6, pp. 689–698,jun. 2004.

[5] Qiang zhang, Jiandong zhao, Dapeng wei , Zhaoyang wu,Zhaoxuan li, "small-sized intelligent humanoid robot based on a multiple sensor fusion," 2012 4th international conference on intelligent human machine systems and cybernetics.

[6] Xin guo, Xiuyun wang, xiaoyun xu,, "design of human-machine interaction module and the communication of control system" proceedings of the 2009 ieee international conference on information and automation june 22 -25, 2009, zhuhai/macau, china.

[7] Ong sing goh, and Chun che fung, 2Kok wai wong "visualchat: a visualization tool for human-machine interaction" 2008 ieee/wic/acm international conference on web intelligence and intelligent agent technology.

[8] Roger k. Moore,, "a human-inspired architecture for speech-based human-machine interaction," in proc. Ieee transactions on computers, vol. 56, no. 9, september 2007

[9] Takanori shibata , Makoto yoshida and Junji yamato , "artificial emotional creature for human-machine interaction".