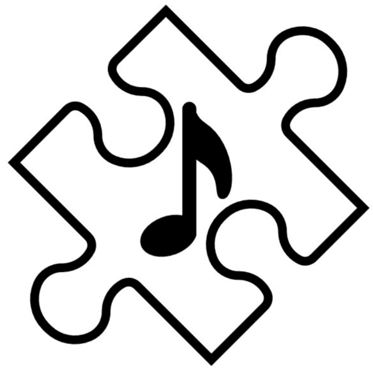
TOWARDS A THEORY OF AUDIO-BASED PUZZLES IN VIDEO GAMES



NICHOLAS SÉGUIN

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctor of Philosophy, Interdisciplinary Research in Music

School of Music
Faculty of Arts
University of Ottawa

Abstract

Audio-Based Puzzles (ABPs) in video games are challenges that require players to rely on auditory perception or knowledge of sound to progress. They may incorporate music, environmental sounds, speech, or visual representations of audio. Studying these experiences that sit at the intersection of problem-solving and auditory perception has implications spanning music cognition, education, and accessibility. Yet ABPs remain understudied, and current literature on the subject is lacking both consistent terminology and theoretical grounding.

In this thesis, I first introduce the Audio-Based Puzzle Mechanics (ABPM) framework, a novel tool for the analysis and design of ABPs based on the gameplay mechanics they employ. I detail the framework's development process, discuss its current limitations, and outline directions for future refinements.

I then report the results of an experimental study in which musicians and non-musicians played two puzzle games: one audio-based and one visual. An analysis of players' performance revealed that musicians solved ABPs faster and with more accuracy than non-musicians, but that no group differences emerged on the visual puzzle game. Participants' subjective experiences of difficulty and enjoyment are also examined, alongside a detailed analysis of gameplay segments associated with progress breakdowns and frustration felt by the players. Based on these results, I offer recommendations for the design of ABPs that can be enjoyed by a wide audience, regardless of prior musical expertise.

Finally, the thesis considers the potential of ABPs as educational tools and as a novel, playful medium for engaging with music. I present the concept of *puzzlification* to describe ABPs' unique player-game-sound configuration, which creates distinct listening experiences and positions the player as a co-author of the game's soundtrack.

Overall, this thesis provides a comprehensive account of ABPs, from existing practices to future possibilities, and establishes a theoretical foundation for interdisciplinary research at the intersection of musicology, psychology, and game studies.

Keywords: Audio-Based Puzzles, Video games, Game design, Auditory perception, Music cognition, Puzzle games, Gameplay mechanics, Music education

Résumé

Les « puzzles » (casse-têtes) basés sur l'audio (ABPs) dans les jeux vidéo sont des défis qui exigent des joueurs qu'ils s'appuient sur leur perception auditive ou leur connaissance des sons pour progresser. Ils peuvent intégrer de la musique, des sons d'ambiance, des voix ou des représentations visuelles de l'audio. L'étude de ces expériences, situées à la croisée de la résolution de problèmes et de la perception auditive, a des implications qui touchent à la fois la cognition musicale, l'éducation et l'accessibilité. Néanmoins, les ABPs demeurent peu étudiés, et la documentation actuelle sur le sujet manque à la fois d'une terminologie cohérente et d'un fondement théorique.

Dans cette thèse, je présente d'abord le cadre « Audio-Based Puzzle Mechanics » (ABPM), un outil novateur pour l'analyse et la conception des ABPs basé sur les mécanismes de jeu qu'ils utilisent. Je détaille le processus de développement de ce cadre, j'aborde ses limites actuelles et j'esquisse des pistes pour de futures améliorations.

Je présente ensuite les résultats d'une étude expérimentale dans laquelle des musiciens et des non-musiciens ont joué à deux jeux de puzzle : l'un basé sur l'audio et l'autre sur le visuel. Une analyse de la performance des joueurs a révélé que les musiciens complétaient les ABP plus rapidement et avec plus de précision que les non-musiciens, mais qu'aucune différence entre les groupes n'apparaissait pour le jeu de puzzle visuel. Les expériences subjectives des participants en matière de difficulté et de divertissement sont également examinées, parallèlement à une analyse détaillée des segments de jeu associés aux blocages dans la progression et à la frustration ressentie par les joueurs. Sur la base de ces résultats, je propose des recommandations pour la conception d'ABPs pouvant être appréciés par un large public, quel que soit son expertise musicale préalable.

Enfin, cette thèse examine le potentiel des ABP en tant qu'outils pédagogiques et en tant que moyen de participation ludique et novateur pour des activités musicales. Je présente le concept de « puzzlification » pour décrire la configuration unique joueur-jeu-son des ABPs, qui crée des expériences d'écoute distinctes et positionne le joueur comme co-auteur de la bande-son du jeu.

Dans l'ensemble, cette thèse offre un compte-rendu complet des ABPs, des pratiques existantes aux possibilités futures, et établit une base théorique pour la recherche interdisciplinaire à la croisée de la musicologie, de la psychologie, et des études sur les jeux vidéo.

Remerciements/Acknowledgements

Cette thèse est le fruit de plusieurs années de travail, et même si j'ai pris un immense plaisir à la rédiger, elle n'aurait pas pu voir le jour sans le précieux soutien de nombreuses personnes.

Je tiens tout d'abord à remercier ma directrice de thèse, Julie Pedneault-Deslauriers, pour m'avoir aidé à naviguer à travers toutes les différentes étapes de ce doctorat. Merci pour les commentaires toujours pertinents sur ma thèse, mes articles, mes présentations et mes demandes de subvention. Merci de m'avoir appuyé tout au long de ce parcours et pour ton intérêt sincère envers ma réussite.

Je remercie également les membres de mon comité consultatif de thèse, Dave Miranda et Geneviève Bazinet, pour leurs commentaires très utiles ainsi que pour les discussions stimulantes qui m'ont poussé à me dépasser. Je remercie tout particulièrement Professeur Miranda pour son aide précieuse concernant les chapitres 4 et 5.

