Using Curriculum to Train Multisensory Foraging DRL Agents

Romulo F. Férrer Filho[1], Alexandre M. M. Santos[1], Halisson R. Rodrigues[1], Yuri L. B. Nogueira[1], Creto A. Vidal[1], Joaquim B. Cavalcante-Neto[1], Paulo B. S. Serafim[2]

[1]Departamento de Computação – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brazil

[2]Gran Sasso Science Institute
L'Aquila – Italy

{yuri,cvidal,joaquimb}@dc.ufc.br, {romulofffufc,halissonr}@gmail.com,

alexandre.santos@irede.org.br, paulo.desousa@gssi.it

Abstract. *Deep reinforcement learning has shown great success in developing agents that can solve complex game tasks. However, most game agents use only visual sensors to gather information about the environment. More recent works have shown that agents that use audio sensors can perform better than vision-only agents. In this paper, we propose a curriculum-based training strategy to develop agents that effectively use audio as a source of information in foraging-based scenarios. First, we demonstrate that agents with both vision and hearing capabilities perform similarly to agents with only a visual sensor, indicating that the first ones ignore the audio. Then, we show that by using a gradually increasing difficult curriculum the agent effectively uses the audio information available, making it more robust to survive in scenarios where visual information is not available. Our results indicate that agents can be trained to effectively use audio as a source of information by using a curriculum-based training strategy, improving their ability to deal with more tasks than agents with only vision.*

Keywords *Deep Reinforcement Learning, Game Agent, Curriculum Learning, Multisensory Agents, Foraging*

1. Introduction

Research on the development of game agents has shown an increased interest since the advent of Deep Reinforcement Learning (DRL) [Mnih et al. 2015]. From mastering traditional board games, such as Chess and Go [Silver et al. 2017], classic Atari games [Schrittwieser et al. 2019], as well as complex digital games, such as Dota 2 [OpenAI et al. 2019] and Starcraft [Vinyals et al. 2019], DRL has showcased its potential. Empowered by deep convolutional neural networks that excel in visual tasks, these agents can learn to play games from scratch, without any prior knowledge of the game rules. As such, the development of agents that can solve complex tasks in virtual environments has become a hot topic in the field of artificial intelligence.

Although the traditional reinforcement learning paradigm is inspired by animal learning, the majority of DRL research has focused on tasks that are primarily visual in

nature. Despite the advances in the development of game agents, they rely most notably on visual input. In some applications, agents receive additional numerical information, such as current energy level, and use proprioceptive sensors to gather information about their own state [Baker et al. 2020]. However, works that leverage audio capabilities in order to gather additional information about the environment are still rare.

In contrast, music and sound effects play a significant role in several aspects of video games. From the simple effects present in the first video game to use sounds, *Computer Space* from 1971, to the proposal of audio-only games [Hugill e Amelides 2016], technological advancement allowed modern approaches to leverage complex sound design development to promote a variety of multisensory interactions, affecting the players in different manners [Grimshaw et al. 2013]. For example, diverse audio employment can greatly increase a player's immersion, by provoking aesthetic and emotional reactions, as well as enabling accessibility and inclusion [Guillen et al. 2021].

Moreover, in the natural world, organisms use multiple sensory modalities to gather information In particular, hearing constitutes an important source of information to assist, for example, in environment navigation and finding resources. For example, in a foraging task, the available audio information is essential to help animals gather food while avoiding obstacles. Inspired by this biological capability, recent efforts have aimed to imbue game agents with auditory perception, enabling them to better understand and interact with their surroundings [Gaina e Stephenson 2019, Park et al. 2021].

Similarly, including audio as a source of information is also beneficial for video game agents. Prior work has shown that agents that listen to audio sources can perform better than vision-only agents [Hegde et al. 2021]. However, simply training an agent that receives both visual and hearing information might not be enough to make the agent effectively use the audio. A difficult challenge is to learn to correctly utilize the diverse information provided by different sensory inputs. Unlike traditional DRL settings, in which typically vision input dominates, multisensory environments require agents to process and integrate information from multiple sources simultaneously.

In this paper, we present a curriculum strategy for training multisensory game agents that effectively use both audio and vision as sources of information in a foraging task. We first show that agents with both vision and hearing capabilities can successfully survive in foraging-based scenarios. Then, we show that agents with only vision have a similar performance to agents with both vision and hearing capabilities in scenarios that are solvable with only visual information, which indicates that the audio is ignored. Finally, we demonstrate how a curriculum can be used to gradually train agents from unimodal to multimodal perception and enable them to navigate environments that might include vision or only sound information. Therefore, although the performance is similar in scenarios that are solvable with only visual information, an agent trained to effectively use audio will also succeed in scenarios with no visual information while other agents will get stuck. Figure 1 shows an overview of our proposal.

We evaluate the effectiveness of our proposed approach in foraging scenarios and provide insights into problems that arise in the learning process in multisensory settings. In particular, we answer the following research question: how can we train a multisensory

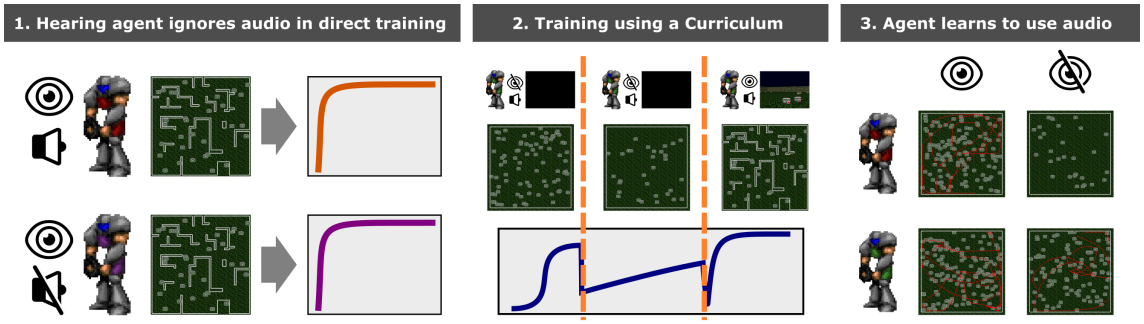


Figure 1. Overview of the proposed work.

exploratory agent that is capable of using both visual and auditory sensors effectively? We highlight three contributions that arise by answering the above question:

1. we show that using a direct training procedure for agents with both vision and hearing capabilities might make them focus only on visual input;
2. we present a training strategy that allows an agent to effectively use the audio as a source of information; and
3. we demonstrate that an agent trained to use audio is more robust to a wider range of scenarios since it will succeed in scenarios with no visual inputs.

