applied sciences

MDPI

Article

A New Technical Ear Training Game and Its Effect on Critical Listening Skills

Sungyoung Kim [1,2,*] and Jacob Cozzarin [3]

[1] College of Engineering Technology, Rochester Institute of Technology, Rochester, NY 14623, USA
[2] Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[3] Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA; jpc4394@g.rit.edu
* Correspondence: sxkiee@rit.edu; Tel.:+1-585-475-5706

Abstract: Technical ear training has proven to be an effective tool for developing the skills of junior audio engineers and enhancing their proficiency in audio and music production. To provide a comprehensive auditory training experience, the authors have created a gamified training program that encompasses four modules: spectral identification, auditory localization, consistent judgment, and memory of a mix balance. Each module is designed to give trainees closed-loop audiomotor training, allowing them to instantly assess their performance and identify areas where they need to improve. This new ear-training game helped players to be more engaged and resulted in the improvement of trained audio engineering skills. Moreover, the game also benefited a non-trained auditory skill, speech understanding in noise.

Keywords: ear-training; gamification; auditory plasticity

check for updates

Citation: Kim, S.; Cozzarin, J. A New Technical Ear Training Game and Its Effect on Critical Listening Skills. *Appl. Sci.* **2023**, *13*, 5357. https://doi.org/10.3390/app13095357

Academic Editor: José A. González-López

Received: 8 March 2023
Revised: 11 April 2023
Accepted: 20 April 2023
Published: 25 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Critical Listening in Various Fields

The audio engineering community has seen a growing interest in "ear-training" in recent years. The advancements in digital audio workstations and audio signal processing tools for audio/music production have made it easier for amateur musicians and engineers to produce music. Despite the ease of access to these advanced technologies, producing good sound quality remains a big challenge. People may know how to use the tools, but they often struggle to produce the desired sound without a strong aesthetic understanding of a sound field. To address this issue, audio engineering schools have updated their curricula with a greater focus on "listening skills" and "musical training" [1]. Contemporary students need to learn how to critically evaluate sounds and make adjustments to control their tools for the best possible sound quality. This shift in pedagogy is reflected in the increasing number of research activities and publications in the field. For example, two recent journal papers have been published in the Journal of the Audio Engineering Society [2,3], demonstrating the growing interest in this area. In one of these papers, Elmosnino [3] provides a comprehensive review of previous and ongoing research in critical listening training, starting from Miskiewicz's pioneering works [4] in the late 1980s and Quesnel's breakthrough with the *Technical Ear Trainer (TET)* [5], a computerized ear-training program in the late 1990s. The paper also introduces the recent curricula of various audio education programs.

The importance of ear training and critical listening extends beyond practical applications in audio engineering. Researchers have studied the relationship between listeners' level of training and experience, and the impact it has on their ability to critically "evaluate" sound quality. In this subjective sound quality evaluation, the experience and training of listeners has a "huge impact on the data and its quality ([6], p. 87)". Experts in the field have a higher sensitivity and ability to perceive differences, making them more reliable in

their assessments compared to untrained listeners. In practical situations, listening skill levels are dependent on the nature of the given tasks. For example, ITU-T Rec. P.800 [7] requires naive listeners, while ITU-R Rec. BS.1534-3 [8] requires experienced listeners. Recent studies on spatial attributes for three-dimensional (3D), immersive sound recording, and reproduction also revealed cognitive differences between naive and experienced listeners [9,10]. According to ([6], p. 91), the difference between the expert and naive listeners lies in that "the small expert panel may have the experience to discriminate repeatably complex stimuli using the scale with confidence". Olive's study [11] indicated that untrained listeners prefer the same loudspeakers (and headphones) as the panel of trained listeners. However, trained listeners are more discriminating and reliable than untrained listeners, supported by their individual F_L statistics for the "loudspeaker" variable. F_L is defined as "the ratio of the loudspeaker effect (mean sum of squares for loudspeaker (preference) ratings) divided by the error variance (mean sum of squares of the residual). This metric accounts for the listeners' ability to discriminate between loudspeakers as well as their ability to repeat their ratings". If a smaller number of "trained or expert" subjects could bring an equivalent, meaningful conclusion quickly, it will save budget and time for researchers conducting a large number of subjective evaluations. To address this, companies that conduct a significant amount of subjective evaluations have formed listening panels consisting of trained or expert listeners, and have developed tools to assess and train their listening skills. For example, Harman International R&D has created a free training program called *Harman How to Listen* [12,13], which is used to train their panel members. Similarly, Yamaha Corporation has collaborated with Kyushu University, which has a rich history of ear-training curricula and its own training software [14], to train its engineers. These two programs are distinct in their training modules, which focus on conceptualization and quantification of auditory attributes. For example, one of the Harman "How to Listen" program's training modules is Attribute Training, in which trainees are asked to quantify a specific auditory attribute such as brightness.

In recent years, researchers have been focused on finding ways to improve the efficacy of ear-training programs. Through categorizing previous ear-training programs and curricula, it was found that the three most frequent practices for critical listening training are verbal communication, sequencing (adaptive adjustment of training difficulties), and software-based technical ear training (as shown in Table 2 of [3]). This highlights the importance of interaction with an instructor and an adaptive aspect of training sequences in critical listening training. However, it is not always possible to have an interaction with a dedicated instructor. To overcome this limitation, researchers have attempted to make ear-training programs adaptive and interactive. The goal of these programs is to guide the trainee in the same manner as a private lesson so that a trainee can practice without an instructor. This field of study is known as Intelligent Tutoring Systems (ITS) [15], in which systems provide customized instructions without the intervention of human beings. In recent e-learning studies, this adaptive and interactive feature is of importance for maintaining a trainee's motivation even in a self-learning scenario [16]. The development of ITS in ear-training as a game has the potential to take the training beyond being a mere teaching aid and provide a more personalized and effective learning experience. Examples of this gamified training include (but are not limited to) music [17], hard-of-hearing listeners' speech perception in noisy environment [18], and auditory localization [19].

