Complex Continuous Meaningful Humanoid Interaction: A Multi Sensory-Cue Based Approach

Gordon Cheng*

Humanoid Interaction Laboratory
Intelligent Systems Division
Electrotechnical Laboratory
Tsukuba, Ibaraki, JAPAN
email:gordon@etl.go.jp
http://www.etl.go.jp/~gordon

Yasuo Kuniyoshi

Humanoid Interaction Laboratory
Intelligent Systems Division
Electrotechnical Laboratory
Tsukuba, Ibaraki, JAPAN
email:kuniyosh@etl.go.jp
http://www.etl.go.jp/~kuniyosh

Abstract

Human interaction involves a number of factors. One key and noticeable factor is the mass perceptual problem. Humans are equipped with a large number of receptors, equipped for seeing, hearing and touching, to name just a few. These stimuli bombard us continuously, often not on a singular basis. Typically multiple stimuli are activated at once, and in responding to these stimuli, variations of responses are exhibited.

The current aim of our project is to provide an architecture, that will enable a humanoid robot to yield meaningful responses to complex and continuous interactions, similiar to that of humans.

In this paper we present our humanoid, a system which is able to simultaneously detect the spatial orientation of a sound source, and is also able to detect and mimic the motion of the upper body of a person. The motion produced by our system is human like – ballistic motion. The focus of this paper is on how we have come about the integration of these components.

A continuous interactive experiment is presented in demonstrating our initial effort. The demonstration will be in the context of our humanoid interacting with a person. Through the use of spatial hearing and multiple visual cues, the system is able to track a person, while mimicking the persons upper body motion. The system has shown to be robust and tolerable to failure, in performing experiments for a long duration of time.

1 Introduction

In viewing everyday life, human interaction can be well regarded as being complex and continuous. The overall outcome of the interaction, typically involves a large number of factors (for example, seeing, hearing and touching), enmeshed together – in a cooperative and competitive manner – in an interplay of production. Physically this interplay must encompass a large number of stimuli, that in turn brings forward a large number of cues, cooperating and competing to gain the attention of an individual. Hence, each individual cue plays some role in influencing the outcome, and not one single cue assumes the sole responsibility for outcomes. This view has been shared across a wide number of disciplines, see [1, 2, 3, 4, 5].

We believe these ideas provide a powerful clue as to how humanoid interaction should be. Nature also provides to us the knowledge that the inner working mechanisms should function as a whole, not purely as individual components.

Our initial aim is to brings forward these views, in producing a simple[1] but yet effective architecture for the integration of a multi sensory humanoid system. The architecture should be able to yield complex and seamless interaction between a humanoid robot and its environment.

In this paper we present our humanoid, a system which is able to simultaneously detect the spatial orientation of a sound source, and is also able to detect and mimic the motion of the upper body of a person. The motion produced by our system is human like, ballistic motion. The focus of this paper is on how we have come about at the integration of these components.

*Currently supported by the Science and Technology Agency(STA) of Japan, as a STA Fellow.

[1] simple in the sense that the use of models will be avoided as much as possible.

Our initial attempt takes into account the following considerations:

seamlessness complex and continuous humanoid interaction should be seamless, that is it should not be obvious to an observer that the system is in one particular mode or another. Therefore, a system should be able to provide a qualitative appearance which is obvious to an external observer. Thus, the sensory information should appears to be cooperating in an interplay at producing the overall outcome of the system.

adaptivity/redundancy in handling of failure in sensory perceptions. Redundancy in the way that if one sensor fails, the system should not come immediately to a halt[2], thus yielding a robust system. We also believe this is the initial prerequisite for a system to support *Self-Preservation*[3].

dominance/competition is related to the issues mentioned above. As discussed, sensory perception tends to be in a way that it is competing for the attention of the beholder. This provides an additional clue that this feature should form part of the integration. By incorporating this characteristic, a system will intrinsicly embrace the property of adaptivity/redundancy without the need for explicitly detecting failures.

flexibility in a way which additional sensor(s) and/or cue(s) can be integrated easily. This compels us to seek a simple and effective internal structure.

basic integrator from the discussion so far, we will need a simple common integrator which is flexible enough that it can yield seamlessness, supports adaptivity/redundancy and also allows dominance at all levels of processing.

natural environment the environment in which the humanoid occupies should remain unmodified – unmodified in anyway to accommodate for any special perceptual need.

multiple input/multiple output consideration should be taken into account for a large number of sensors, and a wide range of concurrent responses should be exhibitable.

human-like motion a humanoid should respond with smooth human-like motions.

self-regulated motion this attribute is exemplified by the way in which our body works. If motion were to be produced by our own body in response to stimuli, we are usually aware of our own limitations. With the help of proprioception, information of the joint limits can be inferred. In other instances, motion of one joint can also influence the motion of others, such as the tonic neck reflex action [1, 5].

