Human-Like Auditory Detection Capability for Intelligent Virtual Agents

Hung Tran
CAE USA Inc.
Tampa, FL
hung.tran@caemilusa.com

ABSTRACT

Intelligent virtual agents (IVAs) represent important components in simulated real-world environments. Significant IVA progress has been made in diverse applications, such as entertainment, gaming, telemarketing, and recently, in military training. The use of IVAs for training is mainly in task collaboration where virtual agents interact with each other or with human users. Typical usage of IVAs in military training is *virtual warfare scenarios*. IVAs with perceptual capabilities, such as vision or hearing, tend to produce results that are more realistic and, consequently, can improve training task performance. Research and development on perceptual models for IVAs focuses largely on visual perception. However, auditory perception represents one of the most fundamental perceptual aspects for human-like behavior in a virtual environment because it improves situational awareness by extending the information and feedback envelope beyond the field of view. Therefore, in an event of auditory detection, IVAs should be able to react to other virtual entities or humans participating in a same virtual scenario.

In this paper, we will present a perceptual model to predict the auditory detection capability of IVAs. Our study will describe the foundation of this perceptual model, which is based on the auditory filters of the human hearing system. We will also present the simulation framework that was used to implement this perceptual model.

When comparing the predicted and observed auditory detection capabilities, the simulation results showed a slight overestimation for the predicted detection thresholds. This overestimation indicated, for the same test conditions, the IVA's detection capability is generally less sensitive than the human hearing capability. Nevertheless, the model proposed in this paper represents a highly promising method for prediction of auditory signal detection capabilities of IVAs.

ABOUT THE AUTHOR

Hung Tran is a Tactical Systems, Aircraft Sensors, and Interoperability Group Leader at CAE USA. He joined CAE USA more than 20 years ago and has worked on the modeling and simulation of several Electronic Warfare systems. Hung is also the lead designer of the Computer-Generated Forces currently used by all United States Air Force C-130J Weapons System Trainers. Hung holds a Bachelor of Science in Electrical Engineering from Polytechnique Montréal at University de Montréal in Quebec, Canada.

Human-like Auditory Detection Capability for Intelligent Virtual Agents

Hung Tran
CAE USA Inc.
Tampa, Florida
hung.tran@caemilusa.com

INTRODUCTION

Simulated real-world environments represent a powerful tool for learning and training. Intelligent virtual agents (IVAs) can participate with people in such environments either to facilitate training tasks or to satisfy learning requirements that would normally require additional human participants. IVAs are computer-generated entities that are not only visually similar to humans but also exhibit human like behavior. Currently, virtual agents are widely used in a diverse array of applications, such as entertainment, gaming, telemarketing, and more recently, in military training. Typical usage of IVAs in military training is virtual warfare scenarios (VWS). In a VWS application, IVAs can represent either virtual team members or enemy combatants. IVA is a new emerging, multidisciplinary field and represents agent-based computing that belongs to a completely new paradigm of software applications. An IVA can be seen as a simulated system situated in a simulated environment and capable of autonomous actions to meet its design objectives. Intelligent and autonomous are difficult concepts to define; however, in the context of virtual agents, they can be defined as the capability of processing situational and environmental data and acting appropriately without the direct intervention of humans. An intelligent and autonomous virtual agent must possess the following characteristics:

- Reactive – A virtual agent should perceive the surrounding environment and respond to the interpretation of the environment in a timely fashion. This characteristic is directly related to the perceptual models used by IVAs.
- Proactive – A virtual agent should not simply act in response to its environment; it should be goal-oriented. Therefore, it must take initiative where appropriate to reach the predefined simulation objectives.
- Social – A virtual agent should be able to interact with other virtual agents or with humans collocated in the same environment.

The design of virtual environments often incorporates IVAs with various degrees of intelligence and autonomy. Virtual agents with perceptual abilities tend to produce more realistic results (Wijermans, Jorna, Jager, Vliet, & Adang, 2013). Yet, due to the lack of perceptual models, most agent-based simulation work has tackled the challenge of perception by providing complete simulation data to the virtual agents. In contrast, Reynolds (1987) argued that global information about the virtual world should not be provided to the IVA because incomplete knowledge models reality more accurately. Consequently, the perceptual model used for an IVA should represent characteristics that reflect the real system (i.e., the system that is simulating) as well as the inherent imprecision of it. An example of such inherent imprecision of our hearing system is the *auditory cone of confusion* (Carlile, 1996), which occurs when our hearing system cannot determine the direction of sounds originating along the circumference of circular conical slice, where the cone's axis lies along the line between two ears. Sound events that originate from a point on this cone are subject to ambiguity.

Human auditory perception of the surrounding environment is a very complex process, as exemplified by the *cocktail party effect* and the *auditory scene analysis. Cocktail party effect* is the phenomenon of being able to focus on one particular stimulus while filtering out a range of other stimuli (Bronkhorst, 2000). *Auditory scene analysis is* the process by which the human auditory system organizes sound into perceptually meaningful elements (Bregman, 1990). This capability allows the hearing system to mentally construct a separate description for each incoming sound source of the surrounding acoustical environment, such as human voices, or music. These auditory phenomena involve complicated mechanisms that go beyond the peripheral aspect of the hearing system, include other senses, and are not yet fully understood by researchers. Therefore, for the purpose of this paper, we adopt a much more simplified view: an acoustic environment is a composition of two main components: the *background environmental noise* and the *aural cues* that convey "useful" information. For example, in a flight simulator, engine noise represents the *background environmental noise*, and cockpit-warning sounds represent *aural cues*; in an urban simulation, the cacophony represents the *background environmental noise* and the warning sounds of an ambulance represents the stimulated *aural cue*.