1.2. General Gain of Critical Listening Training

A recent study by Whitton et al. [20] reported that "*closed-loop*" audiomotor training is an effective training method, especially to improve listeners' speech understanding in the presence of noise. A closed-loop training enables trainees to interact with the given environment and learn to recover from its own mistakes. It is a training approach that involves continuously monitoring a trainee's performance, such as accuracy or error rate, and adjusting the training process in response to that performance in real-time. Their study is distinct from other relevant ones [21,22] because the research was focused on how

much the training efficacy can be expandable to non-trained general auditory function(s). This "plasticity" or "transferability" of perceptual training to general tasks has been a debatable issue among researchers. In particular, opponents who are skeptical of the benefit of specific perceptual training argue that a training-related benefit would not "transfer far beyond the training stimuli [23]". Musical training, in contrast to those opponents' claim, is known to be effective to improve multiple auditory functions. What is the reason for this unique merit of musical training? Whitton and his colleagues suggested that musical training features a multimodal, interactive activity between listening and self-adjustment, the so-called "closed-loop" audiomotor training. For example, a violinist carefully and repeatedly controls his or her fingers to articulate tuned pitch amid other competing instrumental sounds. This move-and-match strategy in a critical listening activity seems to activate neural plasticity [24], which in return results in enhanced multiple auditory functions (as argued in [25,26]). Based on this merit of musical training that diversifies training efficacy in multiple auditory functions, the authors designed and tested a new training paradigm similar to a musical practice/training that features the "closed-loop audiomotor" training. The study result showed that when trainees utilized auditory and motor sensors simultaneously with real-time feedback, the training appeared to improve a non-trained task as well such as music training. Readers should note that music training and performance is not the only way to improve critical listening skills. Recently, Lad et al. [27] comprehensively investigated the relationship between "musical sophistication" (as a general concept of engagement with music) and working memory for frequency, showing a significant correlation between the two. Nonetheless, there is evidence that music training can be specifically effective for improving selective auditory attention, which enables people to separate a target speech signal from non-target maskers (noises) [28].

Therefore, if any critical listening training program could incorporate this "closed-loop audiomotor" feature, the training efficacy does not only increase for the trained tasks, but also stimulates and transfers to non-trained tasks. This motivated us to redesign our previously-developed technical ear training programs [19,29–33] to incorporate two aforementioned features: (1) gamification to increase a trainee's motivation through interactivity and adaptive sequencing, and (2) closed-loop audiomotor training to broaden the efficacy of training to general, non-trained skills.

In the following sections, the authors will describe the design strategy of the new gamified auditory training program, detailed functions of each of the four training modules, a training-induced influence on speech identification in a noisy environment, and discussion.

2. Design and Development of a New Technical Ear Training Game

The authors had discussed how to design a new technical ear training game and embedded the aforementioned two features: "interactivity and adaptive sequencing" and "closed-loop audiomotor training". The initial plan was to make a first-person shooter (FPS) game in an imaginary world where all the visual information is limited or not available. In the game, a player would rotate one's whole body towards a target position and try to shoot the target. If the player failed to shoot the target down, the game would show how much of an error was made. This feedback would help the player to learn and re-strategize for the next trial. With this plot, we implemented an augmented reality (AR) program in which players hunt invisible mice by tracking their sound information [34]. We further developed the game plot as follows: a game player enters a kingdom where they need to hunt "invisible" monsters. Each player would need to train basic skills to obtain arrows. When the player removes all the target monsters, a new challenge will be given in the next mission. The left panel of Figure 1 shows the initial plot of the technical ear training game.

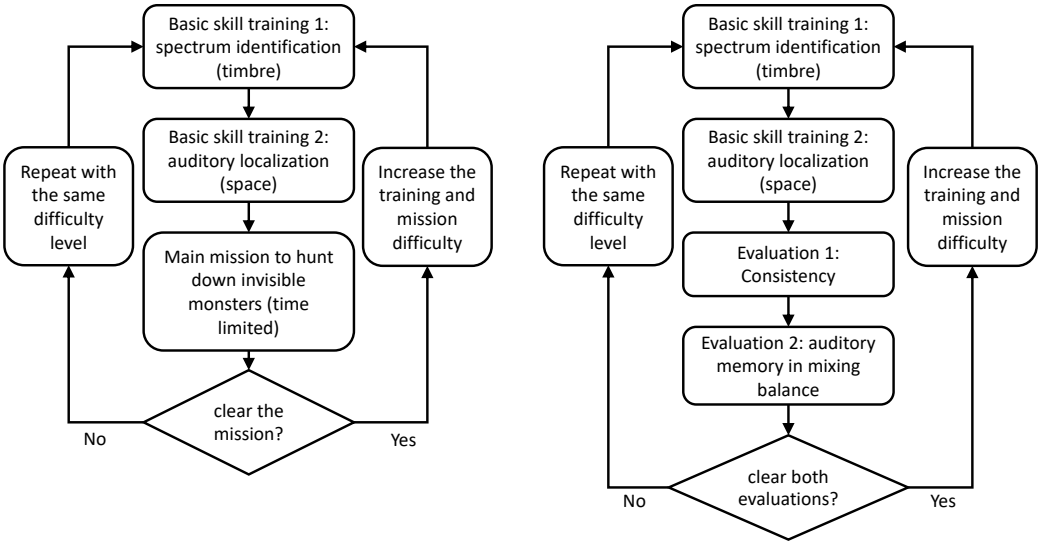

Figure 1. (**Left**) Original plot of the proposed technical ear game structure. Two basic training modules provide players with a chance to challenge the main mission—catching invisible monsters. (**Right**) Modified design of the game with two new modules—evaluating a player's consistency in timbre judgment and memory of a mixing balance of multiple sound sources.

This idea was slightly modified after in-depth internal and external discussions. The right panel of Figure 1 shows the modified plot and strategy of the proposed ear training game. One outstanding critique from the discussions was that the initial design focused too much on auditory localization training. The second basic skill and the main mission were essentially the same auditory localization training. Furthermore, another critique noted that the audio industry (as a potential user of this training game) would need a panel of trained subjects who could conduct a solid quantitative evaluation of their products with high *reliability* as stated in the previous Section 1.1. Bakker [35] termed these trained subjects as "appointed listeners" who can consistently judge regardless of various internal and external conditions. To evaluate whether a trainee could be qualified and serve as an appointed or trained listener, who judges reliably and consistently with a solid, built-in internal reference, a new module of "consistency" evaluation was added.

In addition, we included a module that trains the auditory memory of a mixing balance of multiple sound sources. The module provides a player with a randomly generated mix balance of multiple sound sources and asks them to memorize the mix within a short period of time (less than a minute). Mixing is the one of core tasks for an audio engineer, who carefully mediates the sonic characteristics of each sound source and balances them to satisfy the artistic demands of the performer and client. While this is such an important task in the audio production chain, no formal teaching method for "mixing skill" exists (to the best knowledge of the authors). It is probably because mixing is much closer to an artistic and aesthetic expression of the audio engineer's view of the balance of sound components. Therefore, a junior audio engineer usually mimics what previous mix engineers have done from reference albums. Many audio engineers compare reference albums or tracks and try to replicate the sonic quality in the mixing balance as close as possible to the reference. This fourth module utilizes this concept of a mixing process: to steal (by memorizing) others' mix performance and replicate it as similarly as possible.