I would like to thank Scott Smallwood for taking an interest in my research and for giving me a platform to showcase my work by inviting me to present at the University of Alberta's Sound Studies Institute. It was particularly rewarding to see what people were able to create with my framework during a game jam.

Je souhaite aussi remercier Christopher Moore, mon directeur de thèse par intérim, qui m'a accompagné durant les premières étapes de ce doctorat, notamment lors de mes examens de synthèse et de la rédaction de mon projet de thèse.

Merci au Conseil de recherche en sciences humaines du Canada de m'avoir octroyé une bourse doctorale, qui m'a permis de poursuivre ma passion pour la recherche.

Enfin, je remercie du fond du cœur ma famille, en particulier mes parents Anne-Marie et Mario, pour leur soutien et leurs encouragements, non seulement pendant ce doctorat, mais tout au long de ma vie. Merci énormément à mes proches (Brenden, Émilie, et tous les autres), qui ont su m'apporter de la joie et qui ont toujours été à mes côtés. Et merci à mon chat qui me met toujours de bonne humeur (et bien sûr pour ses précieux commentaires sur l'état de ma thèse).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABP	—	Audio-Based Puzzle
ABPM	—	Audio-Based Puzzle Mechanics
AR	—	Augmented Reality
ARG	—	Alternate Reality Game
AUC	—	Area Under the Curve
CI	—	Confidence Interval
CORGIS	—	Challenge Originating from Recent Gameplay Interaction Scale
GLMM	—	Generalized Linear Mixed Effects Model
GUR	—	Games User Research
I/O	—	Input/Output
IRR	—	Incidence Rate Ratio
ISPR	—	Integrated System of Participation in Research
LMM	—	Linear Mixed Effects Model
M	—	Musician
MDA	—	Mechanics, Dynamics, Aesthetics
NM	—	Non-musician
NPC	—	Non-Player Character
PROMS	—	Profile of Music Perception Skills
ROC	—	Receiver Operating Characteristic
TTRPG	—	Tabletop Role-Playing Game
UXR	—	User Experience Research
VR	—	Virtual Reality

1. Sorting the Pieces: An Introduction

Imagine navigating a maze you cannot see. The walls are there, but the only way to find them is to listen. Each wall emits its own noise, growing louder as you approach and fading as you pull away. As you move through the space, sound becomes your only map. This is the premise of *Maze Sounds* (AMIDEA Games, 2016), one of many games that belong to a growing family of experiences known as Audio-Based Puzzles (ABPs). Puzzles are an integral aspect of many video games; they come in various forms and challenge the player's problem-solving skills in a ludic manner. Some puzzles feature audio in a prominent way, either by making game sound a pivotal component of the puzzle-solving process, or by theming the puzzle around sound or music. These puzzles ask players to engage with sound not as atmosphere or accompaniment, but as the very essence of a problem to be solved.

Studying these ABPs can offer many relevant insights into the relationship between player and game audio. ABPs also hold real promise as tools for music education and play an important role in making games accessible to a wider audience. Despite their presence across decades of games and their potential as objects of study, ABPs have attracted surprisingly little scholarly attention. Building on previous literature, this thesis undertakes two major research endeavours: an analysis of ABPs from 1990 to 2025 using grounded theory, and an experimental study of musicians and non-musicians playing puzzle games. The results of both these projects contribute to the creation of a foundational theory of ABPs. With new and original terminology, analytical methods, and frameworks, the goal is to provide researchers, designers, and educators with a more consistent and common ground to build from.

Research on ABPs is scarce, and as a medium that combines audio, problem-solving, and entertainment, they are highly interdisciplinary in nature. As such, the next section provides an overview of the most important literature relevant to ABPs that serve as a basis for this thesis's goals: research on audio games, on auditory cognition and transfer of musical expertise, on game design and player experience, and on game sound. Following this overview, the next sections will then work to establish the definitions of "puzzle" and "Audio-Based Puzzle" that will be used throughout this thesis. Each of the thesis's subsequent chapters will contain a more focused literature review contextualized to its specific contents.

Literature Review

Audio Games

Current literature on ABPs stems mostly from research on audio games. Audio games are electronic games that either primarily or exclusively utilize sound-based interfaces to construct the game world and facilitate gameplay. As a result, players must rely on their auditory perception to interpret the information necessary for navigating and interacting within the game environment (Rovithis, Floros, et al., 2014).

The history of audio games dates back several decades. Kirke (2018) identifies *Touch Me* (Atari Inc., 1974) as one of the earliest examples of an audio game. In this game, players were asked to memorize a sequence of tones and replicate them using buttons that lit up when playing each sound. A few years later, *Simon* (Baer & Morrison, 1978), a game based on the same principle, was released and attained widespread popularity (Knoblauch, 2016). As a simple and accessible auditory memory task, the gameplay model introduced by *Touch Me* and *Simon* has inspired numerous iterations in subsequent games.

In their early stages, audio games often took the form of text-based adventures adapted for use with text-to-speech interfaces, enabling a purely auditory mode of play. Other titles emerged as sound-only adaptations of well-known visual games, such as *Audio Space Invaders* (McCrindle & Symons, 2000) and *AudioQuake* (Atkinson et al., 2006). Over time, audio games evolved into more complex and ambitious projects, with some reaching mainstream audiences. Notable examples include *Papa Sangre* and its sequel, *Papa Sangre II* (Somethin' Else, 2010, 2013), horror-themed exploration games that rely entirely on spatialized audio cues to guide players through a dark environment. Its availability on mobile platforms made it broadly accessible and contributed to its popularity.