Perceptual aspects that directly affect the design of IVAs include visual, auditory, olfaction, and tactile perception. The majority of past studies on IVA perceptual models are generally focused on visual perception. Auditory perception has received very little attention despite the fact that it represents one of the most fundamental perceptual aspects of human-like behavior in a virtual environment. Auditory perception improves situational awareness (Kukka et al., 2016) by providing situational feedback and information that are not directly within the field of view. Furthermore, auditory cues are important because they improve the realism of the simulated environment. Imagine if the auditory background of our everyday life were removed—we would feel less “connected” to the world (i.e., less presence). Whereas our eyes are completely blind to the “rear” half of the world, our ears do not present such limitations. When we hear something behind us, not only are we *aware* that the object exists in the environment, but we can often also identify *what it is*. Hence, the simulation of sound is important in a virtual environment because it improves the realism and enhances the sense of presence and immersion (Freeman & Lessiter, 2001).

The goal of our current work is to design and implement an agent-based simulation with the capability to detect critical aural cues in noisy virtual training environments. The auditory perceptual model that we propose seeks to introduce a correlation between IVA hearing perception and human hearing perception. This correlation is especially important for situations such as training a soldier to react quickly to critical or dangerous situations. Effective training would require the IVA to react to sounds similar to the way a human would. For example, an IVA with the capability to detect the sound of gunshots would react: either counterattack or just run to hide. However, an IVA lacking a perceptual model would react (or not react at all) in a non-realistic way.

In this paper, we will briefly review how ears receive and process sound, specifically, the capability to detect aural cues when noise is present. We will introduce an auditory detection capability model that is based on human hearing auditory filters. Then we will present an approach to incorporate this auditory detection model into the IVA simulation. Finally, we will assess the effectiveness of the model and provide recommendations for future improvement and work.

PREVIOUS WORKS

In many agent-based simulation systems, the complete global information of the surrounding environment is provided to the IVAs (Shi, Ren, & Chen, 2009; Bandini, Federici, Manzoni, & Vizzari, 2007); therefore, the perceptual mechanism is not required because the IVAs have full knowledge of the simulation environment. This simulation technique may result in unrealistic IVA actions when compared to the actions that a human would take under similar conditions. As we previously mentioned, virtual agents that have perceptual capability tend to produce results that are more realistic. For this specific reason, several perceptual models have been proposed for agent-based simulation. Most studies performed on IVAs primarily focus on visual perception (Noser, Renault, Thalmann, & Magnenat, 1995; Blumberg, 1997; Terzopoulos & Rabis, 1997). For example, a visual sensor that is based on a ray tracing method is integrated with virtual agents so it can analyze the surrounding environment (Pan, Han, & Law, 2005). By casting a laser ray from the virtual agent’s eye position with a specific azimuth, it is possible for a virtual agent to compute the intersection of the ray and a near object, so the line-of-sight of that object can be determined. In addition to visual, auditory perceptual capability can improve virtual agent realism. Some studies have already attempted to address IVA’s auditory perception. Piza, Ramos, and Zuniga (2005) used a 3D noise propagation model to simulate virtual agent hearing capability. Wang, Kapadia, Huang, Kavan, and Badler (2014) used a sound localization model and integrated it into the behavior for virtual agents to solve the autonomous navigation problem. Lehnert (1996) made a significant contribution to the auditory virtual environmental field by providing a comprehensive review of the fundamentals of auditory virtual environments and discussing the necessity of real-time implementation. Herrero and de Antonio (2003) proposed a vertical layering architecture as a starting point to integrate human hearing perception into IVAs.

To characterize the human hearing system, Tzovaras and Dangelmaier (2014) proposed the use of audiogram parameters to represent virtual agent hearing capability. In this study, the authors made the assumption that an audiogram is a standard and valid way of representing a person’s hearing loss. However, many past studies in hearing perception demonstrated that while the audiogram reflected the level of hearing loss, it does not necessarily correlate with other measurable hearing loss parameters. In fact, Laroche, Hétu, Tran, Josserand, and Glasberg (1992) demonstrated that there is no correlation between hearing sensitivity as measured by the audiogram and loss of frequency selectivity, which is a quantification of the auditory detection capability. Consequently, audiograms cannot be used to predict the capability to detect aural cues in a noisy background.

AUDITORY DETECTION MODEL

The human hearing system is sensitive to sounds with frequencies ranging from approximately 20 Hz to 20 kHz. Absolute sensitivity, typically expressed in decibel (dB) sound pressure level (SPL), represents the signal detection threshold of the hearing system, which is the lowest level of sound pressure required to produce an auditory sensation in a silent environment. Generally, the hearing threshold varies with the frequency of the sound. Furthermore, the ear is more sensitive to the middle of the frequency range than the lower and upper extremes (International Organization for Standardization [ISO] 226:2003). Absolute sensitivity can be determined by a hearing test. The result of a hearing test is displayed in the form of an *audiogram*.