Therefore, our work contributes to research on intelligent game agents by providing a framework for training more robust multisensorial agents.

The paper is organized as follows. In Section 2, we provide an overview and important concepts of the relevant topics in this work. In Section 3, we review the existing literature on multisensory DRL. In Section 4, we introduce our proposed approach for training multisensory DRL agents using curriculum learning and present our experimental setup and results. In Section 5, we discuss our findings and provide insights into the underlying mechanisms driving learning in multisensory settings. Finally, in Section 6, we conclude our work and outline future directions.

2. Background

In this section, we introduce the foundational learning methods that drives common research on Deep Reinforcement Learning agents and also present core concepts of the tasks and environments used in this work.

2.1. Learning Methods

2.1.1. Reinforcement Learning

Reinforcement Learning is a learning paradigm that focuses on the interaction of an agent with its environment, aiming to learn an optimal behavior for performing a task. In the traditional setting, it is composed by

- the environment: the world capable of hosting one or more agents, containing states, providing actions, and returning rewards as feedback;
- the agent: the entity capable of making decisions, taking actions, and learning by interacting with the environment; and

- the rewards: the feedback provided by the environment to the agents, indicating whether the action taken was good or not.

The main objective in Reinforcement Learning is for the agent to learn a policy, that is, a mapping of states to actions, in order to maximize the accumulated reward over time. This is achieved through a process of trial and error, where the agent explores the environment, collects the rewards of its actions, and updates its policy accordingly [Sutton e Barto 2018].

The main foundations of Reinforcement Learning include the Bellman Equation, which defines a value function $V(s)$ representing, recursively, the relationship between the immediate reward $R(t+1)$, and the discounted future value, $\gamma V_\pi(S_{t+1})$, starting from a given state $S_t = s$, as

$$V_\pi(s) = \mathbb{E}[R_{t+1} + \gamma V_\pi(S_{t+1}) | S_t = s], \tag{1}$$

where π represents the agent's policy, γ is the discount factor, and $\mathbb{E}$ simbolizes the expected value. This equation allows the agent to learn to estimate the value of each state, guiding it in choosing actions that maximize the long-term reward.

2.1.2. Proximal Policy Optimization (PPO)

PPO is a reinforcement learning algorithm that belongs to the policy gradient family [Schulman et al. 2017]. It is an improvement over earlier policy gradient algorithms and aims to address some of their limitations. PPO directly optimizes the policy function, which maps the agent's state to the probability distribution of actions. The goal is to find a policy that maximizes the expected cumulative reward. The key aspect of PPO is the use of a clipped surrogate objective function that limits the change in the policy during each update. This helps to prevent the policy from changing too much, which could lead to catastrophic performance drops. PPO keeps the updated policy close to the previous one, avoiding destabilizing the learning process.

The clipped surrogate objective function of PPO is formalized as

$$L^{CLIP}(\theta) = \mathbb{E}_t \left[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t) \right], \tag{2}$$

where $r_t(\theta)$ is the probability ratio given by

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}, \tag{3}$$

and $\hat{A}_t$ is the estimated advantage at time t, which can be calculated using the generalized advantage estimation (GAE) method

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l}, \tag{4}$$

where δ_t is the temporal difference error defined as

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t). \tag{5}$$

The clipping term clip$(r_t(\theta), 1-\epsilon, 1+\epsilon)$ limits the probability ratio to avoid excessively large updates, promoting stable and efficient exploration of the policy space.

In comparison with methods based on value functions, such as Deep Q-Networks (DQN) [Mnih et al. 2013, Mnih et al. 2015], PPO directly learns the policy function, which maps states to action probabilities. PPO balances exploration and exploitation based on the stochasticity of the learned policy, while DQN focuses on balancing exploration and exploitation through epsilon-greedy strategies. Overall, PPO is considered more stable and sample-efficient than DQN, as the clipped surrogate objective and proximal updates help to ensure gradual and stable policy improvements.

2.1.3. Curriculum Learning

The idea behind the Curriculum Learning technique is to improve the learning of agents by dividing the problem into stages of increasing difficulty. Presenting the problem gradually mirrors the human learning process, which starts by solving simpler tasks and progresses gradually to more complex ones. This brings more robustness and efficiency to the learning process [Bengio et al. 2009]. The key benefits of curriculum learning are:

1. it mimics human learning, starting with simpler tasks and gradually increasing complexity;
2. it leads to more robust and efficient learning, as the agent is not overwhelmed by the full complexity of the problem from the start; and
3. it can be applied to a variety of machine learning tasks, from computer vision to natural language processing.

By structuring the learning process in this way, curriculum learning has been shown to outperform traditional training approaches in many applications [Wang et al. 2022].

2.2. Environments and tasks

2.2.1. Foraging

Foraging is the act of searching for wild food resources. It involves making decisions about what food to choose, when to move to a new feeding patch, and how to optimize foraging efficiency in order to maximize energy gain [Westneat e Fox 2010]. Learning is crucial in foraging, as animals often change their behavior based on past experiences. This adaptive process, known as foraging innovation, involves animals trying new foods or using new techniques. In the context of reinforcement learning, when an agent performs foraging, it involves inspecting the search space for targets or food, and collecting what it is looking for.