As the right panel of Figure 1 shows, the modified strategy now incorporates two types of activities: (1) training of two basic skills (timbre and space) and (2) evaluation of two applied skills (consistency and mixing balance). Furthermore, the first two modules (basic training modules) provide players with an opportunity to acquire a star token (shown in Figure 2), depending on their performances, as a training incentive. The star token serves as the number of the player's "life" in the evaluation modules. Without a star token,

a player must repeat the training of two basic skills if they failed to complete the evaluations. However, with star tokens, players can challenge the evaluation without repeating two basic skills. Players should complete all four modules to move to the next level.



Figure 2. Main screen of the new ear-training game. A player needs to complete four training modules, the first two basic skills (spectral identification (S) and auditory localization (L)) and two following challenges to prove that one can judge consistently (C) and memorize (M) the given mix balance. The current player in this figure has completed the basic skills (S) and (L), and is going to challenge the consistency (C) evaluation in level 2.

Technical Details of Game Development

The user interface of the game that players interact with is implemented using Unity version 2017.3.1f1. Unity provides the following: a framework to develop the auditory training game into a 2D platform implementation; a scripting framework to program the game's logic using C#; and the ability to build the game for multiple platforms without modification to the original codebase. Unity seemed to be a great way to implement the game once and be able to target it to many platforms such as Android, iOS, Linux, and Universal Windows Platform (UWP) for the greatest amount of future-proofing. Vuforia Engine version 6.2.9 was heavily utilized to integrate an Augmented Reality SDK into the Unity framework for development. The targeted platform (and main test platform) for the game was UWP so that it could be loaded onto a Microsoft Surface tablet. The Surface tablet was an easily accessible platform with the capability for Augmented Reality given it had a back-facing camera. The unity application was set to build for UWP where a final build distribution could be side loaded to a Surface tablet or hosted on a private Windows Store to be downloaded/updated from.

Motivation for storing user data outside of the local game was needed to gather research data on auditory training progression across all users. A separate server was used to host a database where user data and progression in the game would be tracked. A Raspberry Pi 3 Model B (running Raspberry Pi OS for its operating system) would be used as a cost-effective server hosting a MySQL database responsible for handling requests from the game client to insert and update game data. For players to track their game completion data and for new users to create their own accounts to be stored in the database, there needed to be a way for the game client to communicate with the external database. This communication was implemented by having a PHP web server on the Raspberry Pi server receive web form data from the game client and send back response data. The PHP scripts would then handle the connection with the database and perform the appropriate database insert/query and return any information associated with the operation back to the game client.

3. Details of Four Training Modules

As illustrated in the right panel of Figure 1, the proposed new critical listening training game consists of four modules: two basic training modules—spectral identification (S) and auditory localization (L), and two challenge modules to prove that a user can judge with consistency (C) and could memorize (M) an auditory scene and replicate it. We integrated these four modules into a single ear-training game in which users challenge and conquer training tasks to move towards advanced levels (as illustrated in Figure 2). In this section, the details of the function and training goals of each of these four modules will be introduced.

3.1. Spectral Identification

This basic training aims to let players map spectral change in sound to technical parameters of an equalizer, especially the frequency on which the spectral change occurs. This spectral identification and match has been a key training component for many existing technical ear training programs [2,5,13,14,30,36,37], which together evinced the training efficacy of this spectral training. Therefore, the authors included this training in the game to improve trainees' detection of subtle timbral changes. In this module, players are asked to "identify" one boosted or cut spectral band by a notch filter, the center frequency of which varies with a 1/3 octave distance from 63 H̃z to 16,000 Hz. In the beginning, trainees start with the large difference (+12 d̃B peak) and a subset of seven center frequencies (1 octave distance from 125 H̃z to 8 k̃Hz, as shown in the left panel of Figure 3). As the game level increases, the filter gain reduces to +1 d̃B and utilizes the full 26 center frequencies (as shown in the right panel of Figure 3). For a detailed listing of the various parameters that change the difficulty of the training per level, see Table A1.

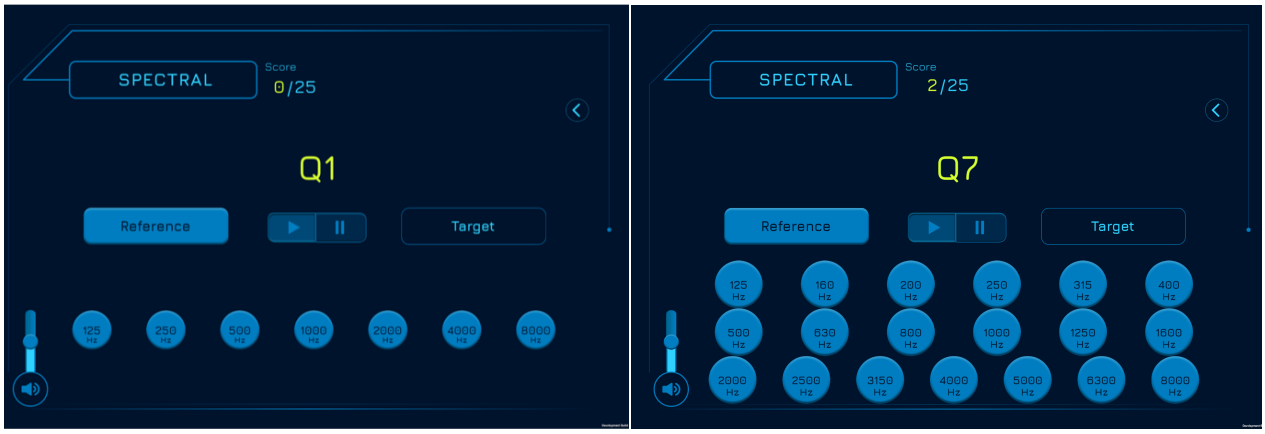

Figure 3. Graphic user interface (GUI) of spectral identification (S) training module. In the beginning, the task is relatively easy with a smaller number of frequency choices ((**right panel**), level 1) with the large gain, yet it becomes difficult for higher levels with more center frequencies ((**left panel**), level 16) and a much smaller filter gain value (+1 d̃B). The score indicates the number of correct answers out of the total number of questions. The right panel indicates that the player is currently on question 7 (out of a total of 25 questions) and has answered 2 questions correctly so far.