Due to their reliance on auditory rather than visual feedback, audio games offer an important avenue for making gaming accessible to individuals who are blind or with low vision. Accordingly, a substantial body of research has focused on designing audio games with the dual aims of accessibility and broad appeal (e.g., Baldan et al., 2013; Brieger, 2015; Carvalho et al., 2012; Ekman, 2007; Mitruț et al., 2015; Östblad et al., 2014). In some cases, these games are also used to simulate aspects of sight loss. For instance, the virtual reality game *Loud and Clear*, developed by Baas et al. (2019), requires players to interact with a series of objects positioned around a 3D virtual environment. Because the game lacks any visual interface, players must rely entirely on sonic information to identify objects and navigate the space, thereby illustrating the auditory abilities developed by blind individuals.

Given the centrality of sound in audio games, it is unsurprising that many of the puzzles found within them are, by necessity, audio-based. In fact, these games offer a wide array of examples that illustrate the variety and potential of ABPs. For example, in *TapBeats*, a mobile game developed by Kim & Ricaurte (2011), players are presented with sequences of percussive sounds (i.e., hi-hat, snare, cymbals, and kick drum) which they must memorize and reproduce by tapping corresponding buttons on their touchscreen. This mechanic draws clear inspiration from *Touch Me* and *Simon*, adapting the format for modern mobile devices. Another example is *Soundoku* (Aline et al., 2021), in which players must arrange musical pitches along interconnected grids so that they progress from lowest to highest pitch. The puzzle design is based on *Blendoku* (Lonely Few, 2013), a visual game in which players organize colours into smooth gradients. *Soundoku* thus offers a sonic analogue, demonstrating how visual puzzle mechanics can be translated into auditory formats.

Despite these innovations, audio games have also faced criticism. Some have noted that development efforts often prioritize technological novelty or accessibility over core elements such as game mechanics and narrative design, factors that are essential to sustained engagement and replayability (Allain et al., 2015; Rovithis, Mniestris, et al., 2014). This imbalance can result in games that, while functional and inclusive, may fail to captivate players over the long term. While not all audio games suffer from these shortcomings, the issue suggests a need for more robust design frameworks.

Auditory Cognition

ABPS are, at their core, problem-solving activities centred around sound. As such, research on auditory cognition provides an important foundation for understanding the cognitive demands they place on players. In particular, the body of literature comparing musicians and non-musicians on auditory and other cognitive abilities offers a useful jumping-off point for further exploration.

There is substantial evidence that musical training is associated with enhanced auditory perception. Musicians consistently outperform non-musicians on tasks involving pitch discrimination and memory, tonal processing, and the detection of subtle acoustic changes, with these advantages observed across age groups and methodological designs (Bianchi et al., 2016; Bidelman et al., 2013; Micheyl et al., 2006; Tervaniemi et al., 2005). There is also evidence to suggest these advantages are not limited to the auditory domain; some studies have found that musicians perform better than non-musicians on measures of visuospatial abilities including spatial memory, mental rotation, and spatial visualization (Hetland, 2000; Pietsch & Jansen, 2012; Sluming et al., 2007). Other studies have observed similar effects with regards to broader cognitive abilities, such as nonverbal reasoning and working memory (Anaya et al., 2017; Schellenberg, 2004; Winston et al., 2022). However, the transfer of musical expertise to other domains remains a debated topic, with other researchers reporting null findings or highlighting methodological flaws in studies that have observed positive effects (Cohen et al., 2011; Forgeard et al., 2008; Sala & Gobet, 2017a, 2017b; Schellenberg & Lima, 2024; Schellenberg & Moreno, 2010)

Despite this large body of research, only one study to date has directly examined the relationship between musical habits and performance on an ABP task, finding no link between the two (Hansen & Hirga, 2017). Thus, it remains largely unclear how musical expertise affects the overall experience of solving an ABP. This gap is significant: ABPs present a unique cognitive context in which auditory discrimination abilities, musical knowledge, and problem-solving skills converge. Whether the perceptual advantages associated with musical training affect players' subjective gameplay experiences and translate into measurable performance differences on ABPs remains an open question that this thesis takes up directly.

Game Design and Player Experience

Beyond their cognitive demands, ABPs are entertaining activities embedded within the broader context of video game play. Research on game design and player experience therefore offers an equally important foundation for understanding how players engage with ABPs and what makes those experiences enjoyable, challenging, or frustrating.

Two psychological frameworks have frequently been used when theorizing optimal game design. Csikszentmihalyi's (1990) theory of flow proposes that activities perceived as too easy induce boredom, while those perceived as too difficult generate anxiety or frustration. When task difficulty aligns appropriately with a person's skill level, they become fully absorbed in the activity. Applied to game design, this suggests that well-designed games should calibrate challenge to player skill, typically increasing in complexity as players develop competencies through continued play (Falstein, 2005; Juul et al., 2009). Self-Determination Theory (Deci & Ryan, 1985; Ryan & Deci, 2000) offers a complementary perspective, conceptualizing motivation along a continuum from amotivation to fully intrinsic motivation. The theory proposes that the satisfaction of three fundamental psychological needs (autonomy, competence, and relatedness) supports sustained engagement and well-being. In the context of video games, Ryan et al.

(2006) demonstrated that psychological needs satisfaction predicted enjoyment, with autonomy and competence emerging as particularly critical in single-player experiences.

More specific insights into the structure of engaging puzzle gameplay are provided by research on learning curves and player progression. Linehan et al. (2014) examined the pacing and difficulty curve of four commercially successful puzzle games, identifying a common structure in which new gameplay mechanics are introduced through simplified challenges before complexity is progressively increased. This pattern allows players to familiarize themselves with new mechanics in easy environments before applying them to harder challenges. Related work on gameplay breakdowns and breakthroughs (i.e., moments where players become stuck or suddenly grasp a solution) has shed further light on players' experiences of puzzle-solving (Iacovides et al., 2015). Of particular methodological relevance to the present thesis, Visani Scozzi et al. (2017) developed a seven-step mixed-methods approach specifically created to evaluate game design and identify significant gameplay obstacles, which is adopted here to analyze player experience in an ABP game.

