Human-Like Auditory Capability for Intelligent Virtual Agents

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

ABSTRACT

Intelligent Virtual Agents (IVAs) are computer-generated entities that are visually similar to humans and exhibit human-like behavior. IVAs are important components in simulated real-world environments. In training, IVAs are used mainly for task collaboration where virtual agents interact with each other or with human users. Additionally, IVAs should be able to 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 to model IVA.

Besides visual, olfaction, and tactile perception, auditory perception represents one of the perceptual aspects that directly affects the design of IVA. To model the IVA hearing capability, an auditory perceptual model was developed that can be used to predict the capability to detect sound cues in noisy environments ("*Human-Like Auditory Detection Capability for Intelligent Virtual Agents,*" I/ITSEC 2018. In addition to the capability to detect sound cues, the human hearing system can also identify the direction from which the sound is coming, estimate the distance of the sound source, and eventually assess the characteristics of the physical surrounding environment affecting sound propagation.

This paper is a continuation of the author's I/ITSEC paper from 2018 (Tran, 2018). To enhance the IVA auditory capability, a perceptual model will be introduced to simulate the capability to localize sound sources. The paper will provide the foundation of the sound localization model, which is based on the Duplex theory of the human hearing system - Interaural Time Difference (ITD) and Interaural Level Difference (ILD). Then, it will explain how this model was integrated to the IVA model to simulate the sound localization capability. Finally, the paper will present the simulation results and assess the effectiveness of this auditory perceptual model when used with the simulation of IVAs. The results of this study indicated that the Duplex Theory is suitable to simulate the capability to localize sound sources, especially when the Signal-to-Noise (S/N) ratio is favorable to the sound localization task (e.g., S/N is equal or greater than 12 dB SPL).

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 was also the lead designer of the Computer-Generated Forces currently used by several 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 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 visually similar to humans and 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 are virtual warfare scenarios (VWS). In a VWS application, IVAs can represent either virtual team members or enemy combatants.

IVA modelling is a new emerging, multidisciplinary field and represents agent-based computing that belongs to a 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 simulation 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 a 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.

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 (Freeman & Lessiter, 2001). 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 this study is to design and implement an agent-based simulation with the ability to detect and localize critical aural cues in noisy virtual training environments. The proposed auditory perceptual model 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 and localize the sound of gunshots would react with either a counter-attack or would run to hide. However, an IVA lacking a perceptual model would react in a non-realistic way or not react at all.

This paper briefly reviews how ears receive and process sound, specifically, the capability to detect and localize aural cues when noise is present. The author introduces an auditory perceptual model based on human hearing auditory filters and an approach to incorporate this auditory perceptual model into the IVA simulation. Finally, an assessment of the effectiveness of the model is presented and recommendations provided for future improvement and work.

AUDITORY CAPABILITY

Human ears convert sound pressure into neural signals, which, in turn, lead to a perceptual experience. The ear is divided into three main parts: outer, middle, and inner. The pinna, part of the outer ear, collects sound energy and guides it into the ear canal. After traveling down the ear canal, sound waves cause the eardrum to vibrate. The eardrum, a cone-shaped membrane that is part of the middle ear, converts the vibrating air into vibrating liquid of the inner ear. The inner ear contains vestibular organs (which provide sense of balance and awareness of spatial orientation) and the cochlea, which are the sense organs for hearing. The two most critical capabilities of our hearing system are the capability to detect sound cues and the capability to localize the sound sources. Both capabilities represent critical elements for human safety and awareness, which is the primary construct for decision-making ability, perception, and effective task performance (Kukka, et. al., 2016).

AUDITORY DETECTION CAPABILITY

Last year's paper, "*Human-Like Auditory Detection Capability for Intelligent Virtual Agents*" (I/ITSEC 2018), presented an auditory perceptual model that can be used to predict the ability to detect sound cues in noisy environments. The model is based on human hearing auditory filters and was used to predict the detection capability of IVAs of any age from 18 to 65 years. To assess the performance of the model, the proposed hearing detection model was integrated into an existing synthetic environment, and the model was used to perform several simulations.

The simulation results obtained from this study are summarized as follows:

1. The simulation results indicated generally a slight overestimation of the masked thresholds, with an average error prediction of less than 2 dB SPL.
2. The auditory detection model predicted correctly that the detection capability is a function of the distance between the origin of the aural cue and the IVA.
3. The auditory detection model predicted correctly the IVA detection capability when the intensity of the aural cue is at least 10 dB SPL greater than the background noise threshold.
4. The predicted masked thresholds increased when the level of environmental noise increased.
5. The predicted masked thresholds increased when the simulated IVA age increased.
6. 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.

Finally, the study concluded that the detection model is suitable to model the IVA capability to detect sound cues in a noisy virtual environment.

AUDITORY LOCALIZATION CAPABILITY

Auditory localization is the element of auditory perception that is the most critical to human effectiveness and safety; especially when visual cues are obscured or completely lacking. This remarkable ability of the human auditory system allows us to detect and analyze sound cues in a complex acoustical environment where surrounding noise is present. In order to localize a sound, the auditory system relies on binaural and monaural acoustical cues. Furthermore, the binaural auditory system plays a main role in improving speech intelligibility in noisy and reverberant environments (Bodden, 1992). Interaural arrival-time and level differences are the main cues evaluated by the binaural auditory system. The "duplex theory" was first proposed by Lord Rayleigh around 1907. Our two ears are separated by a relatively large head, thus for a sound that originated from the horizontal plane, the onset time (time to reach the ear) for that sound is different for each ear. This is referred to as the interaural time difference (ITD). However, these localization cues are effective only for sound with frequencies lower than approximately 1.5 kHz. Another mechanism involved in the sound localization is the interaural level difference (ILD), which represents the difference of sound intensity that reaches each ear. The combination of ITD and ILD represents the binaural cues that the auditory system uses to localize sound originated from the horizontal plane. Figure 1 illustrates the basic mechanism of the duplex theory. From a sound situated at an azimuth of θ, the sound intensity that reached the left ear (ipsilateral ear) is slightly louder than the sound intensity that reached the right ear (contralateral ear). The cause of the different sound intensity between the ears is the shadowing effect of the head, which prevents some of the sound energy from directly reaching the contralateral ear. Similarly, in a condition of free field, where the effect of reverberation is not taken into account, the time for the sound to reach the left ear is shorter than the time to reach the right ear. The mathematical formulas to calculate ITD and ILD are described in detail in Blauert's work (Blauert, 1997). Without getting into the detail of the mathematical calculation of ITD and ILD, the effect of ITD and ILD on the capability to localize sound source is not equal. In general, the effect of ITD is more pronounced when the frequency of the sound is below 1.5 kHz, while ILD is more important for localization when the frequency of the sound is above 1.5 kHz (Comalli & Altshuler, 1976).

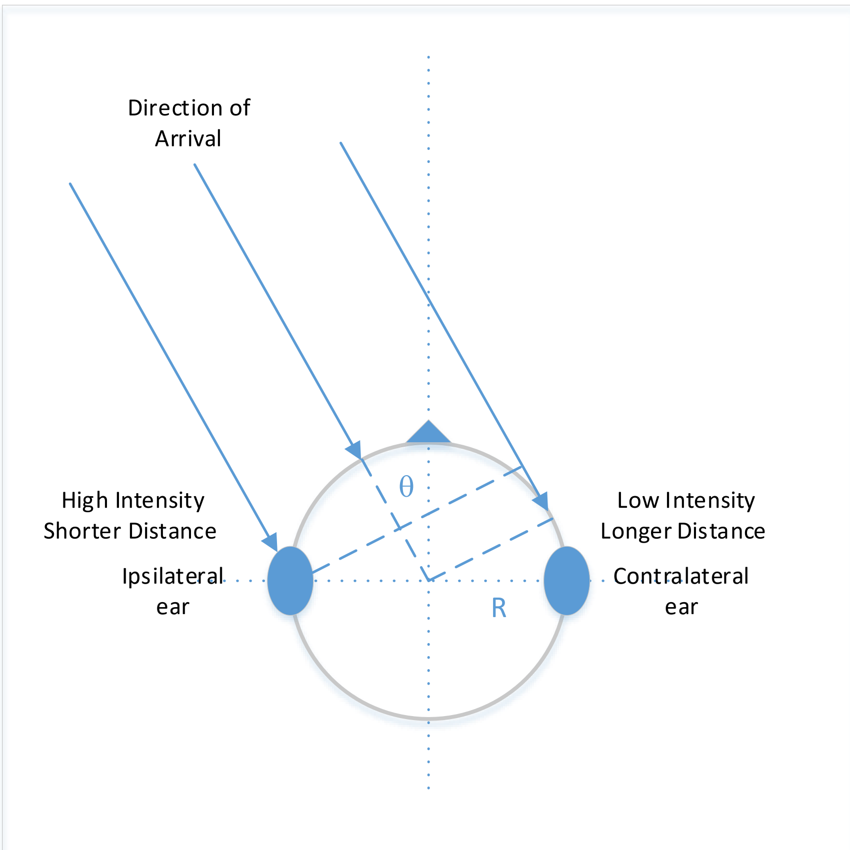

Figure 1. Binaural Localization Cues. Sound Intensity and Time of Arrival at The Near and Far Ear.

While ITD and ILD represent the fundamental mechanism of the duplex theory of sound localization, our capability to localize sound source depends also on the type of sound source and the type of acoustical environment. Furthermore, the movement of the sound source and the presence of other sounds in the environment also affect our sound localization ability.

PREVIOUS WORKS

Human auditory sound localization capability has been the subject of extensive studies. The results of these studies, however, were rarely used in the context of IVA modeling. A combat simulation model was proposed by Michaud (2005). The author of this study analyzed raw data obtained from a sound localization experiment the U.S. Air Force Research Laboratory conducted to determine how to best replicate the human's ability to use the sense of hearing to ascertain direction of sound sources. The effects of the urban environment on the ability of soldiers to localize sound was studied (Scharine, Letowski, 2005). The authors identified and provided a list of factors that affect the soldier's ability to identify the direction of sound sources in an urban battlefield environment. Egger (2013) proposed a model to simulate the localization capability in the human median plane. This model introduced the usage of dual resonance nonlinear filter bank (DRNL) to simulate the behavior of filter width and compression of the auditory filters. While these studies provided interesting data and a concept that can be used to model the capability of sound localization, they did not define or provide an integration interface that will be easily integrated within our IVA simulation model.

An extensive study (Gagné, Tran, Denis, & Leblanc, 1998) investigated the performance of localization with 21 normal hearing listeners and 21 Sensorineural Hearing Loss listeners. This study used four types of sound: pure tones with constant pressure level, pulsed sounds, amplitude modulated sounds, and frequency modulated sounds. The results of this study were published in a comprehensive report by the Institute de Recherche en Sante et Securite du Travail, Canada. The localization capability for the normal hearing listeners indicated:

1- For the horizontal plane, errors of localization are intra-quadrant (i.e., less than 90 degrees). The rate of localization success is approximately 70%. Sounds with frequency of modulation are the best condition for localization, with a greater than 90% rate of success.
2- With a sound test that contains only a single frequency component (e.g., pure tone), the rate of success is better when the test is performed without background noise.
3- With a sound test that contains multiple frequency components (e.g., frequency modulated sound), the rate of success is excellent for both testing conditions, with or without background noise. Furthermore, a signal-to-noise ratio of +12 dB SPL is enough to ensure very good performance of localization in the horizontal plane.

Generally, for all test sounds and for both testing conditions, with and without background noise, the performance of localization decreased for the group of hearing loss listeners when compared to the groups of normal listeners. The observation of this study supports the idea that the hearing system localizes sound based on ITD and ILD indices and has the ability to analyze the frequency content of an incoming sound. The degradation of this ability to localize sound limits the amount of spatial information that hearing-impaired listeners can obtain from the real world. The results obtained by Gagne, et. al., are comparable to those obtained from the study performed by Parhizkaki (2008) in which the author also quantified the localization error by a group of normal and hearing-impaired listeners. The effect of age on the performance of localization is supported by another extensive study by Abel, Giguère, Consoli, and Papsin (2000). Seven groups of 16 listeners, aged 10 to 81 years, participated in this study. The results indicated a decrease in performance as early as the third decade of age. The decrease was largest for low-frequency sounds and smallest for broadband noise.

To simulate the localization of IVA, this paper relies primarily on the data obtained from the study performed by Gagne, et. al. (1998). The authors of this study normalized the localization data obtained to the ITD and ILD for each test condition. By using these normalized results, the distribution of localization error for the horizontal plane can be deducted. Additionally, the effect of age and the effect of noise masking on the sound localization performance can also be estimated with the data for both normal and hearing loss listeners. It is beyond the scope of this paper to describe in detail the mathematical aspect and the computational algorithm of this localization model. Readers interested in more detail on this localization model can refer to the original work of Gagne, et. al. (1998).

SIMULATION FRAMEWORK

The hearing perceptual model, which includes the auditory detection and localization capability, was integrated into an existing synthetic environment. This synthetic environment can generate and enable virtual 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. This simulation framework was described in detail in Tran (2018).

As illustrated in Figure 2, the simulation framework consisted of an offline database environment, a user interface module to control the simulation, and the main simulation module. They are described as follows:

- 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, new characteristics were added 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 perception model was integrated into the simulation model as a pre-processing component of the behavior generation process; consequently, the change in the IVA's behavior is commanded by the computation results of the auditory perception model. The required action based on the behavior generation process is then performed accordingly.

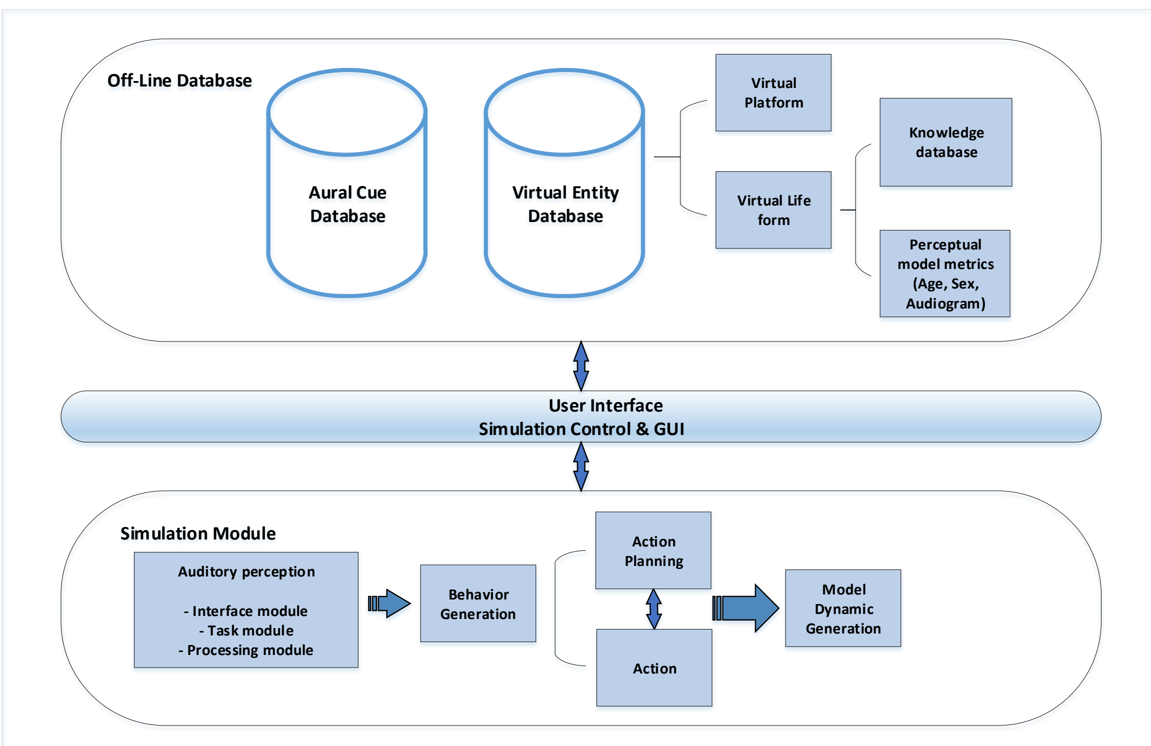

Figure 2. Simulation Framework

The auditory perception model consisted of an interface module, a task module, and a processing module. The interface module handled the acquisition of the following parameters:

1. Environmental data (e.g., characteristics of the aural cue and the background noise)
2. Perceptual model metrics 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 main component of the auditory perception 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 background noise were derived. Subsequently, the loudness summation was calculated to determine if the sounds of the aural cues were detected. Finally, the apparent location of the sound is determined.

Figure 3 illustrates a simplified view of the auditory detection process. It simulates the human capability to detect and to localize sound source as follows:

- 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 must also be at least 10 dB SPL greater than the background noise threshold.
- If the sound passed the detection test, then the apparent location of the sound is computed by applying the localization error distribution.

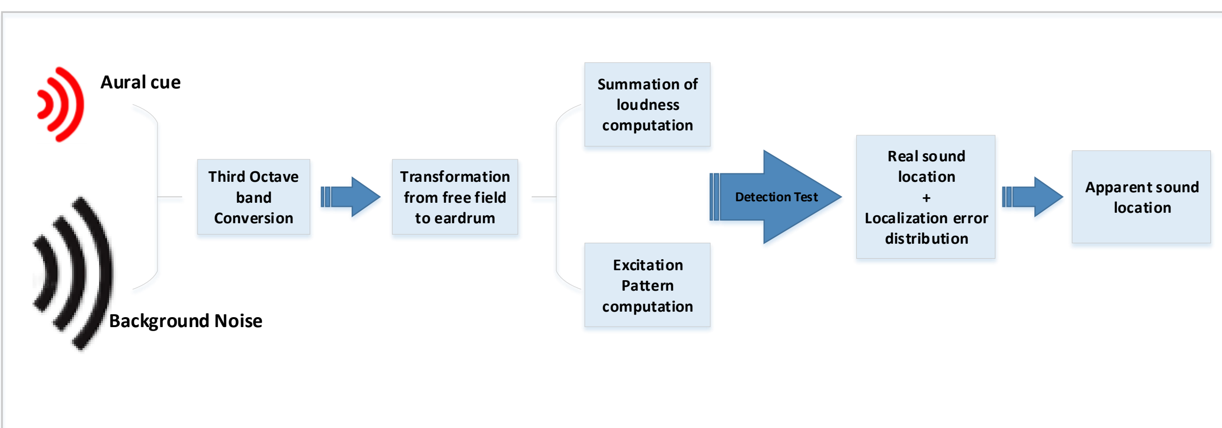

Figure 3. Processing Diagram of the Hearing Perceptual Model

SIMULATION SCENARIO

Human auditory perception of the surrounding environment is a 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, including other senses, and are not yet fully understood by researchers. Therefore, a much more simplified view is used for this paper: an acoustic environment is a composition of two main components: the <u>background environmental noise</u> and the <u>aural cues</u> that convey “useful” information. For example, in an urban simulation, the cacophony represents the background environmental noise and the sound from gunshot represents the stimulated aural cue.

The simulation scenario was iterated in real-time and consisted of a virtual soldier. The gaming area size was set to 1 square mile. The soldier was initially locating at the center of the gaming and heading north. Figure 4 presents a screenshot of the simulation environment. The simulation scenario contained two virtual soldiers, however for the purpose of this study, only one soldier was enabled. During the simulation, a sound of a gunshot was randomly generated within the gaming area. If the soldier detected the sound of the gunshot, the hearing perceptual model computed and generated the apparent location of the gunshot. Using the computed location, the soldier acquired a new heading and moved to the direction of the apparent location of the gunshot. If the soldier did not detect the gunshot, the soldier continued to move with the current direction.

This simulation used the soundwave of an infantry automatic rifle stored in the off-line aural cue database. To emulate environmental noise, a broadband noise similar to a pink noise was used. 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 has approximately an equal quantity of energy per octave bandwidth. The main reason this type of noise was selected for the simulation was to ensure that its effect of masking was approximately equal for a broad area of the audible frequency. To simplify the simulation, the origin of the environmental noise was fixed 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.

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

- Location of the soldier and the gunshot
- Intensity and frequency spectrum of background noise (parameters used to assess and determine hearing masked thresholds)
- Intensity and frequency spectrum of gunshot sounds (parameters used to assess auditory detection capability in the background noise)
- Location of the sound source as perceived by the soldier
- Age of the soldiers (20, 30, and 40 years); (parameter used to simulate a hearing loss condition)
- Computation latency (parameter used to assess computational performance)

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

- Detection thresholds as a function of the distance between the soldier and the gunshot
- Detection thresholds as a function of the soldiers’ ages
- Detection assessment; either “detected” or “not detected”
- Error of localization by the virtual soldier

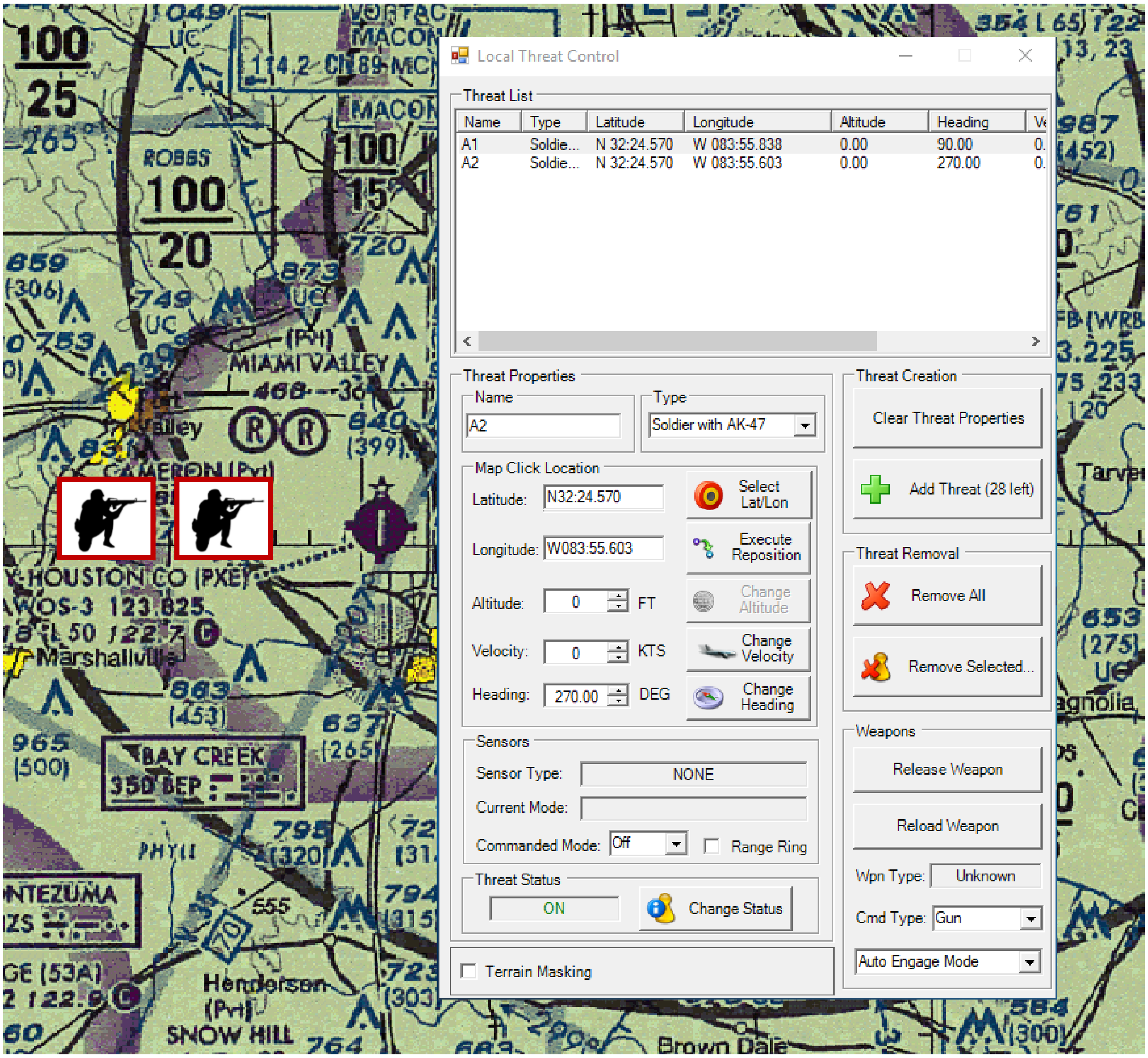

Figure 4. Processing Diagram of the Hearing Perceptual Model

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 simulation results are summarized as follow:

1. No computation latency was observed during the entire simulation. The detection and localization results were computed and available within the same simulation frame and subsequently used by the behavior generation process.
2. The ability to localize gunshots increased when the soldier moved 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 soldier moved father from the center of the gaming area. In fact, if the intensity of the background noise decreased, its effect on the soldier's ability to localize sound gunshots also decreased. Good & Gilkey (1996) reported similar results in their study, in which they quantified the ability to localize sound cues in the presence of a broadband noise at nine signal-to-noise ratios varied over a 23 dB range.

3. The ability to localize 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 ability of detection of aural cues in a noisy background. This is not a surprising result, as the effect of age on the capability of localization is already widely reported. In fact, Rakerd, et. al. (1998) concluded in their study that listeners with clinically relevant levels of presbyacusis (hearing loss due to the effects of age) were greatly compromised in their ability to make spectrally cued localization decisions.

The model, as presented in this paper, is used to simulate the hearing capability of IVAs. It predicts the detection capability in noisy environment of IVAs of any age from 18 to 65 years. Additionally, the model also simulated the capability for IVAs to determine the direction of arrival of sound cues. For the detection capability, 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 provides valid predictions of detectability only when the level of background noise is mostly steady. In order to taking 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. For the capability to localize sound sources, the model is valid only for static and single sound source. When the sound source is not located on the horizontal plane, the application of this model is limited, especially when the sound is located on the cone of confusion. 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 is limited.

CONCLUSION

This paper presents a human auditory perceptual model that can be used to model the behavior of IVAs in a virtual environment. This paper improves the previous work of modeling the capability to detect sound cues in noisy environments, by adding the capability to localize the sound source. The resulting perceptual model simulates the IVA capability to perceive the surrounding acoustical environment. The model was integrated into a Computer-Generated Forces 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 a sound detection and localization model. Additionally, the simulation model also 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 simulating the hearing capability of IVAs. 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.

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. Besides the auditory perception, directions for future research include the development of additional perceptual models, such as visual, olfaction, and tactile perception.

REFERENCES

Abel, S. M., Giguère, C., Consoli, A., & Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *The Journal of the Acoustical Society of America*,*108*(2), 743-752.

Blauert, J. (1997). Spatial hearing: The psychophysics of human sound localization (Revised ed.). London: MIT Press.

Bodden. M. (1992). Modelling human sound-source localization and the cocktail party effect. *Acta Acustica*, 1, 43–55.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bronkhorst, A. W. (2000). The Cocktail Party Phenomenon: A Review of Research 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.

Comalli, P.E.; Altshuler, M.W. (1976). Effects of Stimulus Intensity, Frequency, and Unilateral Hearing Loss on Sound Localization. *Journal of Auditory Research*, 16, 275–279.

Comalli, P.E., Altshuler, M.W. (1976). Effects of Stimulus Intensity, Frequency, and Unilateral Hearing Loss on Sound Localization. *Journal of Auditory Research*, 16, 275–279.

Egger, K. (2013). Implementation and evaluation of auditory models for sound localization. Student Project. Austrian Academy of Science.

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.

Gagné, J.P., Tran, H., Denis, S., Leblanc, M. (1998). Conception de signaux sonores et sur la mesure inductive de la capacité de localization auditive des avertisseurs sonores de danger en milieu industriels. *Institute de Recherche sur la Santé et Securité au Travail*.

Good, M.D., Gilkey, R. H. (1996). Sound localization in noise: The effect of signal-to-noise ratio. . *J. Acoust. Soc. Am.* 99(2), 1108-1117.

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.

Michaud, J.C. (2005). Auditory detection and localization for computer-generated for individual combatants. Naval PostGraduate School Thesis.

Parhizkari, O. (2008). Binaural Hearing – Human ability of sound source localization. Blekinge Institute of Technology.

Rayleigh, Lord (1907). On our perception of sound direction. *Philosophical Magazine,* 13, 214-232.

Scharine, A.A, Letowski, T.R. (2005). Factors affecting Auditory Localization and situation awareness in the urban battlefield. Army Research Laboratory.

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.

Tran, H. (2018). Human-liked auditory detection capability for Intelligent Virtual Agents. *Interservice/Industry Training, Simulation, and Education Conference Proceeding (I/ITSEC) 2018.*

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).