3.2. Auditory Localization

Once players completed the spectral identification (S) training module, they are asked to find the position of invisible sound objects. This training is an extension of the previously described (in the Section 2) AR auditory localization (L) training. To integrate with other training modules, we changed the training platform from Microsoft HoloLens to Surface, a tablet that a player can hold and rotate their whole body towards a target as shown in the right panel of Figure 4. While the player in the figure did not wear a pair of headphones for illustration purposes, players should wear a pair of headphones to accurately trace the target sound position processed through a head-related transfer function (HRTF). The left

panel of Figure 4 shows the module's GUI. The dotted yellow line is the horizontal plane and an invisible target object is located along the line in the beginning. A player needs to rotate his or her body until the gauge points to the target sound object and press the button in the right bottom corner. The current player has four chances in total and did not use any shooting chance yet. If the player catches the target object within the allowed number of chances, the mission will be cleared to the next module. The game checks whether the current position of the cursor is within a given radius of the target sound source. This radius around the target sound source continues to shrink as the levels increase, requiring more precision from the player. Furthermore, as the level increases, the invisible targets are spread over the entire three-dimensional space and a player needs to move to search for the objects both horizontally and vertically. The increased level put additional cognitive loads to make the training more difficult. The target sound object will start moving its location and there will be non-target sound objects that will mask the target. Finding the precise target position amid competing and interfering maskers is a challenging mission, similar to a real-world communication scenario such as speech understanding in noise. This module's training efficacy (to improve the localization accuracy) was validated through previous studies [19,38,39]. For a detailed listing of the various parameters that change the difficulty of the training per level, see Table A2. The main parameter that controls the difficulty is "Target Guess Radius". The values specified in Table A2 are arbitrary 3D distance units and should be used as a comparison of how large the circular area is for obtaining a correct guess between different levels. For example, level 6 has a guess radius of 1.25 and level 7 has a guess radius of 1. Between these two levels, the guess radius decreases by 36%.

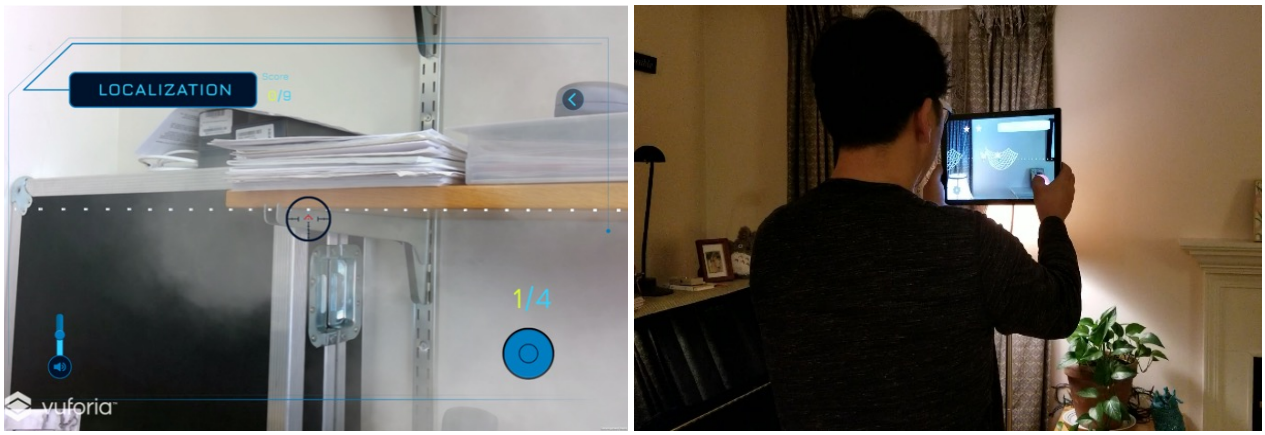

Figure 4. (**Left**) GUI of the localization (L) training module. (**Right**) Illustration of an in situ localization game. To trace the target sound object, the player rotates their entire body to match the sound's position with the gauge and then presses the button in the bottom right corner to shoot the object. Unlike this illustrated image, players should wear a pair of headphones to accurately trace the target sound position. This closed-loop audiomotor training is a core basic training component to improve the precision of a player's spatial awareness.

3.3. Consistency

This module aims to evaluate and improve the player's consistency (C), especially in manipulating the sound spectrum. In a real-world application of spectral manipulation, audio engineers determine a desirable sound spectrum based on their previous experiences built and shaped through extensive amounts of critical listening. These experiences let them judge sound fidelity with internal consistency and external validity. High consistency is one of the criteria for a professional, appointed, trained or expert listener, as supported by a previous study [33]. To assess the player's consistency, the game applies random gains for a high-shelving filter and low-shelving filter. Players are then asked to reverse the filter gains using the dials shown in Figure 5 to equalize the spectrum to a flat state. These gain dials have a step size of ± 1 dB and a range from -12 dB to +12 dB. The deviation between

the equalized spectrum and the reference flat spectrum is then calculated, and the trainee's consistency is determined based on this deviation. Similar to two previous modules, if the deviation is smaller than the threshold value in each level (which decreases as the level goes up), a player moves to the next module, memory (M) of mix balance.



Figure 5. GUI of consistency (C) evaluation module. A player controls two dials to manipulate the sound spectrum to become "flat". The deviation between the manipulated spectrum and the reference flat spectrum is then calculated, and the player's consistency is determined based on this deviation. As the level moves up, the required deviation to pass decreases.

3.4. Memory Game of Mixing

Working memory is defined as a cognitive ability to temporarily store and manipulate a small amount of information in a readily accessible form. It is important for people's everyday functioning and facilitates their abilities to plan, comprehend, reason, maintain attention, and solve a problem. The benefits of working memory "training" have also been reported, showing improved performance including (but not limited to) reasoning tasks and reading comprehension, as well as other cognitive functions in daily life (please refer to Cowan's comprehensive review [40] for more information). Auditory working memory has a complex relationship with a listener's hearing sensitivity and auditory processing capability, as noted in [41]. Auditory working memory and critical listening ability are two domains that interact and influence each other, leading to enhanced neural connectivity and improved auditory functions. Our initiative was to make a new auditory working memory (M) game incorporating the core skill of audio engineering—mixing. Figure 6 shows the GUI of this memory of a mixing balance module. Players first need to identify and "select audio" that is being played. Subsequently, players need to memorize two technical parameters: volume and panoramic-potentiometer (pan-pot). The difficulty increases from one track (the left panel) up to three tracks. The right panel of Figure 6 shows the memory game level 16 in which players are required to memorize types of three audio sources and their volumes and pan-pot information. For a detailed listing of the various parameters that change the difficulty of the training per level, see Table A3. If their replicated values are similar to the given, original mix balance, they clear the level and will move to the next level.