The above considerations may appear to be complex and overwhelming, but our aim is to explore and search for a better and simpler solution. To demonstrate the considerations we have taken above, we present our humanoid in a continuous daily activity of play. This interaction takes place in our laboratory. The interaction takes into play, spatial hearing and multiple visual cues. The response of the humanoid entails a number of self regulating motions, including, but not exclusively to, auditory and visual servoing. A task of mimicking the motion of the upper body of a person by sight, forms part of this interaction.

A discussion of some past approaches to humanoid research is presented in Section 1.1. Section 2 provides a description of our humanoid robot. A session of interaction with our humanoid is presented in Section 3. In Section 4 we present the components that form part of our system architecture. Section 5 presents the integrated system, and a simple architecture which combines these components together into one seamless continuous interactive system. Finally a summary and conclusion is given in Section 6.

1.1 Previous work

In the past, successful humanoid researches have mainly focused toward the development of a human-like robot which performed a particular task. Such as the humanoid of Kato *et. al*, a robot which was able to sight read music while playing a musical instrument, accompanied by an orchestra [6].

At the MIT AI lab they have taken the approach in demonstrating a large number of highly functional sub-systems. Effectiveness of these individual components have proven to be prosperous. However, each sub-system has been developed assuming complete control over a particular system resource. Little focus has been placed on the integration of these sub-systems into a fully coherent functioning system. They share the view that a fully integrated system remains an interesting and important issue in humanoid research. Reportedly, work is underway toward such an integrated system [4].

Recently, the Waseda Humanoid Project has produced some high-level complex multi-modal humanoid systems. They have chosen to develop their systems by approaching the integration problem through modulating each subsystem via high-level mode switching [7]. They also as-

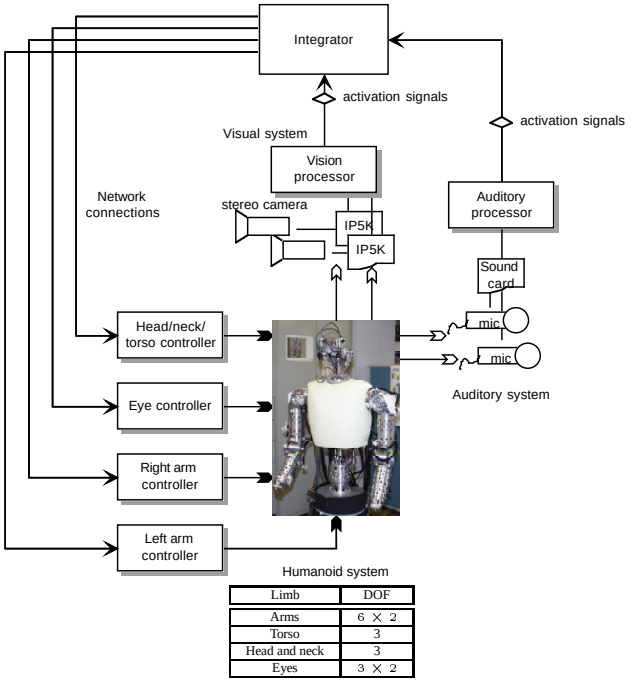
[2] to some level of course.

[3] currently under further investigation.

sumed that each sub-system has the complete control of the system once being active. In this way while the system is in one mode, the other sub-systems do not take part. We believe this is one alternative in which the integration problem can be engineered, especially when the particular task at hand can be clearly defined.

2 Configuration – ETL-Humanoid

In the current phase of development, the upper body of our humanoid robot has been completed. This initial prototype embodied two arms, head and torso, as depicted in Figure 1. This upper body provides 24 degrees of freedom: 12 d-o-f for the arms, 3 d-o-f for the torso, 3 d-o-f for the head/neck and 6 d-o-f for the eyes. Other parts of the body are still under construction. For a detailed discussion of the whole system see [8] and [9]. Motor control and sensor processing is currently performed via a set of six PCs connected to our humanoid.