Overall, very few studies have examined the players' experience when solving ABPs, and fewer still have done so in a way that accounts for differences in musical expertise. How the interaction between auditory challenge and player skill shapes engagement, enjoyment, and frustration in ABP contexts remains largely unexplored, which this thesis seeks to address.

Game Audio

As experiences that revolve around audio, the way players interact with sound in ABPs has implications that extend beyond cognition and into the realm of musical experiences. Research on game audio and player-sound interaction therefore provides an additional important foundation for this thesis.

In the majority of video games, sound is a fundamental component of game design, contributing to immersion, narrative, emotional tone, and the communication of gameplay information. While game audio is sometimes treated as a relatively passive element, influential works in game audio studies and ludomusicology have foregrounded the interactive and participatory dimensions of sound in games (Collins, 2008; Collins, 2013; Summers, 2016). Within the landscape of interactive game audio, music video games represent a distinct genre in which musical activity, such as performing or reproducing music, forms the core of gameplay. Existing scholarship has tended to organize these games in two broad categories: rule-governed, goal-directed play (e.g., rhythm games in which players replicate musical patterns), and freeform, exploratory play (e.g., instrument games that allow for open-ended musical experimentation) (Pilchmair & Kayali, 2007; Kassabian & Jarman, 2016). These classifications, while useful, have consistently excluded music-based puzzle games from their scope, treating them as distinct from music video games.

ABPs engage players with sound and music in ways that are neither about performing/recreating works nor freely exploring. As a result, musical experiences emerging from the act of solving ABPs remain undertheorized. Specifically, this thesis argues that ABPs’ player-audio configuration gives rise to listening experiences and modes of musical participation that differ meaningfully from those found in other music video games, and as such warrant a proper investigation.

Having now highlighted the key research and gaps in the literature regarding ABPs, the next sections proceed to defining puzzles and ABPs to further delineate the scope of this thesis.

What is a Puzzle?

A comprehensive study of ABPs first requires a definition of what is meant by the term "puzzle." At first glance, this might seem like a straightforward question, but it quickly becomes clear that the concept is multifaceted and somewhat ambiguous. While it is tempting to rely on one's intuition ("I know a puzzle when I see one!"), this investigation requires a more rigorous and theoretically grounded definition. Although many elements of the following discussion apply to both analog and digital puzzles, this thesis will focus exclusively on puzzles encountered within video games (including mobile games). Nevertheless, insights from analog puzzles remain relevant and will be drawn upon where useful.

One of the most iconic examples of a puzzle is the jigsaw puzzle: images cut into small, oddly shaped pieces that must be reassembled. While this is only one kind of puzzle, it introduces several important characteristics that recur in broader definitions. First, a puzzle involves a "problem" and a "solver" and thus is fundamentally an interaction between an individual and a structured challenge. Second, the challenge should be non-trivial; puzzles take time and effort to complete. Third, and crucially, puzzles are engaged with voluntarily, usually for entertainment, highlighting their status as a form of play. While this lays the groundwork for a more formal definition of the concept of a puzzle, ambiguity persists. Many forms of play also involve challenging tasks, so what, then, distinguishes puzzles from them?

Scott Kim (2004) attempts to solve this quandary by offering the following succinct definition: "1. A puzzle is fun, 2. and it has a right answer" (p. 1). This definition achieves two objectives. To begin with, it identifies puzzles as a playful activity, and more importantly, it distinguishes puzzles from other forms of play: stories, toys, and games. According to Chris Crawford (1984), the goal of a game is to *win*, while the goal of a puzzle is to *find a solution*. As for toys, they do not have a fixed goal, and stories are consumed rather than interacted with. Games are defined by their conflict, in which there is a winner and a loser, while puzzles, toys, and stories lack this characteristic. Another distinguishing characteristic of puzzles is that they have little to no replay value, unlike games or toys, since once the solution to a puzzle is known, completing it again will not offer the same satisfaction. Kim goes on to explain how these four activities form a hierarchy in which each successive form of play can incorporate the preceding one: stories may be elements of toys, toys may be components of puzzles, and puzzles may be integrated into games.

Notably, however, this hierarchy is unidirectional; a game can incorporate puzzles, but a puzzle by itself cannot be a game. This distinction is important to point out, as terminology from the realm of video games can be somewhat confusing regarding puzzles. The "puzzle game" *genre* is defined by games in which the main activity is puzzle-solving and typically consists of multiple puzzles that become increasingly more complex as the game progresses. Instances of puzzles, however, appear across a wide variety of genres as standalone tasks. This fluidity reflects the broader notion that while puzzles can exist within games, games encompass a broader range of experiences and should not be conflated with puzzles. Adding further complexity, the term puzzle is occasionally used colloquially in game culture to describe many kinds of gameplay challenge, even those not centred on problem-solving. For example, players might refer to difficult platforming sections as "jumping puzzles," or segments requiring precise timing as "timing puzzles." This usage of the word puzzle should be understood more as a synonym for "challenge" than the problem-solving activities that are the focus of this thesis.