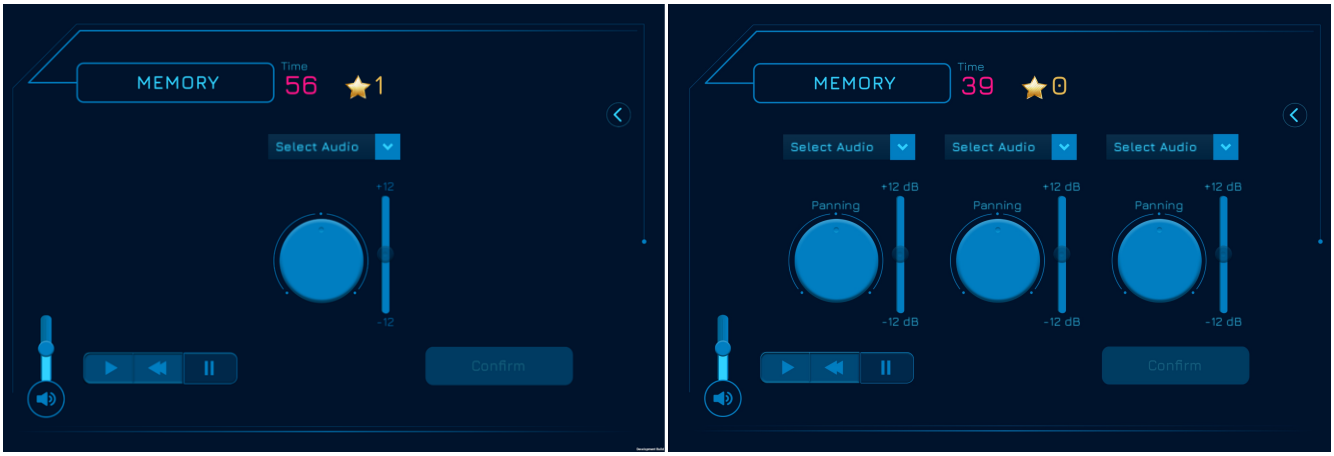

Figure 6. GUI of the memory (M) of a mixing balance module. Players are to memorize computer-generated mix balance of one to six sound sources (depending on the level) and replicate. (**Left panel**) is the UI of the level 1, where players need to memorize a single sound source, its level, and the pan-pot position, and subsequently match them using given controllers (fader and knob). (**Right panel**) is the level 16 where players need to do the same task for three sound sources.

4. Evaluation: Training Benefit in Speech Understanding

The authors observed that all of these training activities are essentially similar to the "closed-loop audiomotor" training. Audio engineers control a physical device, such as a fader or knob, to adjust the timbral and spatial characteristics of target sound objects. In this context, an audio device is equivalent to a musician's instrument. As musicians adjust their tuning using their audiomotor interaction, the audio engineers control the device to achieve the required sonic quality. For example, if a given sound source has an unwanted +3 dB boost at 1 kHz with Q of 2, you need to find the exact amount of variation by adjusting the control devices and listening to sonic variations. Therefore, technical ear training could be equally effective at improving other auditory-related skills, as a benefit of closed-loop audiomotor training. We investigated this hypothesis through a controlled experiment and the details of which will be presented in the following subsections.

4.1. Methods

The authors formed two listener groups: the Control group with 4 participants who studied advanced topics in audio engineering for 14 weeks (without the closed-loop audiomotor training from the proposed game) and the Treatment group with 4 participants who only trained themselves with the proposed game (thus with the closed-loop audiomotor training) for the same 14 weeks. We measured and compared the two groups' speech understanding in noise (SiN), which is a non-trained auditory skill to investigate whether the training benefit has been transferred to a non-trained skill. The two groups' performances were first compared with the Reference group with 34 participants who did not receive any training at all. SiN performance was measured through the Coordinate Response Measure (CRM) corpus by Bolia et al. [42]. The background maskers are speeches that compete and interfere with a given focal target speech. Figure 7 illustrates a task where all listeners are asked to attend to the focal cue of "Ready Tiger" and understand the color and number. In the figure above, the correct answer would be "red" and "5", to which the listeners need to respond by selecting the corresponding button on a tablet (by pressing the red-colored 5 square on the screen). As for the other task attending to the background cue, the correct answer would have been "blue" and "2".

The experiment also adapted spatial separation of background from focal. Two recent studies [28,43] showed that the spatial location of background sounds is an important experimental variable for auditory selective attention. The experiment incorporates this factor by testing six different positions. We used the correct answer ratio of this CRM as a metric of listeners' SiN performance.

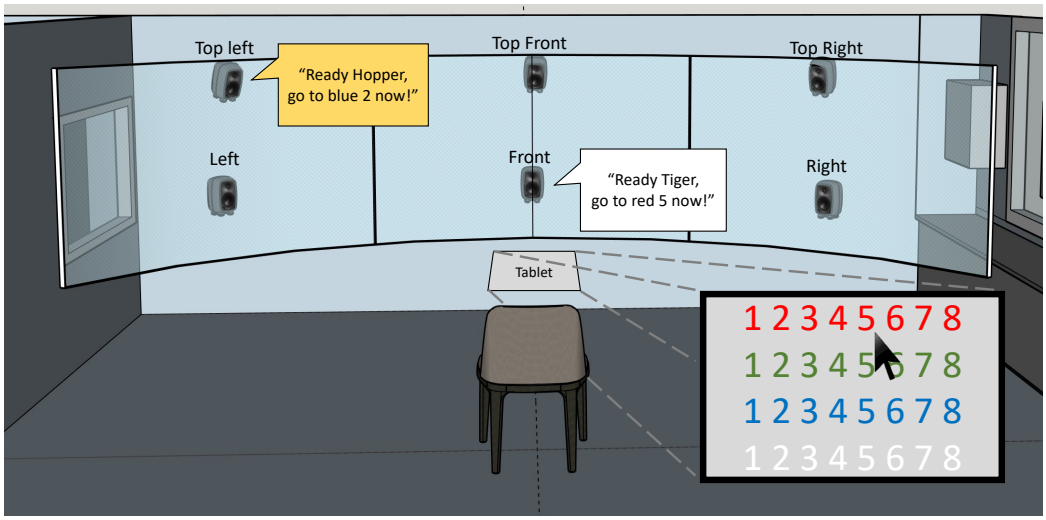

Figure 7. The experimental setup. The focal auditory cue "Ready Tiger" is presented through the front loudspeaker, and the background auditory cue "Ready Hopper" is presented through one of six spatial positions. A participant responds to the color and number of the instructed auditory cue (red 5 or blue 2) using a tablet.

4.2. Results

We analyzed the data with two spatial-location conditions: Collocated vs. Separated condition. The Collocated condition was where both focal and background speeches were reproduced from the same, one location; the Separated condition was where focal and background speeches were reproduced from different locations (as illustrated in the previous Figure 7). In addition, the experiment adapted two tasks: the "Focal" task asked listeners to identify the focal-cued information (color and number) while the "Background" task asked them to identify the background-cued information. The results showed that all three listener groups (Reference, Control, and Treatment) did not make any significant difference in the "Focal" task, for both collocated and separated conditions.