Limb	DOF
Arms	6 × 2
Torso	3
Head and neck	3
Eyes	3 × 2

Figure 1: ETL-Humanoid ('JACK'): In its current form the upper portion of the body has been completed (head, eyes, arms and torso). Currently six PCs are being used: four for motor control and two for vision and audio processing.

3 Complex Humanoid Interaction

An example session of complex humanoid interaction is presented in Figure 6 and 7. These figures show our humanoid continuously interacting with a person in our laboratory. It starts by tracking a person in an unstructured environment, followed by mimicking the upper body of the person. Once the humanoid has lost track of the person the system continues to be influenced by other cues perceived from its surroundings. When a sudden loud noise was detected, the humanoid saccaded toward the sound source. The system then noticed and detected a person and visually re-locked on to the person.

In a continuous sense, this experiment presented was from a single take, which ran for 4 minutes and 33 seconds. Some of our experiments go on for quite a long length of time, some lasting over 20 minutes. This further demonstrates the robustness of our system, and satisfying our aim of producing a continuous interactive system.

The action demonstrated by our system is referred to as *Simultaneous Imitation* (taken from [10]). At this time, we do not claim that our humanoid is currently performing *Imitation learning*, but we believe this is clearly one step toward the stages of *Imitation Learning*.

4 Components

In this section we present the components available to our humanoid. Our discussion will be focused on the topic of interaction. Each component is introduced in the context of providing and facilitating humanoid interaction. A *Basic integrator* is introduced in Section 4.1. A discussion of the auditory processing is presented in Section 4.2. Vision processing is presented in Section 4.3. Motor control of each joint is presented in Section 4.4.

4.1 Basic integrator

The structure we have chosen is a non model-based structure, which only entails two attributes: an action vector and an activation potential, which is associated with each action vector.

The key features of these two attributes are as follows:

action vector providing the magnitude and direction of a given input. e.g. a vector can be used to represent the relative action of the arm, positive for up, negative for down. Its speed being represented by its magnitude.

activation potential provides temporal duration of its associated **action vector**, representing the degree pres-

ence or the absence of a particular stimulus/cue given by the vector, determining its reliability, i.e. confidence.

Inspired by the generality of a biological neural system. The key and central idea of this *Basic integrator* must be applicable across many levels, at both the sensory level and the actuation level, as a neuron would be in a biological system. Due to the complex nature of such an integration, it must be able to satisfy all the requirements stated in Section 1.

We introduce Equation (1), as our *Basic integrator* for use throughout our system. As discussed, the important properties of this integrator is that it is model-free, it can be used at many levels, from sensory processing to the final output of the system.

$$U_i(t) = \frac{\sum_k \alpha_k(t) a_k(t) v_k(t)}{\sum_k a_k(t)} \quad (1)$$

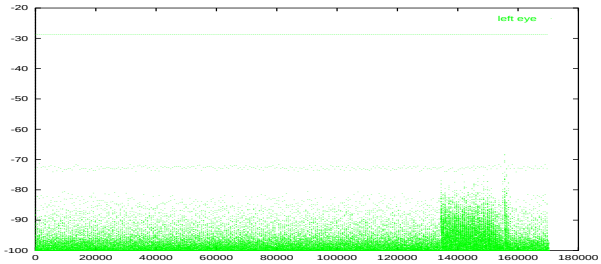
where k is the index for each relevant input. i is the index for each i_{th} output. $U_i(t)$ is the i_{th} output vector at instant t. $a_k(t)$ is the activation potential of the k_{th} input at instant t. v_k is the k_{th} input vector . $\alpha_k(t)$ is the parameter which allows the alteration of the strength of a particular input.

Although, currently not used, the parameter $\alpha_k(t)$ was introduced for the alteration of the overall system behaviour. This is inspired by the daily interaction of a person. Influences from sensory systems tend to be alter based on some selective occasion, depending on the *mood* of an individual at that particular time. Many other factors also comes into play, a well know phenomenon exhibited by a person, is the decay in response to a continuous stimulus over some duration of time [1, 2, 5].

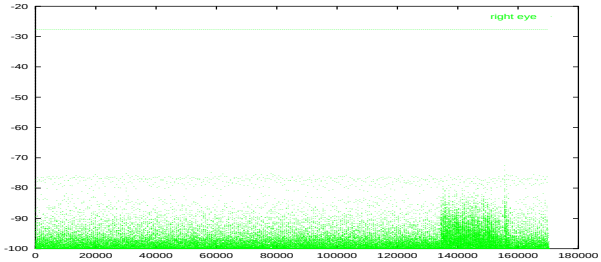
4.2 Auditory response – Spatial hearing

In our examination of auditory processing, we provided the ability for our system to perform left and right spatial discrimination. For instance, auditory servoing can be achieved by moving the head/neck in a pan motion, while minimising the volume of the left and right ears.