While Crawford and Kim's categorization of games, puzzles, toys and stories can help classify most activities, the boundaries between these categories are not always clear. Kim, for example, cites *Solitaire* as a "borderline case between game and puzzle" (Kim, 2004, p. 4) and mentions how the solution to a puzzle might not always be set in stone, such as in poll-based questions where the right answer depends on what everyone else answers. These ambiguous cases are reflected in other similar discussions. For instance, Mark J. P. Wolf, in *The Medium of the Video Game* (2001), defines puzzles as requiring the player to figure out a solution through activities like navigation, tool usage, and object manipulation. This particular definition excludes quizzes and mazes, a point challenged by Mora-Cantallops (2018), who asks: "if a puzzle involves solving an enigma, is not a question the most basic form of such? If a puzzle can involve navigation to figure out a solution, why is a maze not a puzzle?" (p. 3). Framing puzzles solely as problems that require solutions also blurs the distinction between puzzles and games more generally. Under this perspective, puzzles could be considered components of nearly all games, as Costikyan (2002) argues:

> Almost every game has some degree of puzzle-solving; even a pure military strategy game requires players to, e.g., solve the puzzle of making an optimum attack at this point with these units. In fact, if a game involves any kind of decision making, or trade-offs between different kinds of resources, people will

> treat these as "puzzle elements," trying to devise optimal solutions. […] You can't extract puzzle from game entirely. (p. 10)

Taking this argument to the extreme, Mora-Cantallops (2018) states: "one could argue that all video games are essentially puzzles, and would not be far from the truth. Because, from a computing perspective, what is a video game but a puzzle whose solution is a chain of inputs that complete the game?" (p. 2). In short, constructing a conceptual framework that clearly delineates puzzles as distinct from games remains a persistent theoretical challenge.

Mora-Cantallops ultimately proposes that puzzle games be defined by their reliance on problem-solving: "they require players to concentrate, to think their moves, to move towards solving a challenge that was accepted by them. A puzzle is, in the end, a problem with a non-obvious solution" (p. 6). Fernández-Vara (2014) echoes this emphasis on cognition, stating that puzzles provide "a challenge that has one solution and requires thinking rather than skills" (p. 233). Again, this definition is not without its complications. Certain puzzles may accept multiple solutions, revealing the diverse strategies individuals use to approach and resolve problems. Moreover, the notion of skill is elusive as well. Puzzles engage the player's abilities in various ways, be it via memory, pattern recognition or spatial reasoning. This raises a critical question: how can we meaningfully distinguish between cognitive abilities that constitute "thinking" and those that are better characterized as "skills"?

A notion that can help us refine our definition of puzzle is the distinction between static and dynamic systems. According to Crawford, a static system does not react to the player's actions, while a dynamic system will respond in some way to the player's decisions:

> Compare playing a cube puzzle with playing a game of tic tac toe. Compare the sport of high jumping with the game of basketball. In each comparison the two activities provide similar challenges to the player. The key difference that makes one activity a game and the other activity not a game is the interactive element. A cube puzzle does not actively respond to the human's moves; a high jump pole does not react to the jumper's efforts. In both tic tac toe and basketball the opposing players acknowledge and respond to the player's actions. (Crawford, 1984, p. 11)

Karhulahti (2013) further develops this distinction between static and dynamic systems by framing the two concepts in terms of the consequences they produce. In static systems, consequences are determinate, meaning that an action always results in the same outcome. For example, a jigsaw puzzle has determinate outcomes, for placing piece A in spot B always leads to the same result: either the piece fits or it doesn't. Conversely, in dynamic systems, consequences are indeterminate. Chess is a game with indeterminate consequences; while moving piece A to spot B always leads to the same board configuration, it is not possible to predict how the opponent will, in turn, modify the state of the board. Thus, Karhulathi argues that puzzles are static challenge structures, while games are dynamic challenge structures.

Karhulathi also introduces the concept of *kinesthetic challenges*. A kinesthetic challenge involves a nontrivial amount of psychomotor effort, meaning the challenge requires some level of physical and mental coordination. In contrast, a non-kinesthetic challenge only involves mental effort. According to Karhulathi, puzzles are inherently *non-kinesthetic*, though puzzles often co-occur with kinesthetic challenges. This is made possible because of the distinction between solving and executing the solution of a puzzle, which Karhulathi envisions as two different activities. A player can figure out the solution to a puzzle, meaning they understand the various steps necessary to achieve the solution, and yet be incapable of executing these steps themselves. The reverse can also be true, where a player can have difficulty finding the solution to a puzzle, and therefore rely on external help such as online guides, but then execute the solution themselves. In this scenario, the player has not solved the puzzle but has successfully performed the solution. Thus, in Karhulathi's view, "There are no kinesthetic puzzles but only puzzles with kinesthetic executions" (Karhulahti, 2013, p. 10). Returning to Fernández-Vara's definition, we could propose that the "skills" excluded from puzzle-solving primarily refer to kinesthetic challenges, which involve psychomotor coordination rather than cognitive reasoning.

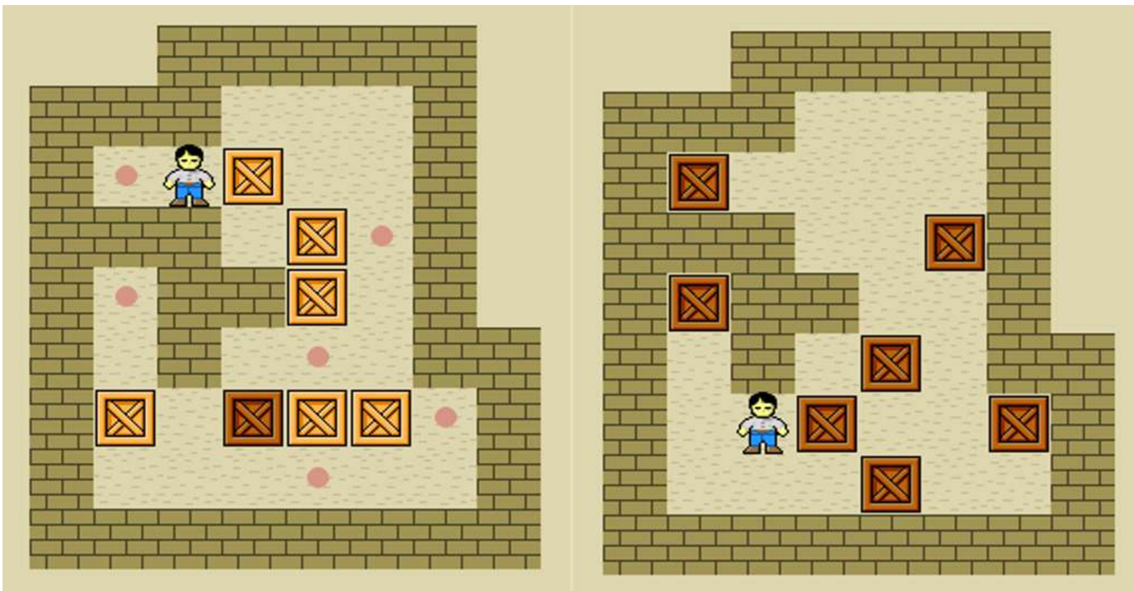
Therefore, for the purpose of this thesis, puzzles are defined as challenges that:

1. Emphasize problem-solving as their core activity;
2. Aim toward finding a solution rather than achieving victory;
3. Operate within static systems that yield determinate consequences;
4. Demand nontrivial cognitive engagement, rather than psychomotor skill.

To illustrate this working definition in the field of digital puzzles, we can turn to some well-known examples. *Sokoban* (Thinking Rabbit, 1982) is a video game in which players must push crates onto designated goal tiles within a confined, grid-based environment. The player can only push (not pull) crates, and the central challenge lies in determining a correct sequence of moves that achieves the goal without rendering the puzzle unsolvable (Figure 1.1). Each level presents a discrete problem to be solved, rather than a contest to be won, and any given move reliably produces the same outcome, highlighting the static, deterministic nature of the system. Furthermore, success depends on spatial reasoning and planning, rather than on reaction time or manual dexterity. The enduring popularity of Sokoban has led to numerous sequels and inspired a broader category of puzzles, commonly referred to as “Sokoban-style,” characterized by similar mechanics of object manipulation that require careful planning within constrained spaces.

Figure 1.1

Beginning and end states of a Sokoban *puzzle.*



By contrast, *Tetris* (Pajitnov, 1985) is frequently labelled a puzzle game, yet it does not fully conform to the criteria outlined above. While it shares some puzzle-like qualities, most notably spatial reasoning and a lack of direct competition, it also introduces elements that do not fit with our third and fourth criteria. The randomized sequence of falling tetrominoes is unpredictable, requiring players to make rapid judgments under time pressure rather than work toward a predefined solution (Karhulahti, 2013). In addition, as the game progresses and pieces fall more quickly, the emphasis on kinesthetic challenges intensifies, marking a shift from reflective problem-solving to reactive execution. A useful point of comparison can be found in *The Talos Principle* (Croteam, 2014), which features puzzles that resemble *Tetris* in visual form but adhere more closely to puzzle logic. In these segments, players are presented with a fixed set of tetromino-like shapes and must fit them into a confined area (Figure 1.2). The challenge lies in discovering a single or limited number of correct solutions, with no random elements or time pressure to undermine the puzzle's static and cognitive nature.

Having established the working definition of puzzle, we may now proceed to defining audio-based puzzles as a distinct category of puzzles in games, identify their core characteristics, and differentiate them from other audio experiences.

Figure 1.2

Beginning and end states of a tetromino puzzle from The Talos Principle.

What is an Audio-Based Puzzle?

Over time, multiple terms have been used to describe puzzles that revolve around sound, including "music puzzles," "sound puzzles," and "audio puzzles." This thesis adopts the term "Audio-Based Puzzle" (ABP) to refer to these experiences for several reasons. First, in contrast to terms like "music puzzle," which imply a focus solely on musical content, "audio-based" is intentionally broader, encompassing puzzles that involve not only music but also other forms of sonic information, such as noises and visual depictions of sound. Second, the term "audio-based" has previously been employed to distinguish between audio games which are purely audio ("Audio-Only Games") and those that include visuals that play a supportive role in order to enable gameplay ("Audio-Based Games") (Rovithis, Mniestris, et al., 2014). This distinction is particularly relevant here, as the ABPs considered in this thesis span both purely auditory and audio-visual contexts. Finally, the term "Audio-Based Puzzle" has been used previously to refer to puzzle experiences centred on sound, such as in Carvalho et al. (2012), Hansen & Hiraga (2017), and Östblad et al. (2014), although its usage does not reflect a formally defined term.

While most puzzles in video games include sound, whether in the form of a soundtrack, sound effects, or feedback, this ubiquity raises the question: when does the use of sound in a puzzle elevate it to the status of an ABP? I propose that a puzzle qualifies as an ABP when sound plays an active, rather than merely supportive, role in its solution. In most games, sound serves to establish atmosphere, signal success or failure, or enhance emotional impact. Such uses do not convey the information necessary to solve a puzzle and are thus insufficient for categorization as ABPs. For example, an error tone that informs the player they have entered an incorrect answer does not constitute meaningful puzzle-solving information; it merely responds to an action already taken. Thus, in addition to the four core criteria for puzzles established earlier, ABPs can be defined as challenges that:

5. Depend on the perception, production, or manipulation of sound, or formal representations thereof, as a necessary condition for progress or solution.

Consider the following examples. In a segment of *Jak II* (Naughty Dog, 2003), players are presented with a set of tombstones that emit short melodies when struck. The objective is to identify matching pairs of tombstones based solely on auditory information. In this case, the only way to find the matching pairs is to listen to the melodies (bar trying every possible combination until successful). Another example appears in *Nine Sols* (Red Candle Games, 2024), where players find four bells that each produce a distinct pitch. To unlock a secret, they must reproduce a four-note melody played earlier in the game by another character. Again, sound plays a pivotal role in solving the puzzle; the only way to know the order in which to play the bells is to memorize the melody, associate the four pitches of the melody with the pitches produced by the bells, then recreate the melody.