However, in the "Background" task, the Collocated condition caused a significant difference between the reference and treatment group ($F(2, 79) = 6.35$, $p = 0.003$). This result indicates that the given training (long-term and audiomotor training) helped the treatment group to better perform SiN in the hardest and thus most demanding task (Collocated condition of the "Background" task). The Separated conditions (both "Focal" and "Background") appear easy for both the Treatment and Control groups due to spatial masking release, resulting in no significant difference (please refer to the following section for more details). This experiment did not account for different masker (background) levels. The authors plan to conduct follow-up experiments and investigate the potential influence of various masker levels.

5. Discussion

The authors designed the auditory localization (L) training module to improve a trainee's auditory spatial sensitivity. Specifically, this training module aims to assist listeners in identifying the spatial location of a target sound amid interfering maskers. This accurate identification of spatial location is an element of auditory spatial attention. Therefore, the authors assumed that the training would support enhancing auditory spatial attention including SiN performance, especially for the "Separated" condition. However, the current result did not show any significant change. It is probable that the training-induced improvement was smaller than spatial release from masking (SRM as introduced in [44]) in this experiment. In other words, "Separated" SiN tests were easy (due to

SRM) both for the control and treatment group, and thus incapable of properly evaluating training-induced effects.

We also analyzed the listeners' response times, the result of which showed a significant change between the two groups in the separated condition. While their correct answer ratios were similar in this experiment, the treatment group responded more quickly after the training. It may indicate that the treatment group has become more confident, potentially through the auditory localization training module (which is in line with previous auditory localization training [45]). The training enabled trainees to reduce cognitive loads, resulting in ease of listening and/or effortless listening. We would like to investigate whether this improves listeners' confidence and would lead to improvement in SiN performance for more cognitively demanding tasks such as low target-to-masker levels.

The current study focuses on training for auditory selective attention as a means to potentially improve speech understanding in noisy environments. However, hearing difficulties can be associated with other types of impairments, such as auditory processing disorder (APD), which is a neurological condition in which the brain has difficulty in phonological processing and interpreting sounds. It remains unclear whether the proposed gamified training approach is effective for individuals with APD across the board. Nonetheless, a previous study conducted by the first author showed that spectral sensitivity training could improve word identification, particularly for words differentiated by vowels [45]. This may suggest that the gamified training, specifically the spectral training module, could potentially benefit individuals with APD.

It is an important research question to investigate the long-term effect of the auditory training program as a preventative tool for young adults. The use of earphones and exposure to high levels of sound is becoming more prevalent, causing people to suffer from partial or permanent hearing loss. Hearing loss may disconnect people from social communication in public places (due to poor SiN performance) and this disconnected communication may subsequently cause cognitive impairments [46]. This is why an auditory training program is drawing researchers' attention: it may be possible that preventative auditory training (with appropriate contents and methods) could help reduce the risk of hearing loss in young adults. In fact, short-term auditory training appeared to improve the speech processing function even for elderly subjects [47]. Furthermore, if the training is embedded in a game and given subconsciously, it may be more appealing to young adults being incorporated into their daily routines. It will be valuable to see how the auditory training program may impact the auditory selective attention performance in young adults over time.

6. Conclusions

The authors have developed a new auditory training game with four modules—spectral identification, auditory localization, consistency, and memory of mix balance—for self-training of technical listening skills required for junior audio engineers and audio engineering students. The game's challenge-and-conquer strategy and adaptive increase in training difficulty, combined with closed-loop audimotor training, have made it an effective tool for improving not only target audio engineering skills but also non-trained auditory functions, such as speech understanding in noise. The fact that the game's benefits extend beyond the specific skills trained, similar to the effects of musical training, supports its potential for bringing generalized improvements in auditory functions. This auditory training game is expected to offer a promising tool for enhancing auditory skills with significant implications for individuals across a range of professions and contexts.

Author Contributions: Conceptualization, S.K.; methodology, S.K.; software, J.C.; validation, S.K.; formal analysis, S.K.; writing—original draft preparation, S.K.; writing—review and editing, S.K. and J.C.; visualization, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yamaha Corporation.

Institutional Review Board Statement: The study was approved by the Institutional Review Board of Rochester Institute of Technology (protocol code HSRO#05101519 and 25 October 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy of individuals' training performance. The ear-training game is also available on request for test purposes only.

Acknowledgments: The authors thank for Justin Levine, Kyunghwan Sul, and Song Hui Chon (Belmont University, TN) for their help in designing and evaluating the earlier version of the auditory training program from which the current version can be completed. This research was supported by Yamaha Corporation and was carried out under the Cooperative Research Project Program of the RIEC, Tohoku University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SiN Speech understanding in Nose
SRM spatial release from masking

Appendix A

This appendix details the various parameter settings for the Spectral Identification, Auditory Localization, and Memory training programs that are part of the technical ear training game. Tables A1–A3 below show the parameter values that are set for each level.

Table A1. Table of all Spectral Identification parameters and their values for each individual level.

Level	Steps between Freqs	Min Freq	Max Freq	Boost to Target Freq	Total Questions
1	1 Octave	125 Hz	8000 Hz	+9 dB	25
2	1 Octave	125 Hz	8000 Hz	+9 dB	25
3	1 Octave	125 Hz	8000 Hz	+6 dB	25
4	1 Octave	125 Hz	8000 Hz	+6 dB	25
5	1 Octave	125 Hz	8000 Hz	+3 dB	25
6	1 Octave	125 Hz	8000 Hz	+3 dB	25
7	1 Octave	125 Hz	8000 Hz	+1 dB	25
8	1 Octave	125 Hz	8000 Hz	+1 dB	25
9	1 Octave	63 Hz	16,000 Hz	+9 dB	25
10	1 Octave	63 Hz	16,000 Hz	+6 dB	25
11	1 Octave	63 Hz	16,000 Hz	+6 dB	25
12	1 Octave	63 Hz	16,000 Hz	+3 dB	25
13	1 Octave	63 Hz	16,000 Hz	+3 dB	25
14	1 Octave	63 Hz	16,000 Hz	+1 dB	25
15	1 Octave	63 Hz	16,000 Hz	+1 dB	25
16	1 Octave	63 Hz	16,000 Hz	+1 dB	25
17	1/3 Octave	63 Hz	8000 Hz	+9 dB	25
18	1/3 Octave	63 Hz	8000 Hz	+9 dB	25
19	1/3 Octave	63 Hz	8000 Hz	+6 dB	25
20	1/3 Octave	63 Hz	8000 Hz	+3 dB	25
21	1/3 Octave	63 Hz	16,000 Hz	+9 dB	25
22	1/3 Octave	63 Hz	16,000 Hz	+6 dB	25
23	1/3 Octave	63 Hz	16,000 Hz	+3 dB	25
24	1/3 Octave	63 Hz	16,000 Hz	+3 dB	50

Table A2. Table of all Auditory Localization parameters and their values for each individual level.

Level	Number of Noises	Target Gain	Target Delay (Seconds)	Moving Speed	Height Angle	Target Guess Radius	Time Limit (Seconds)	Number of Tries	Goal Distance
1	0	0 dB	0	None	0°	1.25	∞	4	3
2	0	0 dB	0	None	0°	1.25	∞	3	3
3	1	0 dB	1	None	0°	1.25	∞	4	3
4	1	0 dB	1	None	0°	1.25	∞	3	3
5	1	0 dB	1	Base	0°	1.25	30	4	3
6	1	0 dB	1	Base	0°	1.25	30	3	3
7	2	−6 dB	1	×2	15°	1	∞	4	6
8	2	−6 dB	1	×2	15°	1	∞	3	6
9	2	−6 dB	2	×2	15°	1	30	4	6
10	2	−6 dB	2	×2	15°	1	30	3	6
11	2	−6 dB	2	×2	15°	1	∞	4	6
12	2	−6 dB	2	×2	15°	1	∞	3	6
13	3	−12 dB	3	×4	30°	1	30	4	6
14	3	−12 dB	3	×4	30°	1	30	3	6
15	3	−12 dB	3	×2	15°	0.75	∞	4	9
16	3	−12 dB	3	×2	15°	0.75	∞	3	9
17	2	−12 dB	2	×4	30°	0.75	30	4	9
18	2	−12 dB	2	×4	30°	0.75	30	3	9
19	3	−12 dB	3	×4	30°	0.75	∞	4	9
20	3	−12 dB	3	×4	30°	0.75	∞	3	9
21	3	−12 dB	3	×4	30°	0.75	30	4	9
22	3	−12 dB	3	×4	30°	0.75	30	4	9
23	3	−12 dB	3	×4	30°	0.75	30	4	9
24	3	−12 dB	3	×4	30°	0.75	30	3	9

Table A3. Table of all Memory game parameters and their values for each individual level.

Level	Panning Knob Step Size	Volume Slider Step Size	Exposure Time (Seconds)	Total Mixing Time (Seconds)	Number of Sound Sources to Mix
1	5	5	60	∞	1
2	7	9	60	∞	1
3	13	9	60	∞	1
4	13	9	45	∞	1
5	13	9	30	∞	1
6	5	5	60	∞	2
7	7	9	60	∞	2
8	13	9	60	∞	2
9	13	9	45	∞	2
10	13	9	60	120	2
11	5	5	60	∞	3
12	5	5	60	∞	3
13	7	5	60	∞	3
14	5	9	60	∞	3
15	7	9	60	∞	3
16	7	9	45	∞	3
17	7	9	30	∞	3
18	7	9	60	120	3
19	13	9	60	∞	3
20	13	9	60	∞	3
21	13	9	30	∞	3
22	13	9	60	120	3
23	13	25	60	∞	2
24	13	25	60	∞	2

References

1. Indelicato, M.J.; Hochgraf, C.; Kim, S. How Critical Listening Exercises Complement Technical Courses to Effectively Provide Audio Education for Engineering Technology Students. In Proceedings of the Audio Engineering Society 137th International Convention, AES, Los Angeles, CA, USA, 9–12 October 2014.
2. Brezina, P. Perspectives of advanced ear training using audio plug-ins. *J. Audio Eng. Soc.* **2021**, *69*, 351–358. [CrossRef]
3. Elmosnino, S. A review of literature in critical listening education. *J. Audio Eng. Soc.* **2022**, *70*, 328–339. [CrossRef]
4. Miśkiewicz, A. Timbre Solfege: A Course in Technical Listening for Sound Engineers. *J. Audio Eng. Soc.* **1992**, *40*, 621–625.
5. Quesnel, R. Timbral Ear Trainer: Adaptive, Interactive Training of Listening Skills for Evaluation of Timbre. In Proceedings of the Audio Engineering Society 100th International Convention, AES, Copenhagen, Denmark, 11–14 May 1996; Preprint 4241 .
6. Zacharov, N. (Ed.) *Sensory Evaluation of Sound*; CRC Press: Boca Raton, FL, USA, 2019.
7. ITU-R. *Recommendation P.800, Methods for Subjective Determination of Transmission Quality*; International Telecommunications Union Radiocommunication Assembly: Geneva, Switzerland, 1996.
8. ITU-R. *Recommendation BS.1534-5, Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*; International Telecommunications Union Radiocommunication Assembly: Geneva, Switzerland, 2015.
9. Howie, W.; Martin, D.; Kim, S.; Kamekawa, T.; King, R. Effect of Audio Production Experience, Musical Training, and Age on Listener Performance in 3D Audio Evaluation. *J. Audio Eng. Soc.* **2019**, *67*, 782–794. [CrossRef]
10. Kim, S.; Howie, W. Influence of the Listening Environment on Recognition of Immersive Reproduction of Orchestral Music Sound Scenes. *J. Audio Eng. Soc.* **2021**, *69*, 834–848. [CrossRef]
11. Olive, S.E. Differences in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study. *J. Audio Eng. Soc.* **2003**, *51*, 806–825.
12. International, H. Harman's How To Listen. Available online: http://harmanhowtolisten.blogspot.com/2011/01/welcome-to-how-to-listen.html (accessed on 10 February 2023).
13. Olive, S.E. A new listener training software application. In Proceedings of the Audio Engineering Society 110th International Convention, AES, Amsterdam, The Netherlands, 12–15 May 2001; Preprint 5384.
14. Iwamiya, S.; Nakajima, Y.; Ueda, K.; Kawahara, K.; Takada, M. Technical Listening Training: Improvement of sound sensitivity for acoustic engineers and sound designers. *Acoust. Sci. Technol.* **2003**, *24*, 27–31. [CrossRef]
15. Sleeman, D.; Brown, J. *Intelligent Tutoring Systems*; Academic Press: New York, NY, USA, 1982.
16. Garris, R.; Ahlers, R.; Driskell, J.E. Games, motivation, and learning: A research and practice model. *Simul. Gaming* **2002**, *33*, 441–467. [CrossRef]
17. Pesek, M.; Vučko, Ž.; Šavli, P.; Kavčič, A.; Marolt, M. Troubadour: A Gamified e-Learning Platform for Ear Training. *IEEE Access* **2020**, *8*, 97090–97102. [CrossRef]
18. Kim, S.; Emory, C.; Choi, I. Neurofeedback Training of Auditory Selective Attention Enhances Speech-In-Noise Perception. *Front. Hum. Neurosci.* **2021**, *15*, 676992. [CrossRef]
19. Chon, S.H.; Kim, S. Auditory Localization Training Using Generalized Head Related Transfer Functions in Augmented Reality. *Acoust. Sci. Technol.* **2018**, *39*, 312–315. [CrossRef]
20. Whitton, J.P.; Hancock, K.E.; Shannon, J.M.; Polley, D.B. Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise. *Curr. Biol.* **2017**, *27*, 3237–3247.E6. [CrossRef] [PubMed]
21. Burk, M.H.; Humes, L.E. Effects of Long-Term Training on Aided Speech-Recognition Performance in Noise in Older Adults. *J. Speech Lang. Hear. Res.* **2008**, *51*, 759–771. [CrossRef] [PubMed]
22. Ferguson, M.A.; Henshaw, H.; Clark, D.P.; Moore, D.R. Benefits of Phoneme Discrimination Training in a Randomized Controlled Trial of 50- to 74-Year-Olds With Mild Hearing Loss. *Ear Hear.* **2014**, *35*, e110–e121. [CrossRef]
23. Owen, A.M.; Hampshire, A.; Grahn, J.A.; Stenton, R.; Dajani, S.; Burns, A.S.; Howard, R.J.; Ballard, C.G. Putting brain training to the test. *Nature* **2010**, *465*, 775–778. [CrossRef]
24. Anderson, S.; Kraus, N. Auditory Training: Evidence for Neural Plasticity in Older Adults. *Perspect. Hear. Hear. Disord. Res. Diagn.* **2013**, *17*, 37–57. [CrossRef]
25. Iliadou, V.; Ptok, M.; Grech, H.; Pedersen, E.R.; Brechmann, A.; Deggouj, N.; Kiese-Himmel, C.; Śliwińska-Kowalska, M.; Nickisch, A.; Demanez, L.; et al. A European Perspective on Auditory Processing Disorder-Current Knowledge and Future Research Focus. *Front. Neurol.* **2017**, *8*, 622. [CrossRef]
26. Iliadou, V.; Kiese-Himmel, C. Common Misconceptions Regarding Pediatric Auditory Processing Disorder. *Front. Neurol.* **2017**, *8*, 732. [CrossRef]
27. Lad, M.; Billing, A.J.; Kumar, S.; Griffiths, T.D. A specific relationship between musical sophistication and auditory working memory. *Sci. Rep.* **2022**, *12*, 3517. [CrossRef]
28. Swaminathan, J.; Mason, C.R.; Streeter, T.M.; Best, V.; Kidd, G., Jr.; Patel, A.D. Musical training, individual differences and the cocktail party problem. *Sci. Rep.* **2015**, *5*, 11628. [CrossRef]
29. Kim, S.; Kaniwa, T.; Terasawa, H.; Yamada, T.; Makino, S. Inter-subject differences in personalized technical ear training and the influence of an individually optimized training sequence. *Acoust. Sci. Technol.* **2013**, *34*, 424–431. [CrossRef]
30. Kim, S. An assessment of individualized technical ear training for audio production. *J. Acoust. Soc. Am.* **2015**, *138*, EL110–EL113. [CrossRef] [PubMed]

31. Kim, S.; Imamura, H. An Assessment of a Spatial ear Training Program for Perceived Auditory Source Width. *J. Acoust. Soc. Am.* **2017**, *142*, EL201–EL204. [CrossRef] [PubMed]
32. Kim, S.; Olive, S.E. Assessing influence of a Headphone Type on Individualized ear Training. In Proceedings of the Audio Engineering Society 138th International Convention, AES, Warsaw, Poland, 7–10 May 2015.
33. Kim, S.; Bakker, R.; Okumura, H.; Ikeda, M. A cross-cultural comparison of preferred spectral balances for headphone-reproduced music. *Acoust. Sci. Technol.* **2017**, *38*, 272–273. [CrossRef]
34. Kim, S.; Choi, I.; Schwalje, A.T. Sound Localization Training Using Augmented Reality. In Proceedings of the Association for Research in Otolaryngology Midwinter Meeting 2019, ARO, Baltimore, MD, USA, 9–13 February 2019.
35. Bakker, R.; Ikeda, M.; Kim, S. Performance and Response: A framework to discuss the quality of audio systems. In Proceedings of the AES 137th International Convention E-Brif., AES, Los Angeles, CA, USA, 9–12 October 2014.
36. Letowski, T. Sound quality assessment: Concepts and criteria. In Proceedings of the Audio Engineering Society 87th International Convention, New York, NY, USA, 18–21 October 1989; Preprint 2825.
37. Corey, J. *Audio Production and Critical Listening*, 2nd ed.; Routledge: New York, NY, USA, 2016.
38. Parseihian, G.; Katz, B.F. Rapid head-related transfer function adaptation using a virtual auditory environment. *J. Acoust. Soc. Am.* **2012**, *131*, 2948–2957. [CrossRef]
39. Chon, S.H.; Kim, S. The Matter of Attention and Motivation—Understanding Unexpected Results from Auditory Localization Training using Augmented Reality. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019.
40. Cowan, N. Working Memory Underpins Cognitive Development, Learning, and Education. *Educ. Psychol. Rev.* **2014**, *26*, 197–223. [CrossRef]
41. Iliadou, V.; Moschopoulos, N.; Eleftheriadou, A.; Nimatoudis, I. Over-diagnosis of cognitive deficits in psychiatric patients may be the result of not controlling for hearing sensitivity and auditory processing. *Psychiatry Clin. Neurosci.* **2018**, *72*, 742. [CrossRef]
42. Bolia, R.S.; Nelson, W.T.; Ericson, M.A.; Simpson, B.D. A speech corpus for multitalker communications research. *JASA* **2000**, *107*, 1065–1066. [CrossRef]
43. Carlile, S.; Corkhill, C. Selective spatial attention modulates bottom-up informational masking of speech. *Sci. Rep.* **2015**, *5*, 8662. [CrossRef]
44. Dirks, D.D.; Wilson, R.H. The effect of spatially separated sound sources on speech intelligibility. *J. Speech Hear. Res.* **1969**, *12*, 5–38. [CrossRef]
45. Chon, S.H.; Kim, S. Does Technical ear Training Also Improve Speech-In-Noise Identification? In Proceedings of the 6th Conference of the Asia-Pacific Society for the Cognitive Sciences of Music, APSCOM, Kyoto, Japan, 25–27 August 2017.
46. Bisogno, A.; Scarpa, A.; Girolamo, S.D.; Luca, P.D.; Cassandro, C.; Viola, P.; Ricciardiello, F.; Greco, A.; Vincentiis, M.D.; Ralli, M.; et al. Hearing Loss and Cognitive Impairment: Epidemiology, Common Pathophysiological Findings, and Treatment Considerations. *Life* **2021**, *11*, 1102. [CrossRef] [PubMed]
47. Morais, A.A.; Rocha-Muniz, C.N.; Schochat, E. Efficacy of Auditory Training in Elderly Subjects. *Front. Aging Neurosci.* **2015**, *7*, 78. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.