An audiogram is a special graph that represents a listener's hearing capability when there is no background noise. When background noise is present, the situation becomes much more complex. *Frequency selectivity* is one of the most important characteristics of the auditory system because it represents the ability to separate or resolve multiple spectral peaks in a complex sound and it contributes to the detection of acoustic signals in a *noisy background*. *Masking* occurs when the listener's ability to detect or recognize one sound is affected by another sound. The most obvious example of masking is when a low-intensity sound is not audible in the presence of another loud sound, but masking can also occur when two sounds have similar frequencies. The basilar membrane within the cochlea of the inner ear acts as a frequency analyzer. Each part of the basilar membrane corresponds to a certain frequency range. High-frequency sound stimulates the parts of membrane close to the outer ear and low-frequency sound stimulates responses of the membrane at the end of the cochlea. Since different sections of the basilar membrane correspond to different frequencies, those sections actually act as a series of band-pass filters commonly known as *auditory filters*. The frequency selectivity of the auditory system is characterized by measuring the bandwidth of these auditory filters. To characterize the frequency selectivity of the auditory system, the equivalent rectangular bandwidth (ERB) of the auditory filter at different frequencies is measured (Glasberg & Moore, 1990). Mathematical formulas were also proposed to estimate the values of ERB (Glasberg & Moore, 1987).

Sensorineural hearing loss (SHL) occurs when there is damage to the inner ear or from the nerve pathway from the inner ear to the brain. There are several causes of SHL, but the most common is due to either a prolonged exposure to intense sound or the natural aging process. The audible frequency range changes dramatically for SHL listeners. Healthy young children may have a full hearing frequency range up to 20 kHz. From the age of 20, the upper limit drops to 16 kHz and continues to diminish gradually. This is commonly known as *presbycusis*, which is a condition of hearing loss due to the effect of age. Generally, the reduction of the audible frequency range is also accompanied by a reduction of sensitivity at all frequency ranges. The reduction of the audible frequency range and the reduction of the hearing sensibility are not the only negative effects due to hearing loss. In fact, SHL also affects the frequency selectivity of the hearing system and compromises our capability to detect critical aural cues in noisy places (Patterson, Nimmo-Smith, Weber, & Milroy, 1982).

The problem of *detectability* of aural cues in a noisy environment has been the subject of numerous studies (Patterson, 1982; Laroche et al., 1992). The detectability of any sound in a noisy environment is determined by the auditory masked thresholds. Masked threshold represents the minimum level of intensity of an aural cue that can be detected in a background noise and is a direct function of the critical bandwidth of the human auditory filtering system. Predicting auditory detection capability is not new and was extensively studied by several authors. In particular, Zwicker and Scharf (1965) proposed a model of loudness summation to predict masked thresholds for young and normal listeners. The ISO 532 (1989) standard is based on their work. Although powerful, this model is applicable only to listeners aged between 18 and 25 years with normal hearing sensitivity and normal frequency selectivity. Therefore, this model cannot be used to predict the detection capability for a population older than 25 years.

A computerized model was developed to predict auditory detection capability (Laroche, Tran, Hétu, & McDuff, 1991; Tran, Hétu, & Laroche, 1992). It was based on the concept of the excitation pattern that would occur at the inner ear when an aural cue was presented in a noisy background and the estimation of the critical bandwidth of the auditory filter system (Zwicker & Scharf, 1965). Masked thresholds can be predicted using the loudness summation and the excitation pattern model. This computerized model has the capability of taking into account the deterioration of the detection capability due a hearing loss, such as the hearing loss due to aging (presbycusis). It is beyond the scope of this paper to describe in detail the mathematical aspect and the computational algorithm of this loudness summation and the hearing-excitation pattern model. Readers who are interested in more detail on this detection model can refer to the original paper published by Laroche et al. (1991).

Beside the problem of detection, a reaction to an aural cue is warranted only if the aural cue is detectable, attracts attention, and can be recognized among other sound signals. To satisfy this requirement, a level of approximately 10 SPL above the masked threshold has been proposed (Wilkins & Martin, 1987; Patterson & Milroy, 1979). Therefore, in the context of predicting the auditory detection capability of IVAs, it is of paramount importance to accurately determine the masked threshold relative to the factors, which affects the detection ability in quiet and in noise. The excitation pattern model as described in the previous section satisfies such requirement. Finally, the excitation pattern model was successfully employed to adjust the intensity of audio warnings installed in several noisy workplaces (Tran & Hétu, 1996; Hétu & Tran, 1996). For this reason, it represents a promising model that can be used to predict the detection capability of IVAs. In this paper, we will present how this model can be adapted and used in a virtual environment in the context of simulating IVA auditory detection capability.

SIMULATION FRAMEWORK

The proposed hearing perceptual model was integrated into an existing synthetic environment. This synthetic environment is currently used on many United States Air Force C-130J training devices as computer-generated forces (CGF). This CGF has the capability to generate and enable virtually synthetic forces (SF). SFs enabled from this synthetic environment were separated into two main components: physical and behavioral. The *physical* aspect represented the movement and states of the SF, while the *behavioral* aspect determined how it would perform the physical actions. In the current state, because of the lack of perceptual models, the SF always had complete knowledge of environmental data; it used this knowledge to analyze the current situation and perform the required actions. For example, a typical scheme would have an enemy sniper using a rifle to shoot at a virtual soldier, where the enemy sniper could be either a human participating in the same virtual training exercise or another virtual soldier. Because of the lack of a hearing model, this simulation assumes that the virtual soldier will always hear the gunshot and will react based on his/her own current condition and state: shoot back, run to hide, etc. Obviously, if the virtual soldier was simulated with an auditory detection model, the soldier may or may not hear the sound of the gunshot. If the sound of the gunshot was not heard, the soldier would ignore this event and continue to perform his/her current task. The objective of this study is to increase the realism of the simulation: the virtual soldier should hear the gunshots only if his/her auditory detection capability is allowed. The characteristics of the hearing system and acoustic environment should be used to determine this auditory detection capability.

As illustrated in Figure 1, the synthetic environment consisted of an off-line database environment, a user interface module to control the simulation, and the main simulation module:

- The off-line database environment provided the definition and parameters of virtual objects, such as the type of virtual object (e.g., airplane, surface-to-air missile site, soldier) and its dynamic characteristics (e.g., maximum speed, maximum altitude). The life form objects used the *knowledge* database to determine methods for performing any specific required actions. The *aural cue* database provided the sound characteristics of aural cues used by the synthetic environment. Aural cues were stored in waveform format (.WAV). To implement the hearing perceptual model, we added new characteristics pertinent to the auditory detection capability to the *life form* database: age, sex, and audiogram data.
- The simulation control module provided user interfaces and mechanisms to control the training scenario, creating and enabling virtual objects. It also allowed users to modify simulation parameters in real-time (e.g., the velocity or the direction of moving entities).
- The main simulation module performed the workflow of the simulation. The auditory detection model was integrated into the simulation model in a way that the behavior generation process was preconditioned by it; that is, the change in the behavior *shoot back or run to hide* happens only when the *sound of gunshot is detected.* The required action based on the behavior generation process is then performed accordingly.

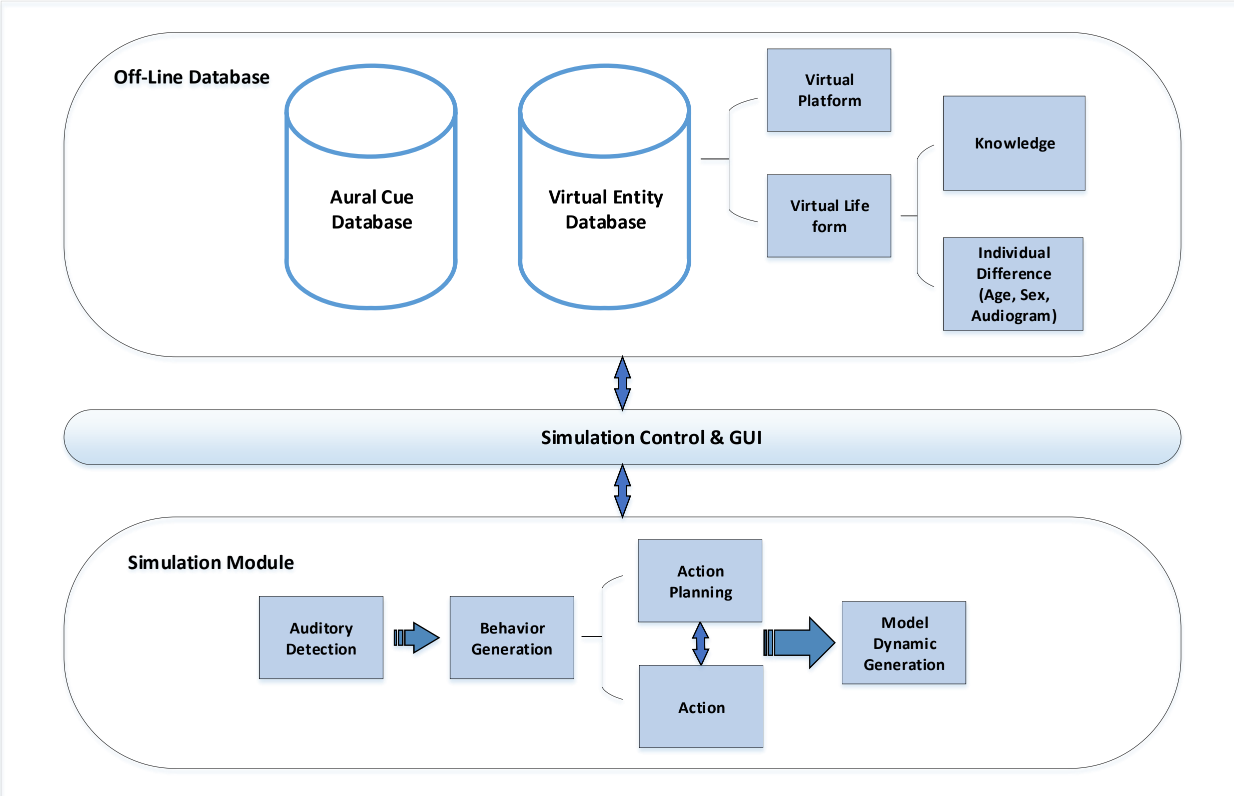

Figure 1. Simulation Framework

IMPLEMENTATION OF THE AUDITORY DETECTION MODEL

To ease the integration, we converted the original auditory detection model as described previously to Visual C++, so it would be compatible with the simulation framework. The auditory detection model consisted of an interface module, a task module, and a processing module. The interface module handled the acquisition of:

1. Environmental data (e.g., characteristics of the aural cue and the background noise).
2. Characteristic of the IVAs (e.g., audiogram, age, and sex).
3. Location of IVAs within the gaming area.

The interface module also interfaced with the existing behavior generation module to plan and perform the required action of the virtual objects. The task module computed the effect of sound propagation. The processing module represented the heart of the auditory detection model. First, it computed the auditory filters bandwidth based on the ages of the IVAs, then using the computed auditory filters bandwidth, the hearing excitation pattern of the aural cues and the background noise were derived. Finally, the loudness summation was calculated to determine if the sounds of the gunshots were detected.

A simplified view of the auditory detection process is illustrated in Figure 2. It uses the following principles and factors to evaluate the auditory detection capability of IVAs:

- Use the intensity and spectral characteristics (converted to third octave bandwidth) of both aural cue and background noise in the prediction of the detection capability.
- Apply a simplified sound propagation model to estimate the reduction of intensity of the sounds due to the distance between the origin of the sounds and the IVAs. This approach is known as the *geometric acoustics*

approach, which provided a reasonably accurate simplification of the effect of sound propagation (Kuttruff, 2000).

- Apply the transformation of "*free field to eardrum"* (Shaw & Vaillancourt, 1985) to both sounds. This process calculates the intensity and spectral characteristics of the sounds at the level of the IVA's inner ear.
- Compute the effect of age on the detection capability by estimating the ERBs of the hearing auditory filters. Valid age for the present simulation framework varies from 18 to 65 years.
- Calculate the excitation pattern of the background noise by using its spectral content as well as the estimation of the ERBs. The excitation pattern of the background noise defines the masked thresholds.
- Use the same approach to calculate and compare the excitation pattern of the aural cue to the one produced by the background noise calculated previously.
- Declare the aural cue as "detected" if its excitation pattern is greater than the one of the background noise for all frequency spectrum of the aural cue.
- Ensure that the aural cue is not only detected, but also recognized and attracted attention. The intensity of the aural cue is required to be also at least 10 dB SPL greater than the background noise threshold.

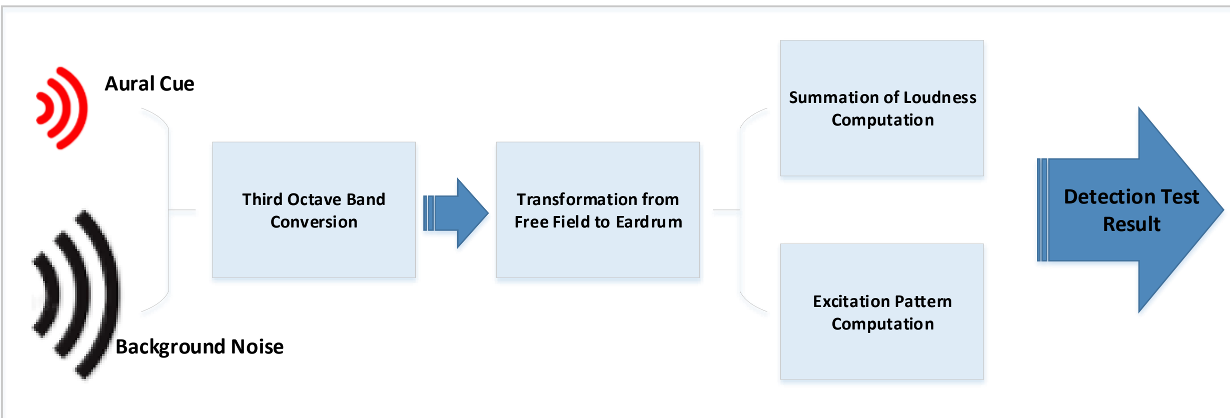

Figure 2. Diagram of the Auditory Detection Processing Capability Model

SIMULATION SCENARIO

The simulation scenario iterated in real-time and consisted of two virtual soldiers. Each soldier's task was simply to use their rifle to fire back whenever they heard a gunshot fired from the other soldier. The gaming area size was set to 1 square mile and the soldiers were free to move randomly within it during the simulation. For the purpose of this simulation, the battle damage model of the synthetic environment was disabled, so the soldiers could not be wounded or killed. Figure 3 presents a screenshot of the simulation environment.

During the simulation, we observed and recorded the following simulation parameters for each gunshot event:

- The *distance* between the soldiers. This parameter reflects the condition that the sound intensity of the gunshots decreases as the distance between the soldiers increases.
- The *intensity* and *frequency spectrum* of the background noise. These parameters will be used to assess and determine the hearing masked thresholds.
- The *intensity* and *frequency spectrum* of the gunshot sounds. These parameters will be used to assess the auditory detection capability in the background noise.
- The *age* of the soldiers (20, 30, and 40 years old). This parameter will be used to simulate a hearing loss condition.
- Computation latency will be used to assess the computation performance.

To assess the performance of the simulation, we computed the following results:

- Detection thresholds as a function of the distance between the soldiers.
- Detection thresholds as a function of the soldiers' ages.
- Detection assessment. This result can be either "detected" or "not detected."

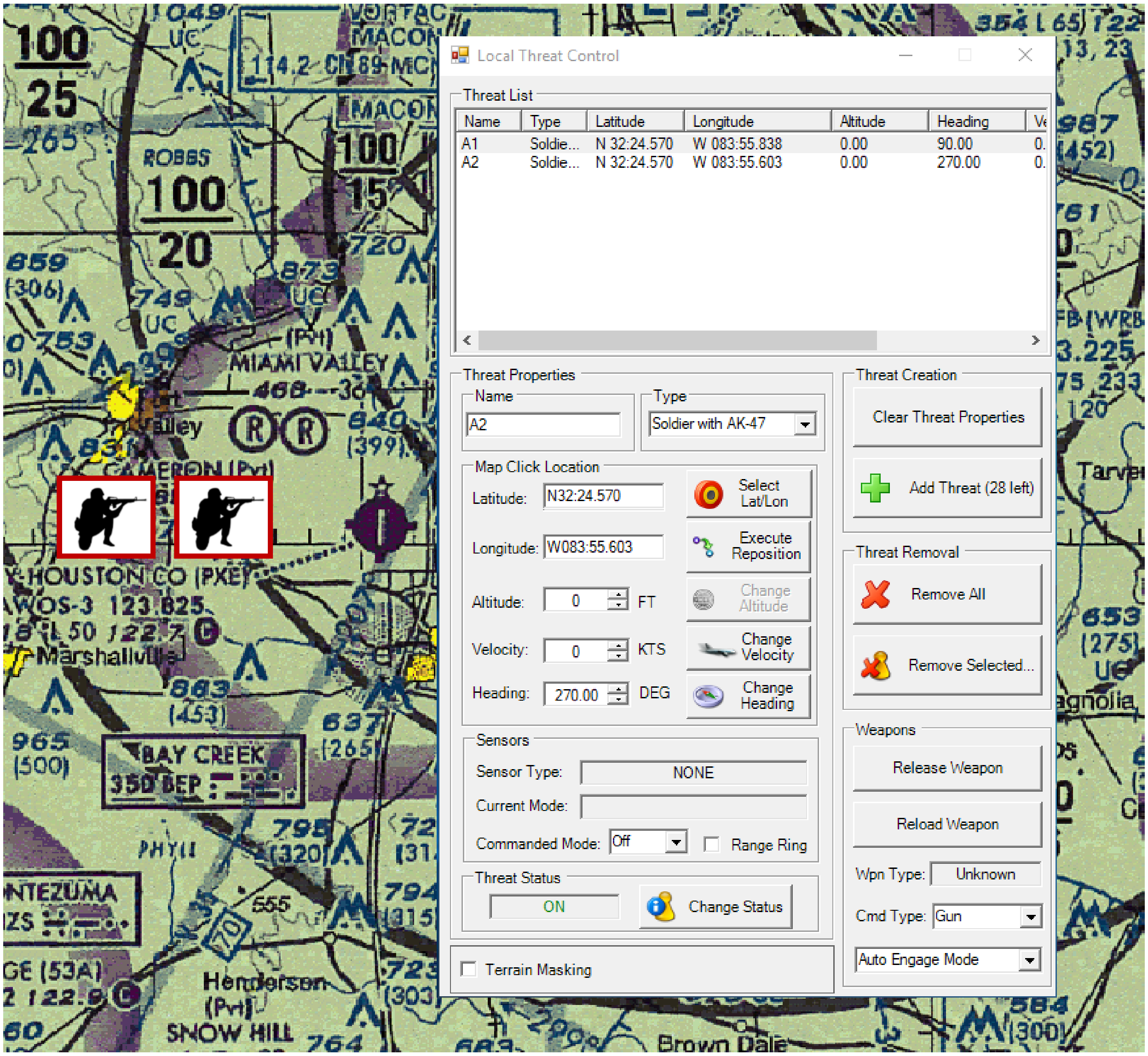

Figure 3. Simulation Control Screen

ACOUSTIC ENVIRONMENT

For the purpose of this simulation, we used the soundwave of an infantry automatic rifle stored in the off-line aural cue database. During the simulation, we assumed that gunshot sounds always originate from the soldiers' locations. The reduction of intensity of the gunshot sounds due to the sound propagation from the shooter soldier to the other soldier was computed using the geometry sound propagation method. To emulate environmental noise, we use a broadband noise similar to a pink noise. The frequency spectrum of this noise is from 100 Hz to 10 kHz. This broadband noise has a very specific characteristic: the power spectral density is inversely proportional to the frequency of the signal. Therefore, this broadband noise will have approximately an equal quantity of energy per octave bandwidth. The main reason that we selected this type of noise for the simulation is to ensure that its effect of masking is approximately equal for a broad area of the audible frequency. To simplify the simulation, we fixed the origin of the environmental noise at the center of the gaming area. Consequently, the reduction of the environmental noise intensity due to the effect of propagation was computed from the center of the gaming area to the location of the soldiers in real-time.

RESULTS AND DISCUSSION

The simulation scenario was executed without interruption for 5 minutes. The results were continuously recorded in a file for post-analysis. The obtained simulation results are summarized as follow:

1. No *computation latency* was observed during the entire simulation. The detection results were computed and available within the same simulation frame and used by the behavior generation process.
2. The detectability of the gunshots decreased as the *distance* between the soldiers increased. This was expected and accounted by the effect of the sound propagation.
3. The detectability of the gunshots increased when the soldier distanced from the center of the gaming area. This result can be explained by the fact that the *intensity* of the background noise decreased when the soldiers moved father from the center of the gaming area. In fact, if the intensity of the background noise decreased, its masking effect is also decreased.
4. The detectability of the gunshots decreased when the *age* of the soldier increased. This result demonstrated the effect of hearing loss due to the effect of age on the capability of detection of aural cues in a noisy background.

During the simulation, the predicted detection thresholds per octave bandwidth were computed. These detection thresholds represent the minimum sound intensity of the gunshots required to generate a "detected" condition. The results were compared to those obtained by the study performed by Hétu and Tran (1994). The authors of this study clinically measured the masked thresholds of 52 listeners between 25 and 65 years old. The results obtained by our simulation generally indicated a slight overestimation of the masked thresholds, with an average error prediction of less than 2 dB. This observation can be explained by the fact that the estimation of auditory filter ERB, as used by our model, is generally greater when compared to those obtained clinically with human subjects. Furthermore, the findings also indicate that errors tended to increase with signal frequency and reach a maximum of approximately 4 to 6 kHz. Laroche et al. (1992) also reported a similar observation in their study.

The model, as presented in this paper, was used to predict the detection capability of IVAs of any age from 18 to 65 years. Nonetheless, the model has some limitations. First, the effects of other types of hearing loss, such as noise-induced hearing loss, was not taken into account. Another restriction comes from the fact that this model ignores the temporal feature of the aural cues and background noise. For this specific reason, the model will only provide valid predictions of detectability when the level of background noise is mostly steady. In order to take into account the temporal characteristics of the background noise, auditory filter time constants can be introduced to the calculation. However, the introduction of these time constants into the model will require an additional signal sampling process and, consequently, will essentially compromise the real-time performance aspect of the model. Finally, the sound propagation computation used by this model is valid only in free fields where the effect of distance can be accounted for easily. If the simulated virtual environment represents a confined space where the effect of the reverberation is significant, the prediction of this model will be limited.

Today, the technology for creating virtual worlds is very advanced and the visual realism of simulated objects continues to improve. Conversely, the development of IVAs for training is still in its infancy. The technology used to create IVAs is far from mature and has not reached the stage where it can be easily integrated into a training system. Part of the difficulty is related to the diversity of IVAs (i.e. different types of IVAs for different types of applications) as explained by Oijen (1982). Another resides in the absence of a standardized method to integrate perceptual models into virtual worlds, like the one that was proposed in this paper. Beside visual and auditory perception, directions for future research include not only the development of additional perceptual models, such as olfaction or tactile perception, but also the criticality of establishing an integration standard to implement these perceptual models into modeling IVAs.

CONCLUSION

While human auditory detection capability in noisy background environments has been the subject of extensive studies, the results were rarely used in the context of IVA modeling. This paper presents a human auditory detection model that can be used to predict the detectability of aural cues for IVAs in a virtual environment. This model has been integrated into a CGF environment designed for military mission training with the ultimate goal of improving the realism of the IVA simulation. The uniqueness of the model presented in this paper resides in the combination of the loudness summation and auditory excitation pattern. This approach takes into account the effect of natural hearing loss due to aging. Using this model, the age of virtual agents can be specified so that the hearing loss (due to the effect of age) will be taken into account when predicting the auditory detection capability. This enhances the modeling of IVAs, enabling a computer-generated combatant to have the ability to acquire additional surrounding environmental data through an auditory perceptual model and, consequently, increase the likelihood of surviving an attack.

Additional work is currently underway to improve this hearing perceptual model by introducing the capability to localize aural cues. Therefore, IVAs will not only be capable of detecting aural cues but they will also be able to determine the *direction of arrival* and the *distance from the origin* of the sound cues.

Although the work of integrating this perceptual model into IVA modeling is still at the proof-of-concept and prototyping phase, it represents a highly promising implementation framework that can leverage to integrate additional human perceptual models. However, it will be essential to define a standard integration interface to facilitate this insertion.

REFERENCES

Bandini, S., Federici, M., Manzoni, S., & Vizzari, G. (2007). Pedestrian and crowd dynamics simulation: Testing SCA on paradigmatic cases of emerging coordination in negative interaction conditions. *Parallel Computing Technologies*, pages 360-369.

Blumberg, B. (1997). Go with the Flow: Synthetic Vision for Autonomous Animated Creatures. Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, CA.

Bregman, A. S. (1990) Auditory scene analysis. MIT Press: Cambridge, MA.

Bronkhorst, A. W. (2000). The Cocktail Party Phenomenon: A Review on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica united with Acustica*, 86: 117–128.

Carlile, S (1996). The Physical and Psychological basis of sound localization. Virtual Auditory Space: Generation and Application. Edited by Simon Carlile. Chapter 2.

Freeman, J., & Lessiter, J. (2001). Here, There and Everywhere: The effects of multichannel audio on presence. *Proceedings of the 2001 International Conference on Auditory Display, Espoo,* Finland, 231-234.

Glasberg, B. R., & Moore, B. C. J. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculation excitation patterns, *Hear. Res.* (28), 209–225.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data, *Hear. Res.* (47), 103–138.

Herrero, P., & de Antonio, A. (2003, July). Introducing human-like hearing perception in intelligent virtual agents. In Proceedings of the second international joint conference on Autonomous agents and multiagent systems (pp. 733-740). ACM.

Hétu, R., & Tran, H. (1996). Psychoacoustic Performance in Workers with NHIL. *Scientific Basic of Noise-Induced Hearing Loss*. Thieme Medical Publishers, Inc. Chapter 22. 264-285.

International Organization for Standardization. (1989). *Acoustics – Method for calculating Loudness level* (ISO Standard No. 532).

International Organization for Standardization. (2003). *Acoustics – Normal equal loudness level contour* (ISO Standard No. 226).

Kukka, H., Goncalves, J., Wang, K., Puolamaa, T., Louis, J., Mazouzi, M., & Barco, L. R. (2016). Utilizing Audio Cues to Raise Awareness and Entice Interaction on Public Displays. *Proceedings of the 2016 ACM Conference On designing Interactive Systems*. Brisban, QLD, Australia. 807-811.

Kuttruff, H. (2000). *Room Acoustics.* Spon Press.

Laroche, C., Hétu, R., Tran, H., Josserand, B., & Glasberg, B. (1992). Frequency Selectivity in workers with noise-inducted hearing loss. *Hearing Research*, 64, 61-72.

Laroche, C., Tran, H., Hétu, R., & MacDuff, S. (1991). "Detectsound": A computerized model for predicting the Detectabbility of Warning Signals in Noisy Workplaces. *Applied Acoustics*, *32*, *193-214*.

Lehnert, H. (1996). Fundamental of auditory virtual environment. *Artificial Life and Virtual Reality*, N. M. Thalmann, D. Thalmann (edts), John Wiley and Sons Ltd.

Noser, H., Renault, O., Thalmann, D., & Magnenat N. (1995). Navigation for Digital Actors based on Synthetic Vision, Memory and Learning. *Computers and Graphics.* Pergamon Press, Vol.19, No 1.

Oijen, J. V. (1982): Cognitive Agents in Virtual Worlds: A middleware Design Approach. *Netherlands Organization for Scientific Research*. SIKS Dissertation Series No. 2014-35.

Pan, X., Han, C. S., & Law, K. H. A Multi-Agent Based Simulation Framework for the Study of Human and Social Behavior in Egress Analysis. In: Soibelman, L., PenaMora, F. (ed.) *ASCE International Conference 2005,* Cancun, Mexico, July 12-15 (2005).

Patterson, R.D. (1982). Guidelines for auditory warning systems on civil aircraft. CAA Paper 82017. Civil Aviation Authority, London.

Patterson, R.D., & Milroy, R. (1979). Existing and recommended levels for auditory warnings on civil aircraft. Contract report 7D/S/0142, MRC Applied Psychology Unit, Cambridge, UK.

Patterson, R.D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.* 72(6), 1788-1802.

Piza, P., Ramos, F., & Zuniga. F. (2005). Virtual sensors for dynamic virtual environments. In Computational Advances in Multi-Sensor Adaptive Processing, *1st IEEE Int. Workshop*, pages 177-180.

Reynolds, C. W. Flocks, herds, schools: a distributed behavioral model (1987). *Computer Graphi*cs, 21(4), 25–3.

Shaw, E.G.A., & Vaillancourt, M.M. (1985). Transformation of sound-pressure level from free field to the eardrum presented in numerical form. *J. Acoust. Soc. Am.* 78(3), 1120-1123.

Shi, J., Ren, A., & Chen, C. (2009). Agent-based evacuation model of large public buildings under fire conditions. *Automation in Construction*, 18(3):338-347.

Terzopoulos D., & Rabie, T.F. (1997). Animal Vision: Active Vision in Artificial Animals. *Journal of Computer Vision Research*, 1(1):2-19.

Tran, H., & Hétu, R. (1996). La planification de la signalization acosutique en milieu industriel: Critère de conception des avertisseurs sonores de danger. *Canadian Acoustics*, *24(2)*, *3-17*.

Tran, H., Hétu, R., & Laroche, C. (1992). Computerized assessment and prediction of the audibility of sounds warning signals for normal and hearing-impaired individuals. *Mattila, M., Karwonski, W. (eds)*. Computer Application in Ergonomics, Occupational Safety and Health. Elsevier Science Publishers. 105-112.

Tzovaras, D, & Dangelmaier, M. (2014). Virtual and Augmented environments and realistic user interaction to achieve embedded accessibility designs. VERITAS D4.1.13. Grant agreement #247765.

Wang, Y., Kapadia, M., Huang, P., Kavan, L., & Badler, N. I. (2014). Sound Localization and Multi-Modal Steering for Autonomous Virtual Agents. *Symposium on Interactive 3D Graphics and Games* (I3D '14), 23-30.

Wijermans, N., Jorna, R., Jager, W., Vliet, T., & Adang, O. (2013). CROSS: Modelling Crowd Behavior with Social-Cognitive Analysis. *Journal of Artificial Societies and Social Simulation*. 16(4).

Wilkins, P., & Martin, A.M. (1987). Hearing protection and warning sounds in industry – a review. *Applied Acoustics*, 21, 267-293.

Zwicker, E., & Scharf, B. (1965). A model for loudness summation. *Psychological Review*, 72, 3-26.