Importantly, the proposed definition does not require that puzzle-solving information be presented *exclusively* via sound. Rather, ABPs may incorporate both visual and auditory cues. For instance, the game *Aquaria* (Bit Blot, LLC, 2007) features a puzzle nearly identical to that in *Nine Sols*, where the player must replicate a melody found at another location in the game to proceed. However, in *Aquaria*, the melody is represented visually using symbols that correspond to in-game pitches the player can produce by making their avatar sing (Figure 1.3). While the premise of the puzzle is the same, the player can solve the puzzle using only these visual cues, without relying on auditory information. In this latter case, sound is not *necessary* to solve the puzzle. At the same time, the presence of sound means one *could* rely on sonic information, entirely or partially, to solve the puzzle.

Figure 1.3

Singing puzzle from Aquaria. *Players encounter a set of four symbols (left). They must then reproduce this set at another location in the game by making the avatar sing the pitches associated with each symbol (right).*

Finally, this brings us to the concept of "formal representations of sound." Some puzzles are thematically and functionally centred on sound but do not deliver information sonically. Instead, they may use representations such as musical notation or descriptive text. Because interpreting these representations requires an understanding of their sonic referents, such puzzles should, in my opinion, still be considered ABPs. Take, for example, the "water harp" puzzle from *Remnant 2* (Gunfire Games, 2023), in which players must input a melody based on notated symbols rather than an auditory sequence. Again, just as in *Nine Sols* and *Aquaria*, the player is tasked with replicating a melody found elsewhere in the game. However, in this particular puzzle, the player only has to contribute the sequence of notes in their notated form, with the melody being played *after* the correct input is provided. At no point during the puzzle-solving process has the player had to listen to any sounds, yet the symbolic information presented refers explicitly to sound and musical structure. For this reason, these puzzles are included in the proposed definition of ABPs.

Of course, the cognitive demands of ABPs vary significantly depending on how sonic information is presented and what players are tasked to do with it. Puzzles that rely solely on sound will differ in player experience and difficulty from those that present information through a mixture of modalities. While this thesis proposes an inclusive definition that encompasses these variations, they can still be differentiated when analyzing the unique properties and perceptual challenges associated with different types of ABPs. These distinctions will be explored further in Chapter 2, which introduces a taxonomy of audio-based puzzles.

Why Study Audio-Based Puzzles?

Designing audio games that are both accessible and engaging remains an open challenge, which more focused research into ABPs can help address. As entertaining challenges that inherently lend themselves to non-visual design, ABPs offer fertile ground for developing gameplay that is both accessible and compelling. Yet their potential extends beyond game design; ABPs sit at a unique intersection of problem-solving and auditory perception, and studying them can shed light on how people perceive, process, and interact with sound in goal-directed contexts. This has meaningful implications for music cognition research and for music education, as ABPs offer a structured yet playful way to engage with musical concepts.

The chapters that follow are organized around two major research projects. The first is a survey of a sample of ABPs from 1990 to 2025, undertaken to establish a common analytical framework for the field. In Chapter 2, I introduce the Audio-Based Puzzle Mechanics (ABPM) framework, a taxonomy of ABP mechanics developed to assist in the analysis and design of such puzzles. In Chapter 3, I then address the framework's limitations and present broader findings from the survey, tracing the evolution of ABPs across three decades.

The second project is an experimental study comparing the experiences of musicians and non-musicians when playing an ABP. Chapter 4 details the study's mixed-method protocol. In Chapter 5, I then report the quantitative results, examining differences between groups in terms of performance, enjoyment, and perceived difficulty. In Chapter 6, I turn to the players' subjective experiences, using qualitative data to analyze the design tensions inherent to ABPs and offering recommendations for more inclusive game design.

Finally, Chapter 7 steps back to consider the broader implications of both projects. I introduce the concept of *puzzlification* to theorize the relationship between player, sound, and game in the context of ABPs. I also speculate on what that relationship might mean for the future of music education. It is my hope that, taken together, these chapters will advance our understanding of ABPs while offering practical tools and a shared foundation from which future research can grow.

2. A Framework for the Design and Analysis of Audio-Based Puzzles

Puzzles constitute a core component of many video games and are a main staple of the adventure and puzzle genres, though individual puzzles also appear across a wide range of other genres. Among these, some puzzles are centred specifically on audio, whether in the form of music, sound, or noise. Such puzzles are referred to as Audio-Based Puzzles (ABPs) and are defined here as challenges that:

1. Emphasize problem-solving as their core activity;
2. Aim toward finding a solution rather than achieving victory;
3. Operate within static systems that yield determinate consequences;
4. Demand nontrivial cognitive engagement, rather than psychomotor skill, and;
5. Depend on the perception, production, or manipulation of sound, or formal representations thereof, as a necessary condition for progress or solution.[1]

As an example, in each level of the maze game *Hear Your Path* (Solo Philosopher, 2022), players are presented with a piece of music and must navigate a grid by playing one note at a time to replicate the piece. At each step, several notes are offered, and selecting an incorrect one leads the player down the wrong path. The game's objective is to reach the exit with as few mistakes as possible, and the game features works by notable classical composers such as Beethoven and Chopin. It thus meets all the criteria of an ABP: identifying the correct note serves as the core problem-solving activity; the goal is to reproduce the piece, not to defeat an opponent; the game lacks

[1] For more information on how this definition was established, refer to Chapter 1.

randomness and high psychomotor demands; and, crucially, progress depends on attentive listening.

While puzzles have been examined in game studies and design literature (e.g., Adams, 2009), these classifications typically address puzzles in a more general fashion (e.g., logic, lateral thinking, memory, etc.) without elaborating upon the distinct properties of audio-based puzzle-solving. ABPs differ from other puzzle types in that their core informational medium is sound, often requiring auditory discrimination or musical knowledge. These characteristics introduce unique challenges and design considerations, making a dedicated taxonomy for ABPs both relevant and necessary. Nevertheless, scholarly literature on the subject remains sparse, and no comprehensive overview or systematic categorization currently exists, a gap which this chapter seeks to address.