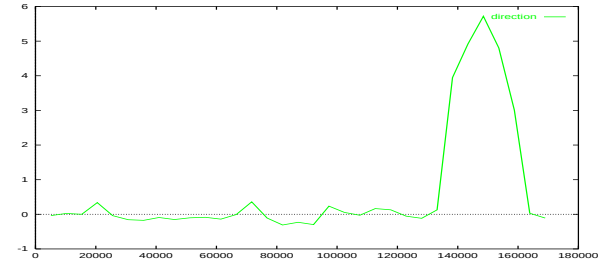
The technique we have employed is a process of interaural processing, for a comprehensive coverage of the subject see [11]. The sound source from each ear is processed by a Fast Fourier Transform (FFT), producing a power spectrum for each channel (see Figure 2a and b), the next process is then simply by taking the difference of each spectrum with its corresponding frequency. In yielding an output of the direction and magnitude of the sound source, see Figure 2c. The importance of this final stage is that it produces a magnitude and spatial orientation of the sound



a) Power spectrum from left ear



b) Power spectrum from right ear



c) Magnitude and spatial orientation of hearing

Figure 2: Auditory processing: a) and b) shows the power spectrum of sound input taken from the left and the right ear, c) shows the spatial orientation of the sound source.

source, which can be use as a vector for the integration in the final process. The activation potential is calculated based on a threshold, activation is increase if the threshold is reached. An active thesholding is currently under investigation.

The current processing is performed using a SoundBlaster™ card installed on a PC running the Linux/OS. With this configuration we were able to sample the stereo sound channels at 22kHz, and outputting a result at 5Hz.

4.3 Visual response

The human visual receptors are the single most developed and heavily utilised organ of our perceptive system. Therefore we chosen to provide as many visual cues as possible to our humanoid system. Currently the system response to the following cues: motion detection, disparity, skin detec-

tion, at the higher level person detection and upper body motion tracking (head, left and right arm).

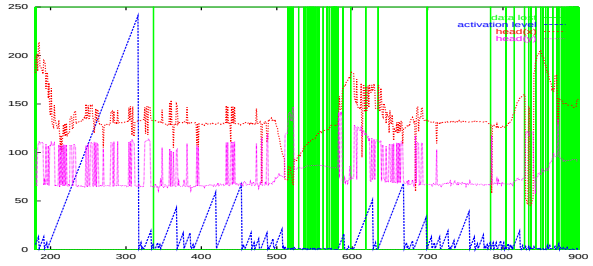
In the current stage of our research, we have integrate the head and upper body motion detection with the auditory response, as discussed in the previous section.

The skin colour detection is based on Colour distance and Hue extraction. The extraction is performed on a pair of Hitachi IP5005 vision processor cards, installed on a single PC running the Linux/OS. The vision processing is performed in real-time, at 30Hz.

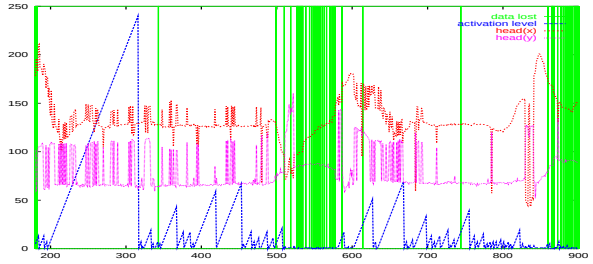
Figure 3 shows the output from our head detector, the upper two figures shows the detection performed by each eye. The figures show both the location of the head, and its corresponding activation. The tracking of the head is facilitated by a probability distribution, introduced to reduce the problematic noisy data.

Figure 4 shows the results of tracking the motion of a human body. First the left and right eye inputs are processed. Once processed the output is then merged. Since we are only interested in determining the arm motion of the person being tracked. We can take advantage of its derivative information. The derivative provides the trajectory information of the arm, moving up or down, and/or, side to side. This is used as the action vector, and its activation potential will be used in the final stage of integration.

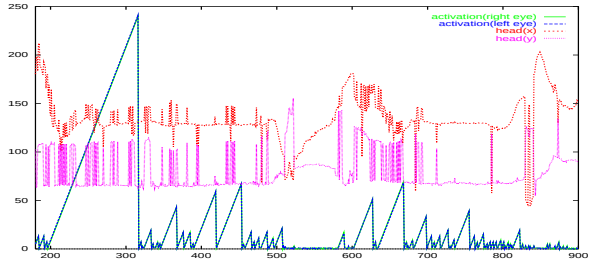
The activation potential of these processes is calculated based on the presence of each of the visual cues. The activation increases as long as the cue exists. The upper portion of the Figures 3 and 4 show the activation level of each signal, and the loss of these signals is indicated by vertical lines.



a) Left eye – head detection



b) Right eye – head detection



c) Combined – head detection

Figure 3: Head detection: a) and b) shows the output of the left and right eye tracking the head of a person, the vertical lines show the loss of data while tracking, caused by sensory noise, c) shows the final combined output. The activation potential is provided on the bottom of each plot.

4.4 Motor control – Humanoid motion

The aim at the level of motor control is to provide flexibility in the way the motor can be used. At each joint the following motor control schemes have been implemented on our humanoid system: current(force), velocity, angular and position control. These schemes run with a conventional **P**roportional **I**ntegral controller. The controls of each joint can be commanded via any of the above schemes, in a flexible way. The motor can be controlled in a mixed fashion. For instance, in the current experiment (see Section 3) the motion of the arm is driven at both velocity and current level. The motion of the arm is control via velocity, but once no motion is required the arm is commanded to fall into a zero current loop. Hence, allowing the arm to be free and compliant, allowing human-like ballistic motion to be achieved.

Self-regulating motion is achieved through the monitoring of the encoder at each joint. The joint limits are set in two ways, *a priori* at the start, and through physical interaction. Physical interaction, is done by taking advantage of the compliance of the system while it is not in motion. A person may physically move the robot, the system monitors this movement. The new limit of the joints is determined by the upper most position reached.

4.5 Motor mapping and motor output

The current mapping of cues to motor action has been done *a priori*, although neurological experiments have shown that this maybe an innate ability that is available to us. We wish to leave this part of the system open for further studies.

During this early stage of development we wish to focus on the issues of integration. Therefore, we have selected a

simplified mapping scheme for motor output. The current mapping between human and humanoid is done directly. The corresponding arm motion of the person is mapped directly to the output of the corresponding humanoid arm motion. The control mappings are as follows, spatial hearing and detected head motion, controls the head/neck/torso motion. The rotation of the torso and the head allows the humanoid to keep track of the person in the horizontal direction. While the neck moves in the vertical direction to ensure a full view of the person is seen. Each arm is mapped in the same way, the vertical motion of the detected arm is mapped to the motor joint at the elbow and at the shoulder (vertical – allows the shoulder to move in the forward/backward direction), the horizontal motion of the arm is mapped to the second motor on the shoulder (horizontal – allows the shoulder to move in the outward/inward direction).

These mappings produced a number of motion primitives, individual arm motion, up and down, and side to side. Some motions that have been realised based on these primitives include, swinging each arm in and out of phase; swinging side to side while moving the arms up and down. The production of these motions are shown in Figures 6 and 7.

5 Putting them together

As discussed in Section 1.1, past approaches have tended to allow each sub-system to take the complete control of system resources once it has been active. Our development so far has not taken this assumption, rather we have chosen to integrate them together based on their action vectors and their activation potentials. By using the *Basic integrator* given by Equation (1) and the motor perception mapping discussed in Section 4.5. Figure 5 shows the output of this final processing in determining the humanoid motion. Figures 6 and 7 shows the motion performed by the humanoid robot while observing a person.

6 Summary and Conclusions

This paper presented a number of ideas in the integration of a multi-sensory humanoid system, which is able to yield a large number of simultaneous responses. Our humanoid system was able to interact via auditory, physical and multiple visual stimuli. Human-like motion was produced in response to the stimuli. A complex interaction of mimicking the upper body motion of a person was exhibited by our system. The humanoid in its interaction has shown to be robust and continuous.

The key ideas of this paper can be summarised as follows:

Integration should be seamless in such a way that no one part of the overall system should be allowed to dictate the system resources, rather it should be integrated in a continuous manner.

The mechanism used should be able to combine and yield a mix of adaptivity, redundancy and flexibility.

Acknowledgments

We wish to acknowledge kindly the support of the COE program funded by the Science and Technology Agency(STA) of Japan.

References

[1] J. A. S. Kelso, ed., *Human Motor Behavior: An Introduction*. Lawrence Erlbaum Associates, Publishers, 1982.

[2] A. R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*. Avon Books, 1994.

[3] D. C. Dennett, *Kinds of Minds*. Science Masters series, Basic Books, 1996.

[4] R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson, "The Cog Project: Building a Humanoid Robot," in *IARP First International Workshop on Humanoid and Human Friendly Robotics*, (Tsukuba, Japan), pp. I–1, October 26–27 1998.

[5] R. A. Schmidt and T. D. Lee, *Motor Control and Learning: A Behavioural Emphasis*. Human Kinetics, third ed., 1999.

[6] I. Kato, "Wabot-2: Autonomous Robot with Dexterous Finger-Arm," in *Proceedings of IEEE Robotics and Automation*, vol. 5 of *2*, 1987.

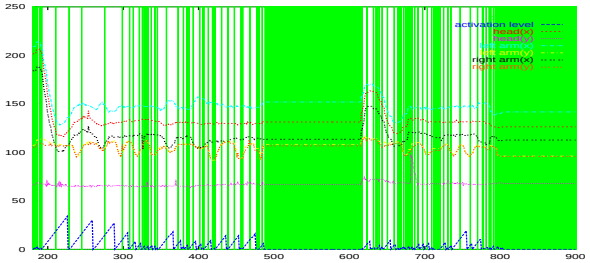
[7] S. Hashimoto *et. al.*, "Humanoid Robots in Waseda University – Hadaly-2 and WABIAN," in *IARP First International Workshop on Humanoid and Human Friendly Robotics*, (Tsukuba, Japan), pp. I–2, October 26–27 1998.

[8] Y. Kuniyoshi and A. Nagakubo, "Humanoid Interaction Approach: Exploring Meaningful Order in Complex Interactions," in *Proceedings of the International Conference on Complex Systems*, 1997.

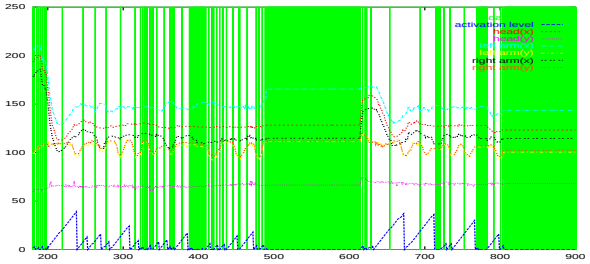
[9] Y. Kuniyoshi and A. Nagakubo, "Humanoid As a Research Vehicle Into Flexible Complex Interaction," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'97)*, 1997.

[10] S. Schaal, "Is imitation learning the way to humanoid robots?," *Trends in Cognitive Sciences*, 1999.

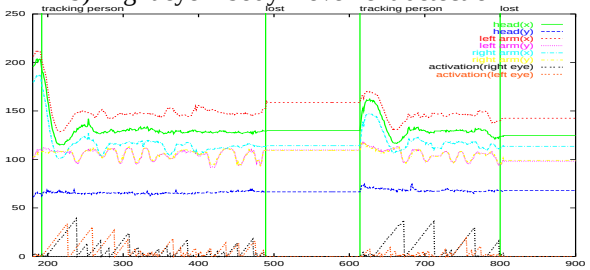
[11] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, revised ed., 1999. Second printing.



a) Left eye – body movement detection

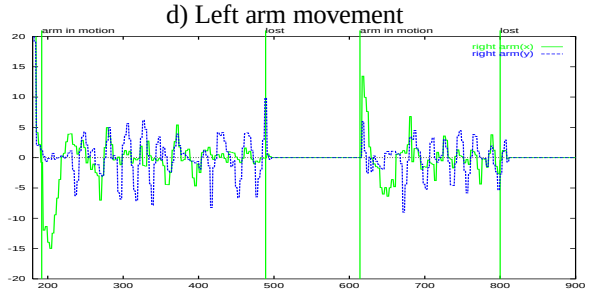


b) Right eye – body movement detection



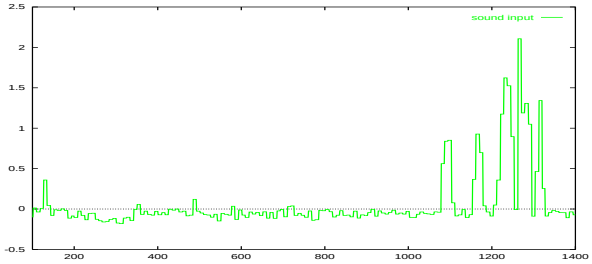
c) Combined – body movement detection

d) Left arm movement

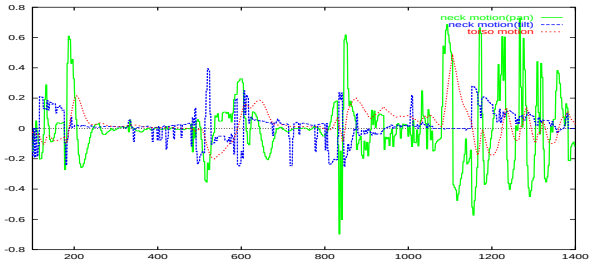


e) Right arm movement

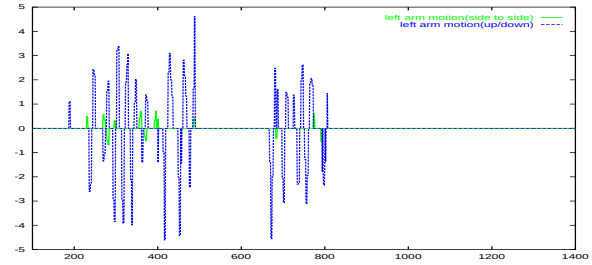
Figure 4: Body motion processing: a) and b) shows the upper body motion of a person, c) shows a combined version of these data and their activation potential, d) and e) shows the final determined motion.



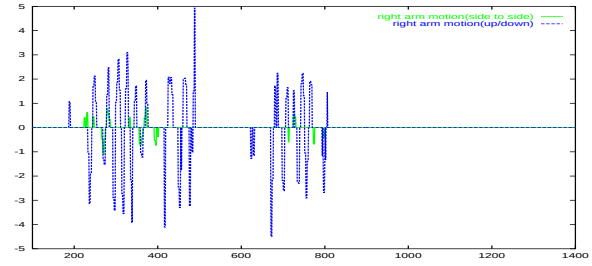
a) Auditory influence



b) Body motion – head/neck/torso



c) Left arm motion



d) Right arm motion

Figure 5: Humanoid motion: a) shows the auditory influence, b) is the upper body motion without the arms, c) and d) shows the motion of the arm.

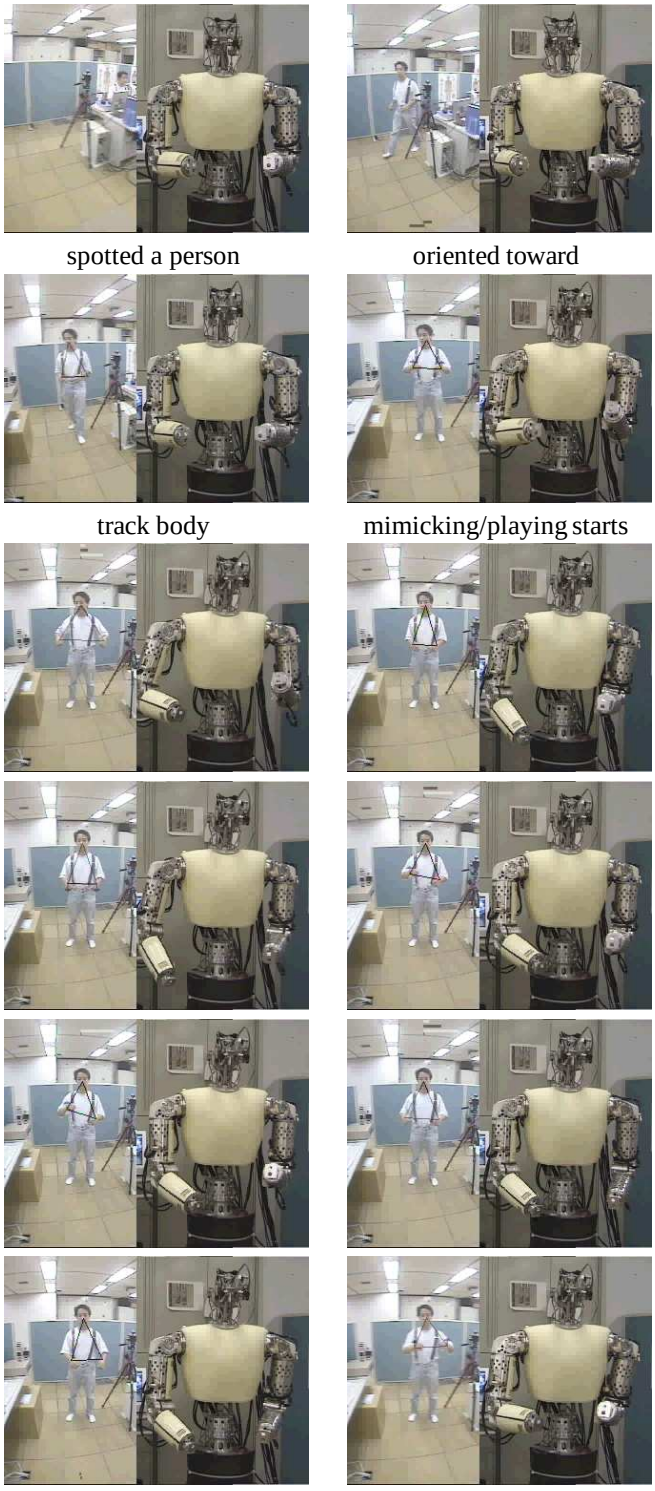


Figure 6: Interaction experiment – part one

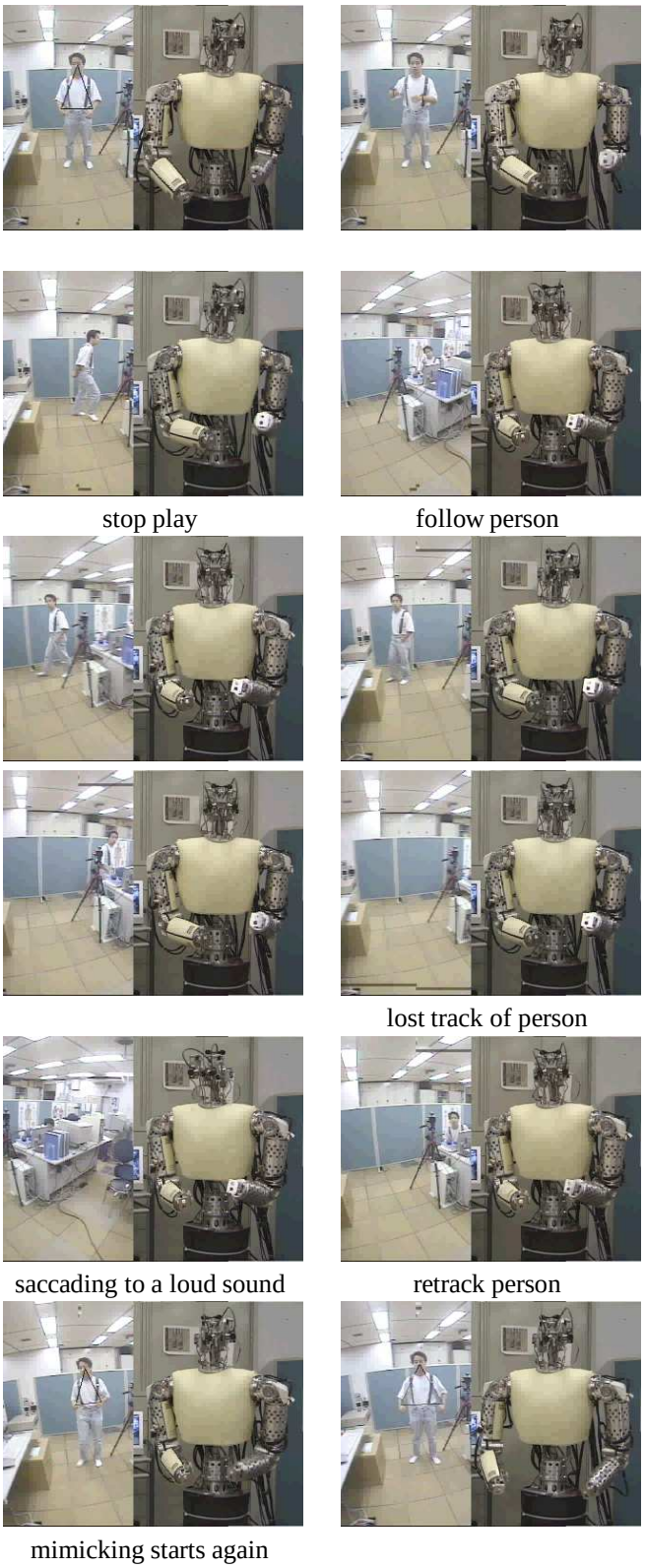


Figure 7: Interaction experiment – part two