2.2.2. VizDoom

VizDoom is a framework for the development of artificial intelligent agents based on the classic first-person shooter game Doom [Kempka et al. 2016]. It provides a highly customizable and visually realistic environment for the training of reinforcement learning agents. Some of its main contributions include the facilitated creation of new customized

scenarios, the ability to create levels with varying complexity, and the integration and use of deep neural networks, allowing agents to be trained to navigate and interact within the environment. It also stands out for its great customization capacity combined with an extensive community of mods, which allows for rapid prototyping and generation of custom levels. All these factors were essential for its widespread adoption by the reinforcement learning research community.

2.2.3. Sample Factory

The Sample Factory is a reinforcement learning system capable of controlling agents in a 3D environment using only screen pixels at a very high frame rate [Petrenko et al. 2020]. It employs an asynchronous processing approach, with worker processes collecting samples independently and sending them to the learner, which updates the model in parallel. This allows the system to achieve extremely high sampling rates, reaching up to 100,000 FPS, far surpassing traditional methods. In addition to its ease of use, with a simple interface, it integrates with various 3D simulation environments, including VizDoom.

In terms of performance, the authors report excellent results in the VizDoom environment, outperforming previous reinforcement learning methods in terms of efficiency and learning speed. The experiments showed that the system is capable of achieving performance comparable to human players in Doom deathmatch tasks, demonstrating the method's ability to learn complex behaviors from pixels alone.

The key features and contributions of the Sample Factory system are:

- high-performance reinforcement learning using an asynchronous, highly parallelized architecture;
- integration with a wide range of simulation and virtual environments;
- demonstrated state-of-the-art performance on complex reinforcement learning tasks; and
- capabilities to train agents presenting complex behavior directly from raw pixel inputs, without the need for manual feature engineering.

The Sample Factory's innovative design and impressive results make it a valuable tool for reinforcement learning research.

3. Related Works

Many of the initial works that show intelligent agents capable of winning games use the agent's vision as an input. However, other modalities, such as sound, are often neglected [Hegde et al. 2021], resulting in the under-utilization of Deep Reinforcement Learning algorithms based on sound sensors. Despite this, some works started to explore the application of sound sensors in problems involving intelligent agents in games [Latif et al. 2023].

[Hegde et al. 2021] used ViZDoom [Kempka et al. 2016] as a training environment for multi-sensory intelligent agents. In the paper, the authors present a modified version of the environment, in which three audio encoders were implemented and tested against each other. The first one uses two one-dimensional convolutional

layers on a downsampled input. Next, a Fourier Transform is applied to the audio buffer. After that, a Max-Pooling layer is used to downsample the data. The resulting output is then fed into two fully connected layers, each with 256 neurons. Lastly, a spectrogram encoder was also tested, which is very common in works of speech processing [Garcia-Romero et al. 2020]. In this scenario, the samples are also transformed to the frequency domain through a Short-Term Fourier Transform (STFT), and the generated spectrograms are fed to two two-dimensional convolutional layers. All agents were trained using the Proximal Policy Optimization (PPO) algorithm.

The authors evaluated their agents in four different scenarios. One consists in recognizing a specific sound in a room. Two of those are command recognition tasks, distinguished by the interval at which the command is given. The last one is a duel scenario, where one multi-sensory agent fights another with just one sensor (vision). The results of the experiments show a prevalence of hearing agents in relation to deaf ones in all scenarios.

[Giannakopoulos et al. 2021] used agents solely with audio in sound-based navigation and sound source localization tasks. They also investigate the possibility of reusing learning from one task to another. To do this, they created two environments using Unity3D,

- audio-based navigation: an agent is placed in a room containing 1 to 5 sound sources (besides being empty). One of these sources is designated as the target, and the agent must move toward that source without colliding with the others; and
- sound source localization: the previous scenario is repeated, but the agent is now fixed in the center of the room, and its movements are limited to rotation and adjusting the height. The goal is to locate all sound sources in the room by pointing the agent's microphone toward each one.

The study compares a human tester and a random agent to the autonomous agent trained with the PPO algorithm, both of which are outperformed by it. Additionally, they perform an analysis of the transfer learning application between the two presented tasks. From this, the authors conclude that the main challenge for a listening agent is sound source localization since it is a necessary skill for both tasks.

[Woubie et al. 2019] utilized sounds along with the screen pixels to train a set of agents in a localization task, which consists of finding a sound source randomly positioned within rooms in a 3D environment. The authors used ViZDoom for the experiments to incorporate audio into the agent and two approaches were evaluated. In the first approach, the pitch value is extracted from an audio clip and sent directly to the Convolutional Neural Network. The second approach uses a normalization of the raw audio signal. Thus, the authors created three agents, one that uses only vision and two others that use both vision and sound, each following one of the aforementioned audio capture approaches. The agents were trained using Deep Q-Networks [Mnih et al. 2015]. The experimental results show that the multi-sensory agents were able to learn to perform the task more easily, despite all of them being able to complete it.

All of the works described above make use of sound-based Deep Reinforcement Learning agents and discuss the implications of adding a sensor during agent training. However, the authors did not include experiments using a curriculum to train multisensory agents, which are explored in this paper. More specifically, these works do not evaluate

if audio information is ignored when the task presents visual information. In the context of a foraging environment, we found that achieving an agent that effectively uses sound inputs required the addition of a curriculum a strategy not proposed by other works to reach this goal. We summarize a comparison with prior work on Table 1.

Related Work	Method	Sensor Type	Curriculum
[Hegde et al. 2021]	**PPO**	Multisensory	No
[Giannakopoulos et al. 2021]	**PPO**	**Sound**	No
[Woubie et al. 2019]	DQN	Multisensory	No
Our approach	**PPO**	**Multisensory**	**Yes**

Table 1. Comparison of Related Work.

4. Experiments and Results

In this paper, our goal is to train an agent that effectively uses both visual and auditory information present in a foraging-based scenario. The most obvious way to start this kind of experiment is to directly train an agent capable of seeing the screen as well as hearing the audio emitted by the environment. As we show below, this type of direct training does not produce an agent with the desired abilities. In particular, since the task can be solved using only visual information, i.e. the pixels of the screen, the trained agent appears to completely ignore any audio input.

In order to successfully train an agent to use its audio capabilities, we developed a curriculum-based strategy using three scenarios of increasing difficulty. Moreover, since the vision seems to dominate over sound inputs, the agent cannot be able to see in the first curriculum stages. The complete set of experiments as well as the results obtained at each evaluation stage are described in detail in the following subsections.

4.1. Training Scenarios

The ViZDoom [Shao et al. 2018] platform provides two standard foraging scenarios called *Health Gathering* (HG) and *Health Gathering Supreme* (HGS), which have been used before as a testbed for DRL agents [Akimov e Makarov 2019]. These scenarios consist of a room surrounded by walls, with an acid floor, that causes damage over time to the player. The agent's objective is to stay alive as long as possible, to a maximum of 2100 in-game tics. To achieve this goal, the agent needs to learn to identify and gather *Medikits*, which increase its health points, and avoid collecting poison crates, that damage the player even further. However, the basic versions of both scenarios are soundless and we had to modify them to include audio. Two of the three scenarios created were based on HG and one on HGS. They all share the same implementation of in-game sound, as done by [Hegde et al. 2021], featuring 3D sounds, echo effects, and reverberation, which became standard in ViZDoom.

In order to make the scenarios solvable by using audio, we made individual medikits emit an assigned specific sound, which was kept looping as long as the item existed in the game world. The agent should then learn to recognize, locate, and collect these items to survive longer in each episode. The game engine additionally allows the control of the sound attenuation factor. We chose the ATTN_STATIC [ZDoom 2020] standard, which causes the sound to diminish rapidly as distance increases, requiring the

agent to get closer to the sound source to identify it. We also increased the maximum duration of each episode to 4200 tics, doubling the original 2100 from ViZDoom, and removed the poison crates. We kept the map size the same. Lastly, to define the number of medikits available on each map, we trained an agent on each scenario and evaluated its performance while decreasing the number of medikits on each run according to the following values: 30, 20, 15, 7. For instance, with 7 medikits, not even the easiest scenario was solvable. We chose 30 as it was the number that made all scenarios solvable, but not too easy when combined with the changes in medikit spawn rate, explained below, along with the main differences between the three scenarios.

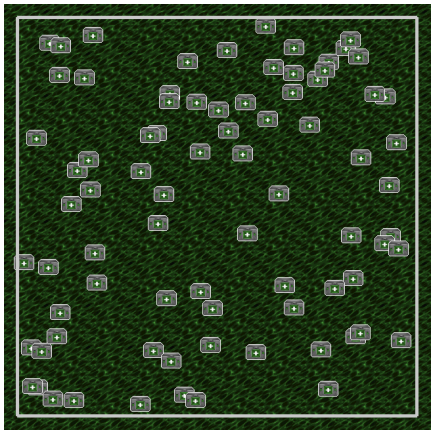
- **Health Gathering Sound + 30 (HG+30)**: Similar to the original scenario, but with sound. It starts with 30 medikits randomly scattered around the room and new medikits spawn every 25 in-game steps. This scenario has plenty of resources for the agent to collect, allowing it to receive rewards more frequently, making it the simplest and easiest one.
- **Health Gathering Sound = 30 (HG=30)**: It differs from the previous one by keeping only 30 medikits at a time in the game. It does not spawn new ones with time. New medikits only appear when the agent collects one that already exists, thus maintaining the same amount all the time. This scenario is more difficult, due to having a limited amount of resources for the agent to gather, which requires it to explore further in order to receive a reward.
- **Health Gathering Supreme Sound + 30 (HGS+30)**: This scenario is based on HGS, which differs from the previous two by having internal walls, which creates separate rooms. Here the spawn mechanic is the same as **HG+30**. As such, although abundant in resources, the agent must navigate through this maze-like scenario to find them, and using the sound to locate medikits across a wall might help the agent.

Figure 2 shows a top-down visualization of the three scenario, and Figure 3 shows the agent point of view on **HGS** and **HG+30**, which is the same as **HG=30**.

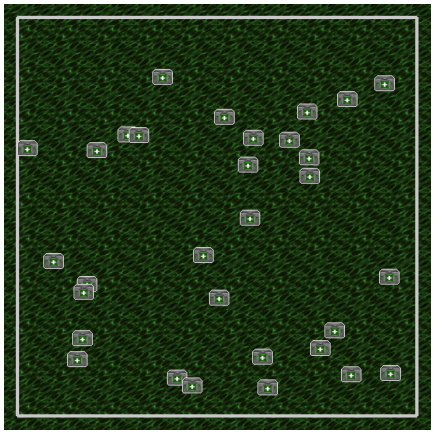
4.2. Training Strategy

One of our main goals was to develop an agent that leverages both the image and the sound of the environment to accomplish the foraging task. We implemented an agent equipped with sensors for both inputs, using the same Fast Fourier Transform audio encoder developed by [Hegde et al. 2021], but adding two more fully connected layers of size 256×256, and the standard ConvNet image encoder from Sample Factory [Petrenko et al. 2020]. We trained this agent, which we call *Image and Sound*, on all three scenarios individually to ensure it could learn from those, which it did, achieving peak performance and surviving all 4200 tics with around 400 million training steps for each scenario. The trajectory of an example run of this agent in the **HGS** scenario can be seen in Figure 4.

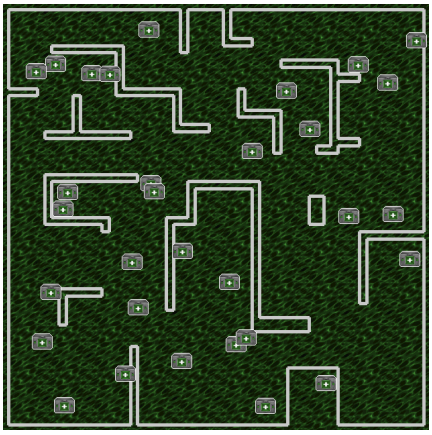
We then proceeded to check whether or not the audio information was relevant to that agent. To do so, we ran the same agent in the **HG=30** scenario in which is possible to play using audio only. What we observed was that when we removed the visual input, the agent was unable to play. It just stood at the spawn point in the center of the map, rotating around itself until it died. Figure 5 shows the agent's movements in this scenario.



(a) Health Gathering +30 (HG+30).

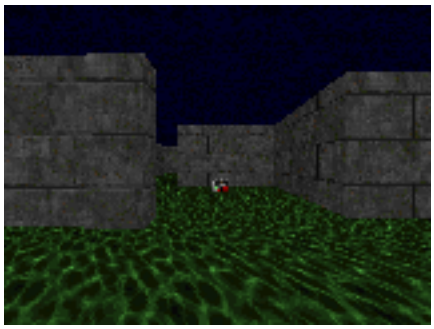


(b) Health Gathering =30 (HG+30).

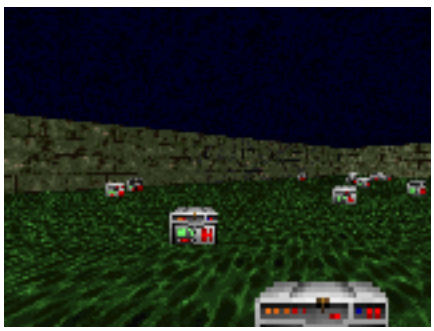


(c) Health Gathering Supreme +30 (HG+30).

Figure 2. Map visualization of the *Health Gathering* scenarios.



(a) Health Gathering +30 (HG+30).



(b) Health Gathering +30 (HG+30) and Health Gathering =30 (HG=30).

Figure 3. Agents' point of view. The main difference here is the internal walls of HGS seeing in Figure 3a.

Then, we performed another experiment to evaluate the behavior of the first agent. We trained a new agent using only the visual input and compared its performance to the previous one while using both inputs. In order to execute the evaluation, the audio input was zeroed, and a new agent, which we called *Image Only* was trained from scratch. After 400 million steps it was already achieving peak performance as the previous agent did.

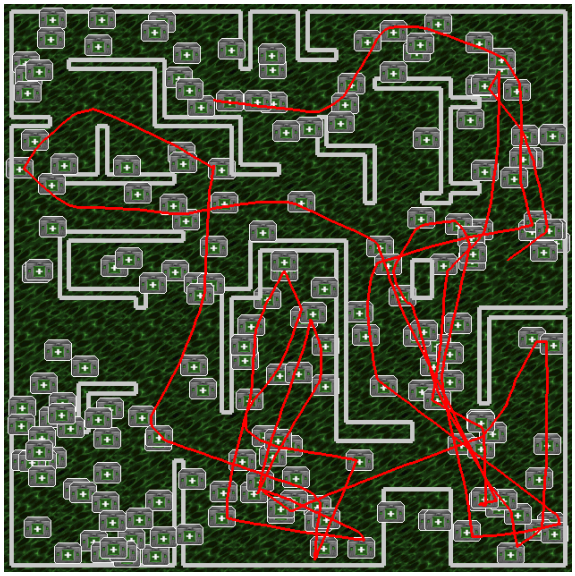
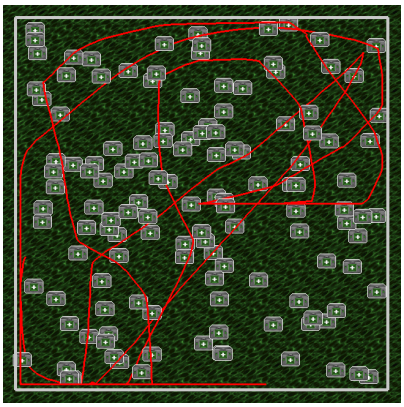
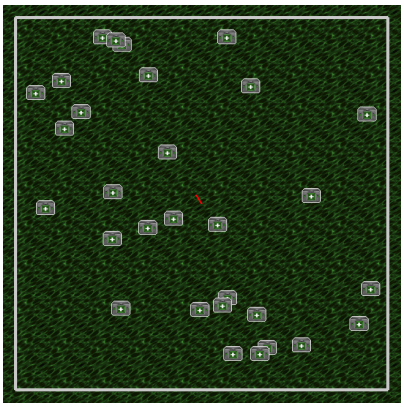


Figure 4. *Image and Sound* agent trajectory while running in the HGS scenario.



(a) *Image and Sound* agent with full access to inputs.



(b) *Image and Sound* agent with image input off.

Figure 5. The *Image and Sound* agent evaluation.

We measured the average reward received by both agents during training, which is shown in Figure 6.

Finally, we trained an agent using a curriculum strategy developed to force it to learn how to guide itself using both visual and audio information. The curriculum consisted of a combination of the three previously mentioned scenarios, from the easiest to the hardest one: **HG+30** $\rightarrow$ **HG=30** $\rightarrow$ **HGS+30**. However, in the first two training scenarios only the sound sensor is active, forcing it to learn how to survive using only the sound emitted by the medikits. When it gets to the third scenario, we include the vision sensor as well, needed to learn how to avoid the internal walls of the maze.

The *Curriculum* agent trained for 500 million steps on the **HG+30** scenario, 1 billion steps on **HG=30**, and finally 600 million steps on **HGS+30**. It achieved satisfying results in each scenario individually, and by the end of the entire training, it was able to achieve the same performance as the two previous agents, learning to survive the

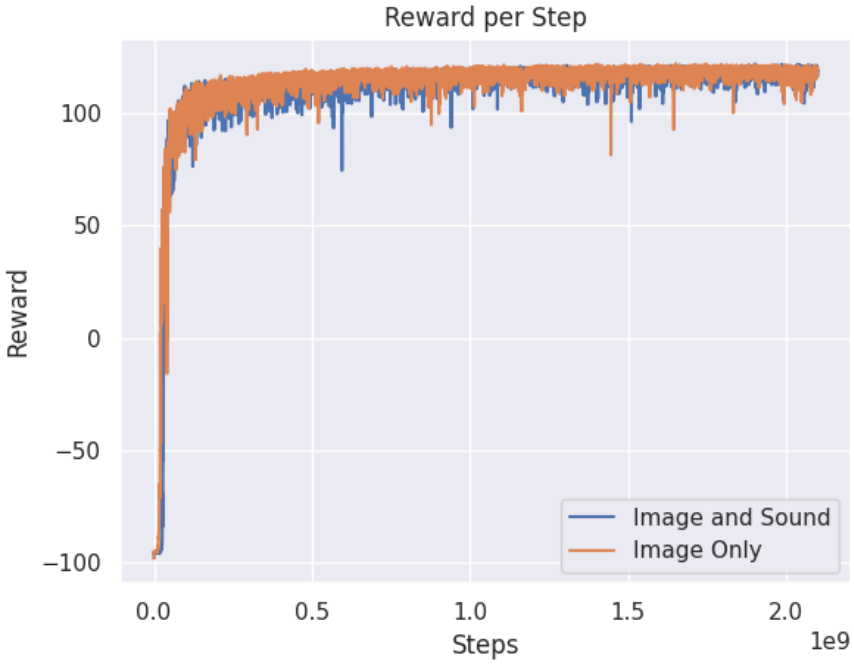


Figure 6. Reward received during training for both *Image and Sound* and *Image Only* agents.

entire episode duration. Figure 7 shows the reward received during the entire curriculum execution, and Figure 8 shows the trajectory of the agent running in the **HGS** scenario.

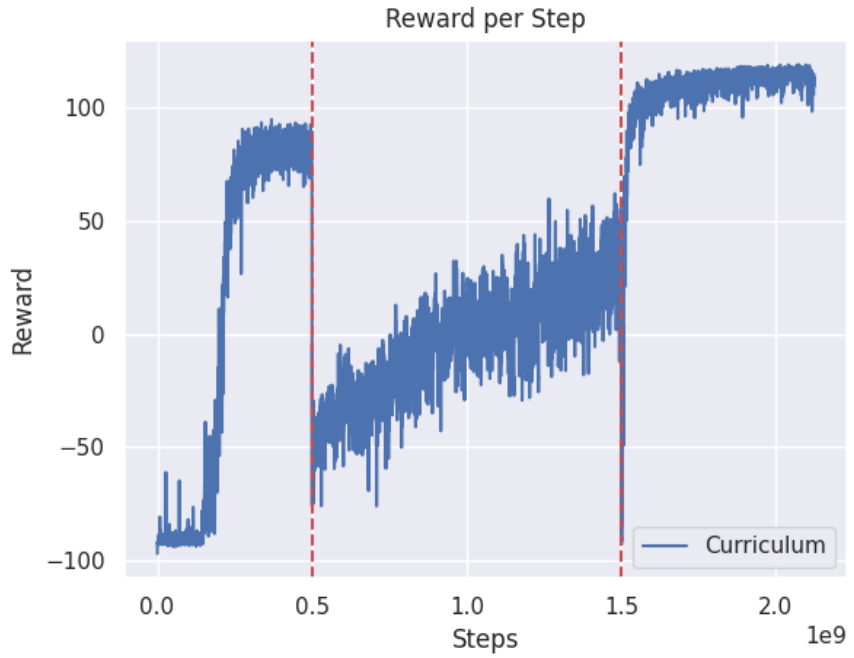


Figure 7. Reward received during training by *Curriculum* agent. The vertical red lines indicate when the curriculum phase changed.

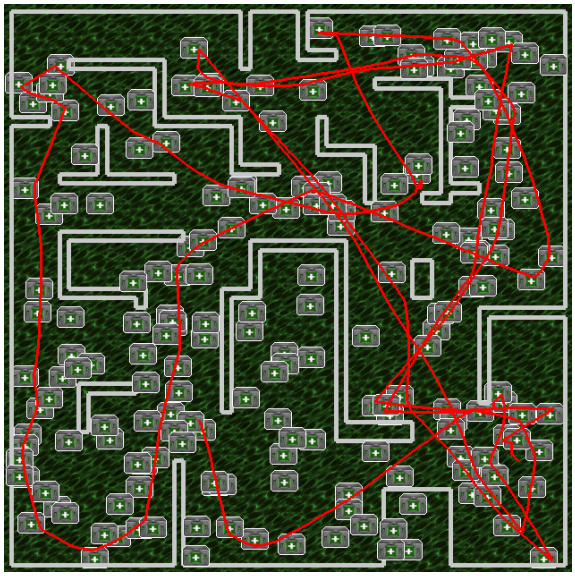
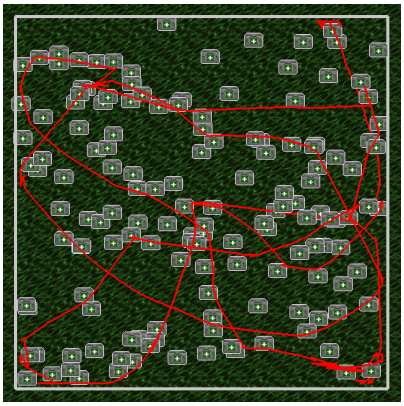
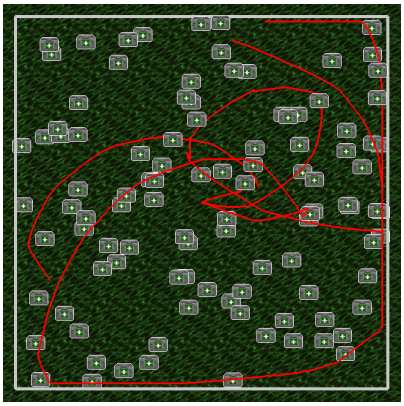


Figure 8. *Curriculum* agent trajectory while running in the HGS scenario.

With our *Curriculum* agent trained, we repeated the previous blind test on the scenario **HG=30** and plotted the traversed path by this agent. The main difference when compared to Figure 5 is seen on the 5b versus 9b. By forcing the agent to first learn to use only the audio input, the final agent is capable of using both image and sound to guide itself and survive the foraging scenario.



(a) *Curriculum* agent with full access to inputs.



(b) *Curriculum* agent with image input off.

Figure 9. The *Curriculum* agent evaluation.

All agents were trained using the Sample Factory library, due to its capacity to instantiate multiple scenarios simultaneously, greatly speeding up the training process. All experiments ran in a local machine equipped with an NVIDIA RTX 2070 Super Graphics Card, 32GB of DDR4 Memory, and an Intel i7 10700 CPU.

5. Discussion

As mentioned in Section 4, the first agent, referred to as *Image and Sound*, was able to learn to complete the foraging task, but lacked the ability to guide itself by the audio

input when the image was not available. Our hypothesis is that given vision is sufficient to complete the *Health Gathering* scenarios, as shown by the second agent, named *Image Only*, if we start the training with sensors for both inputs activated, the agent would rely solely upon the image one. We believe this happens because the image input is more valuable for the agent in this task, thus the agent learns to ignore the audio input, focusing only on what it sees.

The blind experiment also reflects this behavior, in Figure 5b the now blind agent is just turning around itself, in a behavior we call *Turrent Behavior*, and it only moves slightly forward, barely leaving the spawn point, although it has the sound sensor active. Figure 10, shows the exact same behavior for 5b, but now performed by the *Image Only* agent, showing that both consider solely the image input for deciding which action to perform.

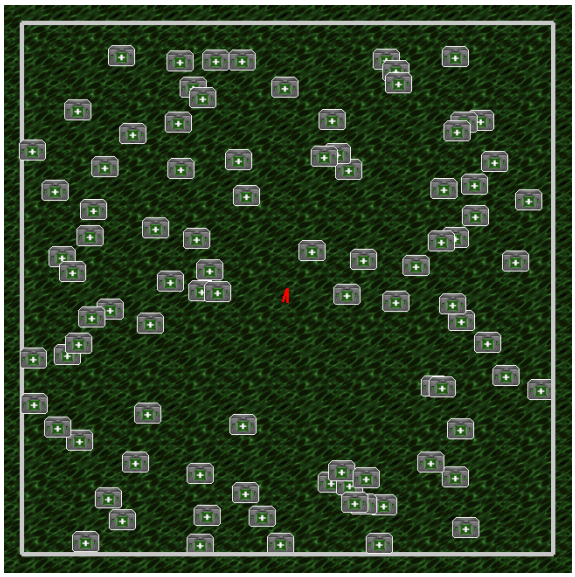


Figure 10. Trajectory of the *Image Only* agent in the Blind evaluation.

On the other hand, training an agent with the proposed *Curriculum* still achieves the same final performance of surviving all 4200 tics, but running the same blind evaluation shows that it indeed learned how to use the sound to locate the *Medikits*. When we deactivate the agent's vision, it loses some performance, but it is still capable of surviving, not replicating the *Turrent Behavior* of the previous ones, as shown in Figure 9.

5.1. Evaluation of action distribution

Table 2 shows the difference in action distribution between the three agents (*Image and Sound*, *Curriculum*, and *Image Only*) running blind for 1000 episodes, once again showing that the *Curriculum* one is capable of leveraging the sound to decide which action to perform. The *Turrent Behavior* is also exemplified in Table 2, where around 98% of all actions performed by the *Image and Sound* and *Image Only* agents were the *Turn Right* action as if it was looking for a visual input to start moving forward. In turn, the *Curriculum* agent has a much broader distribution of actions.

We also evaluated the action distribution in the **HGS** scenario while running all agents with their respective sensors fully active. Table 3 shows that all three agents,

	Image and Sound	***Curriculum***	*Image Only*
No Action	0.41%	**5.83%**	0.48%
Left	0.48%	**1.93%**	0.27%
Right	98.29%	**30.72%**	98.15%
Forward	0.23%	**57.20%**	0.09%
Backwards	0.59%	**4.32%**	1.01%

Table 2. Distribution of actions on HG=30 scenario with blind agents.

when running at their full capacity, have a very similar distribution of actions performed, highlighting how all three of them achieve similar results and behaviors. However, we note that the agents with audio capabilities perform the backward action more often, which indicates that they might be able to move towards the medikits without having to turn and see them. Moreover, the *Image Only* performs more often the forward action and less frequently stands still, which corroborates the fact that it has to move in order to have medikits in its field of view.

	Image and Sound	*Curriculum*	*Image Only*
No Action	2.28%	3.06%	1.57%
Left	18.17%	16.05%	14.57%
Right	20.84%	22.15%	24.81%
Forward	57.66%	57.66%	58.20%
Backwards	1.05%	1.08%	0.84%

Table 3. Distribution of actions on HGS scenario.

When comparing the distribution of *Left* and *Right* actions only, it is notable how the *Curriculum* agent is more balanced, followed by the *Image and Sound*, while the *Image Only* prefers to turn *Right* most of the time. This distribution can be seen in Table 4. These action distribution results suggest that agents with audio sensors have a more fluid behavior, moving and turning towards medikits more evenly. This contrasts with vision-only agents, which are frequently moving, in order to see the medikits, and have a preferred rotation direction.

	Image and Sound	*Curriculum*	*Image Only*
Left	42.02%	**46.58%**	37.00%
Right	57.98%	**53.42%**	63.00%

Table 4. Distribution of Left and Right actions on HGS scenario.

5.2. Summary

The experiments yielded several significant findings, which are summarized below.

- ***Image and Sound* Agent Performance**: The initial agent, equipped with both visual and auditory sensors, successfully learned to complete the foraging tasks in all scenarios, reaching peak performance and surviving all 4200 tics after about 400 million training steps per scenario. However, in the absence of visual input, it failed to navigate effectively, highlighting an over-reliance on visual cues.

- ***Image Only* Agent Performance**: A second agent trained solely on visual inputs achieved similar performance to the multisensory agent. This indicates that visual information alone is sufficient for the foraging tasks in the designed scenarios, explaining the multisensory agent's failure when deprived of visual input.
- **Curriculum Training Strategy**: To address the limited use of auditory cues, we introduced a curriculum training strategy that forced the agent to rely initially on auditory information before integrating visual input. The agent was trained in three phases, starting with sound-only input in simpler scenarios and progressing to combined input in the most complex scenario. The curriculum-trained agent demonstrated robust performance across all scenarios. Notably, it maintained the use of auditory cues in vision-deprived tests, thus avoiding the "Turret Behavior" observed in the other agents.
- **Action Distribution and Behavioral Analysis**: The action distribution analysis showed that the curriculum-trained agent exhibited a more balanced action distribution compared to both the multisensory and visual-only agents. This indicates a more refined decision-making process, effectively leveraging both visual and auditory inputs.

Our findings support the effectiveness of curriculum training strategies in fostering multisensory integration in DRL agents.

6. Conclusion

In this study, we developed a deep reinforcement learning (DRL) agent capable of multisensory foraging using both image and sound inputs. Using the ViZDoom platform, we designed a series of increasingly difficult scenarios to evaluate the agent's performance under different conditions and training strategies.

A direct training approach produced an agent that learns to complete a foraging task using both sensors but fails to succeed in a scenario with only audio input. Our experiments demonstrate that a curriculum strategy can train an agent to effectively survive in an audio-only scenario. By initially emphasizing less dominant sensory inputs, we can develop agents capable of adaptive and robust behavior in varying sensory environments. This research contributes to the fields of autonomous game agents and foraging and opens new avenues for future work in multisensory integration for other complex tasks.

Future work could explore additional sensory modalities and more intricate curriculum strategies to further enhance agent adaptability and performance. New scenarios could be implemented to leverage the agents' capacity to learn from different degrees of audio information. Also, in our work, only one audio file was used during the experiments, an interesting question is whether or not an agent can learn different patterns of audio simultaneously and variations of the same audio (pitch, speed, tone).

The work presented in this paper focuses on demonstrating that a well-designed curriculum can significantly improve the learning and adaptation capabilities of DRL agents, ensuring they can utilize all available sensory information to make informed decisions in complex, dynamic environments. However, applying the same curriculum strategy to different applications and neural network configurations could improve even further the agent's capabilities.

References

Akimov, D. e Makarov, I. (2019). Deep reinforcement learning in vizdoom first-person shooter for health gathering scenario. *MMEDIA*, pages 1–6.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., e Mordatch, I. (2020). Emergent tool use from multi-agent autocurricula. In *International Conference on Learning Representations (ICLR)*, pages 1–28.

Bengio, Y., Louradour, J., Collobert, R., e Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48.

Gaina, R. D. e Stephenson, M. (2019). "did you hear that?" learning to play video games from audio cues. In *2019 IEEE Conference on Games (CoG)*, pages 1–4.

Garcia-Romero, D., Sell, G., e Mccree, A. (2020). Magneto: X-vector magnitude estimation network plus offset for improved speaker recognition. In *Proc. Odyssey 2020 the speaker and language recognition workshop*, pages 1–8.

Giannakopoulos, P., Pikrakis, A., e Cotronis, Y. (2021). A deep reinforcement learning approach for audio-based navigation and audio source localization in multi-speaker environments. *CoRR*, abs/2110.12778.

Grimshaw, M., Tan, S.-L., e Lipscomb, S. D. (2013). Playing with sound: The role of music and sound effects in gaming. In *The Psychology of Music in Multimedia*. Oxford University Press.

Guillen, G., Jylhä, H., e Hassan, L. (2021). The role sound plays in games: A thematic literature study on immersion, inclusivity and accessibility in game sound research. In *Proceedings of the 24th International Academic Mindtrek Conference*, Academic Mindtrek '21, page 12–20, New York, NY, USA. Association for Computing Machinery.

Hegde, S., Kanervisto, A., e Petrenko, A. (2021). Agents that listen: High-throughput reinforcement learning with multiple sensory systems. In *2021 IEEE Conference on Games (CoG)*, pages 1–5.

Hugill, A. e Amelides, P. (2016). *Audio-only computer games: Papa Sangre*, page 355–375. Cambridge University Press.

Kempka, M., Wydmuch, M., Runc, G., Toczek, J., e Jaśkowski, W. (2016). Vizdoom: A Doom-based AI research platform for visual reinforcement learning. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8. IEEE.

Latif, S., Cuayáhuitl, H., Pervez, F., Shamshad, F., Ali, H. S., e Cambria, E. (2023). A survey on deep reinforcement learning for audio-based applications. *Artificial Intelligence Review*, 56(3):2193–2240.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., e Riedmiller, M. A. (2013). Playing atari with deep reinforcement learning. *ArXiv*, abs/1312.5602.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. a., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A.,

Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., e Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

OpenAI, Berner, C., Brockman, G., Chan, B., Cheung, V., Debiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., de Oliveira Pinto, H. P., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F., e Zhang, S. (2019). Dota 2 with large scale deep reinforcement learning. *ArXiv*, abs/1912.06680:1–66.

Park, K., Oh, H., e Lee, Y. (2021). Veca: A toolkit for building virtual environments to train and test human-like agents. *arXiv preprint arXiv:2105.00762*.

Petrenko, A., Huang, Z., Kumar, T., Sukhatme, G. S., e Koltun, V. (2020). Sample factory: Egocentric 3d control from pixels at 100000 FPS with asynchronous reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 7652–7662. PMLR.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., e Silver, D. (2019). Mastering atari, go, chess and shogi by planning with a learned model. *ArXiv e-prints*, pages 1–21.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., e Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint 1707.06347*.

Shao, K., Zhao, D., Li, N., e Zhu, Y. (2018). Learning battles in vizdoom via deep reinforcement learning. pages 1–4.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., e Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *ArXiv e-prints*, pages 1–19.

Sutton, R. S. e Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, 2nd edition.

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., et al. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354.

Wang, X., Chen, Y., e Zhu, W. (2022). A survey on curriculum learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4555–4576.

Westneat, D. e Fox, C. W. (2010). *Evolutionary behavioral ecology*. Oxford University Press.

Woubie, A., Kanervisto, A., Karttunen, J., e Hautamäki, V. (2019). Do autonomous agents benefit from hearing? *CoRR*, abs/1905.04192.

ZDoom (2020). Acs playsound documentation. `https://zdoom.org/wiki/PlaySound` [Accessed: (28/05/2024)].