Among the few works addressing the properties of ABPs, Rovithis's (2012) classification of audio-based games stands out. The classification divides audio games into 11 categories, one of which being puzzle games. While Rovithis's taxonomy is focused on audio games and not games in general, it offers valuable insight into common themes in the realm of ABPs. In particular, Rovithis identifies four recurring puzzle types in audio games: "the repetition of a pattern ('Beat Master'), discovering hidden sounds ('Did you hear that?'), placing sonic fragments in the right order ('Audio Puzzle') and matching similar sounds in pairs ('Hear Hear')" (Rovithis, 2012, p. 161). Additionally, the category includes applications that "focus on a specific property, and make riddles out of it, teaching players to recognize pitch, intervals, scales and chords" (p. 161).

Outside of academic sources, the website *TV Tropes*, which aims to describe and collect examples of popular narrative devices or conventions in various forms of media, provides further observations on common ABP concepts. Relevant to this paper, instances of ABPs appear in the categories "Songs in the Key of Lock"[2] (doors or other obstacles are unlocked by music), "Simon Says Mini-Game"[3] (iterations on the memory game *Simon* [Baer & Morrison, 1978]), and "Musical Trigger"[4] (events initiated by playing music).

While both Rovithis's work and *TV Tropes* offer a great starting point for a categorization model of ABPs, their classifications are centred on themes rather than on the underlying mechanics that govern puzzle operation. For instance, overlap can be observed in some of Rovithis's labels: in puzzles from the *Beat Master*, *Audio Puzzle* and *Hear Hear* categories, players are given an audio input (a pattern, a song, or a sound) and must then select or produce a corresponding output (the same pattern, the complete song, or a matching sound). The differences between listening to a series of sounds and selecting the same pattern from multiple options, and finding matching pairs of sounds, are minimal; both involve listening to a reference sound and identifying its identical counterpart.

A similar issue arises in *TV Tropes'* thematic groupings, which do not meaningfully distinguish between puzzles with different operational structures. Consider the following two puzzles from "Songs in the Key of Lock." In *Resident Evil* (Capcom,

[2] Available at https://tvtropes.org/pmwiki/pmwiki.php/Main/SongsInTheKeyOfLock. Accessed August 20th 2024
[3] Available at https://tvtropes.org/pmwiki/pmwiki.php/Main/SimonSaysMiniGame. Accessed August 20th 2024
[4] Available at https://tvtropes.org/pmwiki/pmwiki.php/Main/MusicalTrigger. Accessed August 20th 2024

1996), a secret door opens when the player character performs the first few measures of Beethoven's *Moonlight Sonata* on a nearby piano. To do so, the player must first look around to find and pick up the musical score and then navigate the menu and select the score while next to the piano, which prompts an automatic performance. By contrast, in *The Elder Scrolls III: Morrowind* (Bethesda Game Studios, 2002), the player encounters stalactites that emit a four-note sequence, which must be replicated by striking stalagmites that produce pitched notes. In the first example, interaction is minimal and does not challenge the player's auditory perception, while in the second, success depends on accurately perceiving and reproducing a sequence of pitches. While the puzzles' themes of unlocking a passage using music align, the tasks demand entirely different skills from the player.

These examples underscore the need for a more precise framework. In this chapter, I introduce the **Audio-Based Puzzle Mechanics** (ABPM) framework, which focuses on the mechanics of ABPs rather than on their themes. Mechanics are specific rules, actions, or systems within a game that structure player interaction with puzzle elements, defining what the player must do to achieve the puzzle's objective. They form the fundamental building blocks of puzzles and thus shape the player's problem-solving approach and overall experience. By classifying ABPs according to their mechanics, the ABPM framework enables clearer distinctions between puzzle types and the specific skills they require. The framework was constructed through an examination of existing ABPs found in commercial and independent games, taking inspiration from similar endeavours in the realm of analog puzzles (Hufkens & Browne, 2019; Krekhov et al., 2021) and video game mechanics (Toups et al., 2014; Wuertz et al., 2018). A constant comparative approach was adopted to identify and group recurring mechanics into distinct categories. The resulting model provides a robust conceptual basis for future research and serves as a practical tool for examining the design and reception of ABPs.

Method

To develop this framework, I utilized a constant comparative method on puzzles from a corpus of 148 puzzles from 114 games.[5] This method is adapted from grounded theory (Glaser & Strauss, 1967) and is inductive, meaning that general conclusions are reached based on a body of specific observations. It is divided into three activities: coding, memoing, and integrating (Bernard, 2017, pp. 225–232). Coding involves carefully going through the data and assigning codes to every chunk of data—in this case, puzzle gameplay. As codes are assigned to more and more data, related concepts are grouped into larger categories. Memoing comprises writing field notes about identified concepts throughout the research procedure. These notes consist of thoughts, observations, insights, and hypotheses regarding the concepts and how they might be related to one another. Memoing acts as an intermediate step between the initial coding and the complete analysis. The final aspect of the constant comparative method, integrating, consists of creating conceptual models linking the coded ideas. These models are generated early in the research process; then, subsequent data is held up to the model to see if it conforms to it or not. This method continuously refines the initial model until it reaches its final iteration (Charmaz, 2000).

The selection of ABPs for the development of this framework was conducted through a search designed to capture a diverse sample. Initial searches were conducted

[5] For a complete list of games surveyed in this study, see Appendix A.

during the summer of 2024, with additional instances of puzzles incorporated subsequently as they were encountered. Search activities included:

- General web searches, using keyword combinations such as “puzzle” or “riddle” paired with “music,” “audio,” or “sound”;
- Exploration of online game distribution platforms, such as *Steam* and *Epic Games*, using both keyword searches and category browsing;
- Consultation of other online resources, such as *TV Tropes* and *AudioGames.net*;
- