Modelling Intelligent Virtual Agent Skills
with Human-Like Senses

Pilar Herrero and Angélica de Antonio

Facultad de Informática. Universidad Politécnica de Madrid.
Campus de Montegancedo S/N.
28.660 Boadilla del Monte. Madrid. Spain
{pherrero,angelica}@fi.upm.es

Abstract. The research work presented in this paper represents a significant advance, specifically in the area of perception, within the multidisciplinary field of Intelligent Virtual Agents (IVAs) and multi-Intelligent Virtual Agent Systems (mIVAS) .Within the cognitive research area there are some studies underwriting that human perception can be understood as a first level of an "awareness model". Bearing in mind these researches, we have developed a human-like perceptual model based on one of the most successful awareness models in Computer Supported Cooperative Work (CSCW), called the Spatial Model of Interaction (SMI), which has been applied to Collaborative Virtual Environments (CVEs). This perceptual model extends the key concepts of the SMI introducing some human-like factors typical from human being perception as well as it makes a reinterpretation with the aim of using them as the key concepts of a IVA's human-like perceptual model.

1   Introduction

Nowadays, virtual environments often incorporate human-like embodied virtual agents with varying degrees of intelligence, getting what we call Intelligent Virtual Agents (IVAs).

An IVA may simply evolve in its environment or it may interact with this environment or even communicate with other IVAs or humans, but, in order to make this interaction possible, an IVA has to be aware of its environment. This awareness can be understood as the result of an IVA perceiving its surroundings and therefore those object/agent which are in the environment and nearby to the observer IVA.

The research that we present in this paper is precisely oriented towards endowing IVAs with perceptual mechanisms that allow them to be "realistically *aware*" of their surroundings. We propose a perceptual model, which seeks to introduce more coherence between IVA perception and human being perception. This will increment the psychological "*coherence*" between the real life and the virtual environment experience. This coherence is especially important in order to simulate realistic situations as, for example, military training, where soldiers must be trained for living and surviving risky situations. A useful training would involve endowing soldier agents with a human-like perceptual model, so that they would react to the same stimuli as a human soldier. Agents lacking this perceptual model could react in a non-

realistic way, hearing or seeing things that are too far away or hidden behind an object. The perceptual model we propose in this paper introduces these restrictions inside the agent's perceptual model with the aim of reflecting more faithfully a human-like perception.

Having in mind that the physical perception can be understood as the first level of an "awareness model" [2] the first goal for our research was to select a model of awareness which could be valid for our purposes. As the "Spatial Model of Interaction" (SMI) [1] used the properties of the space to get knowledge of the environment, it was based on a set of key awareness concepts – which could be extended to introduce some human-like factors - and it had been tested with successful results in CSCW multi-user environments, this model had the essential qualifications for our purposes and we selected it in order to develop the perceptual model we propose in this paper.

2   Key Concepts in the Perceptual Model

As we mentioned in previous sections, the key concepts of our Perceptual Model are based on the main concepts of a CSCW awareness model known as *The Spatial Model of Interaction (SMI)* [1].

2.1   Key Concepts in the SMI

The spatial model, as its name suggests, uses the properties of space as the basis for mediating interaction. It was proposed as a way to control the flow of information of the environment in CVEs (Collaborative Virtual Environments). It allows objects in a virtual world to govern their interaction through some key concepts: awareness, focus, nimbus and boundaries. In this paper we are going to concentrate on the three first, leaving the boundaries out of the paper's scope.

Focus. "The more an object is within your focus the more aware you are of it". The focus concept has been implemented in the SMI as an "ideal" triangle limited by the object's aura. In the same way, it is possible to represent the observed object's projection in a particular medium, this area is called *nimbus*: "The more an object is within your nimbus the more aware it is of you". The implementations of these concepts –focus and nimbus- in the SMI didn't have in mind human aspects, thus reducing the level of coherence between the real and the virtual agent behaviour. *Awareness* between objects in a given medium is manipulated via *focus* and *nimbus*, controlling interaction between objects.

