V-TOY

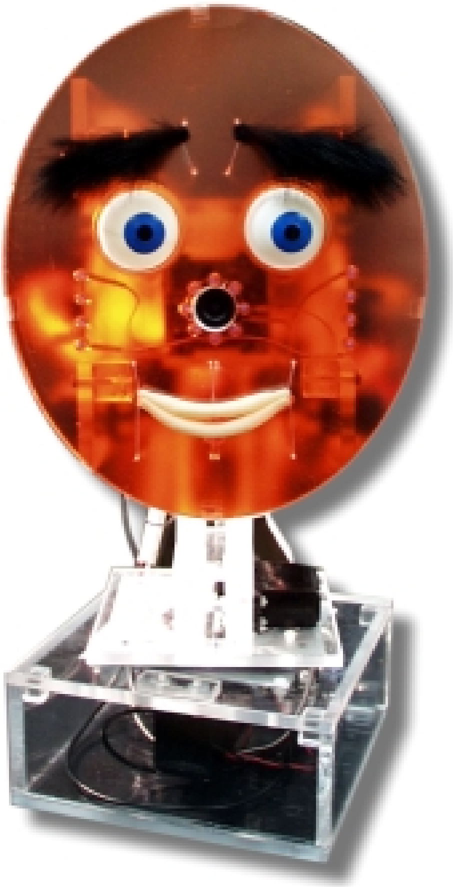
Visually Interactive Toy

Background

Toys are currently deaf and blind. They are unable to recognize the presence of humans, understand their communicative messages, or react to these messages. V-TOY is a prototype toy of the future. It was created with the premise that physically and sensory enabled interactive toys offer a more compelling human experience. Robotics, vision, acoustics, natural language processing and computational behaviors are employed to create a pioneering interactive experience.

V-TOY is an audio-visually interactive robot that reacts to human presence and "understands" simple communication messages. It replicates the functions of the upper part of the human body with a rotating neck, two controllable eyes, two cameras, deformable eyebrows and a mouth. V-TOY detects a person walking into its space and greets and engages the person visually and verbally.

V-TOY is an ongoing project between University of Maryland at College Park and IBM-Almaden for enabling robots to better perceive and understand human actions and react to them by executing audio-visual-motor behaviors.



Impact

Scientific contributions:

- Building a simple robot that is functionally similar to a human head, yet cartoon-looking. The prototype is a test-bed for the capabilities needed by audio-visual toys to participate in a social environment.
- Developing algorithms for understanding human movement and actions. These algorithms focus on the interaction between people and V-TOY.
- Testing interactive strategies between people and audio-visual toys. Determining which interactions people are comfortable with and defining future behavioral capabilities that V-TOY needs to acquire.

Social contributions:

- The development of intelligent toys (i.e., toys having perceptual, auditory and speech processing capabilities) transforms toys from passive to interactive. Interactive toys have the potential to be more beneficial to a child's development than passive toys.
- Computers, which are rapidly becoming embedded in our environment, can be made more human-friendly with audio-visual interactive capabilites. This involves perceptual understanding of human body and facial movements, as well as understanding how humans react to robotic emulation of humans.

UNIVERSITY OF MARYLAND

IBM®

Physical Capabilities

Eyes are constructed from ping-pong balls each having two degrees, azimuth and elevation, of rotational freedom. This enables V-TOY to establish eye contact with people in the camera's field of view and track them as they move.

Eyebrows have one degree of freedom that corresponds to the corrugators supercilii and medial frontalis muscles, alllowing V-TOY to raise and lower its eyebrows.

Mouth is controlled by four servos. The first two enable the left and right mouth corners to move up and down. The second two enable the upper and lower middle lip to move up and down enabling the mouth to open and close.

Neck has two servos controlling the pan and tilt positions of V-TOY's head.

Interfacing through an RS232 port driven from a Java application.

Sensors include two video cameras, a microphone array, and speakers.

Visual Capabilities

Detection of human presence. V-TOY initiates interaction once it detects the presence of a human in its space. Two video cameras analyze the environment searching for faces. When multiple people are detected, it chooses to engage the closest person.

Face expression analysis. V-TOY employs motion analysis to observe human facial expressions. It tracks the motion of the person's face and measures the non-rigid motion of face regions such as the eyebrows and mouth.

Gaze Analysis. V-TOY establishes and maintains eye-contact with the person by detecting the person's pupils. It uses two infrared time-multiplexed light sources, synchronized with the camera frame rate.

Color analysis. V-TOY employs color analysis to enhance its perception of people. Skin color and background color analysis guides V-TOY in its search for faces. Color analysis enables V-TOY to perceive the different colors of clothes.

Face recognition. V-TOY recognizes faces of familiar people and acquires information about unfamiliar ones.

Behavior Capabilities

Tracking humans. V-TOY uses its eyes to maintain eye-contact with a person and rotates its neck to track the person.

Collecting and learning information about its environment. V-TOY seeks to expand its knowledge of people it interacts with. It uses its face database to learn about new people in its environment and learns about new colors by labeling the colors of a person's clothes.

Mimicking face expressions. V-TOY observes and mimics the facial expressions and actions of the interacting person.

Verbal dialogues. V-TOY carries out simple verbal dialogues during interactions.

Audio Capabilities

Speaker localization. V-TOY integrates audio cues along with video cues to localize people. If a person is out of sight, V-TOY focuses its attention on that person using the acoustic information detected by a microphone array.

Natural speech recognition. V-TOY employs Via Voice™, a speech recognition and generation software, to understand simple commands and respond to them.

Team Contact Information

Yaser Yacoob (Computer Vision)
Ramani Duraiswami, Dimitry Zotkin (Acoustic Localizaion)
UMIACS
University of Maryland
College Park, MD 20742
http://www.umiacs.umd.edu/~yaser

Ismail Haritaoglu (Computer Vision)
David Koons (Multimodal Interaction and V-TOY Designer)
Alex Cozzi (Computer Vision and Software Engineering)
Myron Flickner (Computer Vision, Attentive Environments)
IBM Almaden Research Center
650 Harry Road, San Jose, CA 95120
http://www.almaden.ibm.com/cs/blueeyes