2.2   Making the Perceptual Model More Human-Like

There are many factors that contribute to our ability as humans to perceive an object, some of which are directly working on the mental processes, being not easily modelled or reproduced in a virtual world. We have selected the two most useful senses: sight and hearing and we have analysed some human perception key concepts

to determine which of them could be introduced in our agent's perceptual model. These concepts, selected for being the more representative of human visual and hearing perception, are [3,4]:

- Sense Acuity: It is a measure of the sense's ability to resolve fine details In a visual medium, it is known as Visual Acuity. In a hearing medium, it is know as Auditory Acuity There are two different kinds of auditory acuity: spatial acuity and frequency acuity. Both acuities are inter-related.
- Sense Transition Region: It is the interval in the space between perfect and null perception. This factor plays an important role in a visual medium where it is known as Lateral Vision. The Lateral Vision corresponds to the visual perception towards the extremes of the visual focus. In a hearing medium this concept can be understood as the cone in the space known as Cone of confusion. The cone of confusion is a cone extending outwards from each ear. Sound events that originate from a point in this cone is subject to ambiguity.
- Filters: Allowing the selection, from all the objects in an extensive focus, of only those that the agent is especially interested in.
- Directivity of sound: This factor , associated to a hearing medium, represents the directional characteristic of the sound source propagation.

2.3   Reinterpreting the SMI's Key Concepts

Neither the SMI nor its implementations considered aspects of human perception. Thus, if the SMI were applied just as it was defined by Benford, the level of coherence between real and virtual agent behaviour would be minimum. We have decided to identify the factors concerning human-like perception which provide more realistic perception, and introduce them into the SMI. In this section, we are going to describe how the key concepts defining the SMI have been modified to introduce these human factors.

Focus. Benford introduced the focus concept in 1993 as "*The more an object is within your focus the more aware you are of it*" [1]. This concept meant that the observing object's interest for each particular medium could be delimited. According to this definition, the focus notion is the area within which the agent is perceiving the environment. We have analysed how sensitive perception works in humans beings, and, from this analysis, we have decided to select some factors that should have an effect on the physical area delimiting the observing object's interest.

Visual Focus. In a visual medium, these factors are the *Sense Acuity* and the *Sense Transition Region*. Starting from the focus concept in the spatial model, and bearing in mind previous implementations, where focus was implemented as a triangle, we have defined a new mathematical function to represent the human-like visual focus concept.

   This mathematical function describes the focus as a double cone centered on the agent's eye and which length is limited by agent's visual resolution acuity. Each of these two cones is dependent on the opening angle, which will be the agent's foveal angle of vision in the foreground region and the agent's lateral angle of vision in the lateral region of perception.

Auditory Focus. A human listener can hear a sound arriving to him in two different ways: combining information from the two ears, *"Binaural Hearing"*, or taking information from one ear or from each ear independently, *"Monaural Hearing"* .Modern psychoacoustic research has turned its attention to binaural hearing [4] and, for this reason, we are also going to focus on this kind of hearing along this research.

Sound waves usually travel out in all directions from the source of sound, with an amplitude that depends on the direction and distance from the source. We have represented the IVA's hearing focus concept by an sphere circumscribing the agent's head.

Nimbus. Benford introduced the nimbus concept in 1993 as *"The more an object is within your nimbus the more aware it is of you"* [1]. This concept meant that the observed object's projection for each particular medium could be delimited. The nimbus concept, as defined in the Spatial Model of Interaction, has always been implemented as a circumference in both visual and hearing media. The radius of this circumference has an "ideal" infinite value, although, in practice, it is limited by the object's aura. Just as with the above-mentioned focus concept, the nimbus concept in the Spatial Model of Interaction does not consider any human factors, thus hypothetically reducing the level of coherence between real and virtual agent behaviour.

Visual Nimbus. In a visual medium, we are going to represent the nimbus of an object as an ellipsoid or a sphere, depending on the conic by which the object is circumscribed, centred on its geometrical centre. The way of practically determining which conic has to be associated with each object in the environment is to look for the bounding box that has been associated to this object in the environment. If the bounding box is a rectangle, we will approximate the nimbus as an ellipsoid; if the bounding box is a circle, then we will approximate the nimbus as a sphere. The nimbus radius, or its eccentricity if it is an ellipsoid, will depend on two factors: the object's shape and the furthest distance at which a human being would be able to distinguish the object. This distance is determined by visual acuity, which depends on the object's size; thus, indirectly, the nimbus conic will depend on the object's size as well.

Auditory Nimbus. In a hearing medium, the nimbus delimits the physical area of projection of a sound source for a particular medium. Before modelling the nimbus shape, it is important to analyse the way of projecting the sound, but also the factors that can have an effect on the physical projection of the sound.

We start from the assumption that the sound is propagated in the medium by a spherical wavefront, but even if this occurs, it could happen that the sound amplitude, and therefore its intensity, weren't the same in all the directions. For this reason, in this model we interpret the nimbus concept as the region within which the sound source is projected with the same intensity.

Starting from this interpretation of nimbus we can determine which factors have an influence on this concept and its representation within an environment. From all these factors, in this research we have just considered directivity of sound, living the rest of the factors, as for example, the presence of non-linear effects or the homogeneity of the medium, for future research and extensions of this work.

We have also centered our research in the projection of human voice. The directivity pattern associated to the human voice is a cardioid representing the sound pressure level (SPL) - in decibels (dB) - versus horizontal angle and being the direction of speaker's mouth located at 0 degrees.

3   Clarity of Perception

In the previous sections we have introduced the key concepts of our perceptual model. In this section we are going to concentrate on how to use these concepts in order to endow IVAs with a human-like perceptual model.

In order to make this possible, we have selected an appropriate agent architecture in which to introduce this perceptual model. From all the possible architectures, we have selected a vertical layered architecture which is composed by three different blocks: Perception, Central Processing and Action. We have also proposed several modules to form the agent's perception block being one of them the *Sensitive Perception.* The *Sensitive Perception* module simulates the typical perceptual process by which organisms receive sensations from the environment, depending on some relevant sensorial concepts. More details of this design and its implementation are given in [3,4]. The IVA's sensitive perception module will calculate what we have called *Clarity of Perception,* taking into account the focus of the IVA and the nimbus of those objects that are surrounding this agent.

Clarity of Perception is a measurement of the ability to perceive an object (or a sound) inside the agent's visual focus, as well as the clearness of this perception. Once an agent's focus intersects with an object's nimbus, this agent is aware of this object and therefore the sensitive perception module will calculate the clarity of perception for this object.

Following the research conducted by Levi et al. [5,6] in a visual medium and Shinn-Cunningham [7] in a hearing medium, we propose several mathematical function to describe the variation in the clarity of perception with the eye/ear-object/sound distance. More details are given in [3,4].

4   Perceptual Model Implementation

This model has been implemented in Visual C++ as a library. This library has been integrated with MASSIM_AGENT. MASSIM_AGENT was the prototype system built using the MASSIVE-3 – a CVE system - and the SIM_AGENT – a toolkit for developing agents. MASSIM_AGENT was the result of a collaboration established between the Mixed Reality Laboratory (MRL) at the University of Nottingham and the Universidad Politécnica de Madrid (UPM). More details of its implementation are given in [3,4].

5   Some Scenarios for Human-Like Perception

In order to prove the usefulness of the proposed perception model, lets consider that, as it was previously mentioned, mIVAS can be used to simulate risky situations, as for example, a world-wide war, where the soldiers' training plays a very important role. In this kind of systems, soldiers can be trained for living and surviving the worse real-life situations. To get a useful training, it is important to endow soldier agents with a human-like perception model. Different scenarios and situations can be raised where human-like perception plays a very important role. In this section we are going to describe some of them.

5.1   Visual Perception

Lets imagine that a soldier agent is at the battlefield. He is placed at a physical position given by the co-ordinates (x,y,z)= (1,0,0) in the space, in meters, with an orientation of 90º related to the x axis of co-ordinates. This soldier is endowed with a visual acuity (in Snellen notation) equivalent to 20/20 and his foreground angle of vision is $\theta=30º$ while his lateral angle of vision is $\theta'=65º$. Lets also imagine that a fighter plane, a Focke-Wulf BMW whose size is (length x wide x height)=(20,15,7), in meters, appears in the air space. Introducing all these values in the implemented visual perceptual model, we get the foreground and lateral soldier's cone of vision. In the same way, we get the nimbus geometry associated to the plane, which in this case is an ellipsoid, and the plane's nimbus. In the same way, the perceptual model calculates the maximum distance of resolution for an object of the plane's size (D_m), which in this case is 64.20 m. When the plane is placed at co-ordinates (x,y,z)=(1,25,0), in meters, away from the soldier, it is possible to perceive the object in the foreground area of perception. The clarity of perception in this area gets the maximum normalised value 1. The soldier can perceive most of the plane's details. When the plane is placed at co-ordinates (x,y,z)=(-40,25,25), in meters, away from the soldier, it is possible to perceive the object in the lateral area of perception. The clarity of perception in this area is 0.5. The soldier can perceive just a few details of the plane. The soldier will probably perceive an object but he will not be able to identify its physical details. As the plane is placed in the lateral area of perception, the details that this soldier can get from the plain will be associated to its movement. When the plane is placed at co-ordinates (x,y,z)=(15,70,25), in meters, away from the soldier, it is possible to perceive the object in the foreground area of perception. The clarity of perception in this point is 0.033 (very low). Maybe the soldier can perceive the plane's shape and its movement but this soldier can make a mistake confusing the coming plane with a friendly plane instead of recognising it as the hostile plane that it really is.

5.2   Auditory Perception

Let us imagine that a soldier agent (A) is in the countryside, keeping watch. He is placed at a position (x,y,z)=(0,0,0) in the space, in metres. The soldier azimuth angle is 80 degrees and his elevation angle is 87.5 degrees. At this moment, a couple of people approach him and one of them starts speaking when at co-ordinates

Fig. 1. The war scenario in MASSIM_AGENT

(x,y,z)=(1,2,0). As they are having a moderate conversation, the intensity of the sound that this person is emitting is 60dB. In this case, the intensity of the sound that is reaching soldier agent A, and which is calculated by our perceptual model, is 11.22 dB. Then the method that implements this perceptual model determines that this intensity of sound is between the standard thresholds of hearing and, therefore, the agent can hear it. However, the clarity of perception with which this agent can hear this sound is *Very_Low*. Soldier A can hear a sound, but he cannot understand what they are speaking about. If the same couple continue walking and the same soldier continues speaking with the same intensity, when he reaches co-ordinates (x,y,z)=(0.50,0.75,0) the perceptual model determines that the intensity of the sound that is reaching soldier agent A is 42.63 dB, determining that the sound can be heard by soldier agent A because it is between the standard thresholds of hearing, and calculating that this sound can be heard by soldier A with a clarity of perception *Medium*. Soldier A can perceive something of the couple's conversation but not too much. If they continue walking under the same conditions, when the speaker reaches co-ordinates (x,y,z)=(0.50,0.50,0) the perceptual model determines that the intensity of the sound that is reaching soldier agent A is 51.77 dB, returning a clarity of perception *High*. The soldier A can perceive the couple's conversation.

6   Conclusions

We have developed a human-like perceptual model for IVAs based on one of the most successful awareness models in Computer Supported Cooperative Work (CSCW), called the Spatial Model of Interaction (SMI) [1]. This perceptual model extends the key concepts of the SMI introducing some factors typical from human being perception as well as it makes a reinterpretation of the key concepts and introduces some new concepts – as the Clarity of Perception (CP)- with the aim of using them as the key concepts of an IVA's human-like perceptual model.

References

1. Benford, S.D., and Fahlén, L.E. *A spatial Model of Interaction in Large Virtual Environments*. Proc. Third European Conference on Computer Supported Cooperative Work (ECSCW'93), Milano, Italy. Kluwer Academic Publishers, 1993.
2. Endsley. M., *Design and evaluation for situation awareness enhancement*. Proceedings of Human Factors Society and Annual Meeting, volume 1, 1988.
3. Herrero P., De Antonio A., Benford S., Greenhalgh C., *A Hearing Perceptual Model for Intelligent Virtual Agents*. Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia, July, 2003.
4. Herrero P., De Antonio A. *Keeping Watch: Intelligent Virtual Agents Reflecting Human-Like Perception in Cooperative Information Systems.* The Eleventh International Conference on Cooperative Information Systems (CoopIS 2003). Catania, Sicily. Italy, 2003.
5. Levi, D.M., Hariharan, S. & Klein, S.A. *Suppressive and Facilitatory Spatial Interactions in Peripheral Vision: Peripheral Crowding is neither size invariant nor simple contrast masking*. Journal of Vision, 2, 167-177. 2002.
    `http://www.journalofvision.org/2/2/3/`
6. Levi, D.M., Klein, S.A. & Hariharan, S. *Suppressive and Facilitatory Spatial Interactions in Foveal Vision: Foveal Crowding is simple contrast masking*. Journal of Vision, 2, 140-166. 2002. `http://journalofvision.org/2/2/2/`
7. Shinn-Cunningham, BG. *Distance cues for virtual auditory space* Proceedings of the IEEE 2000 International Symposium on Multimedia Information Processing. Sydney, Australia. December 2000.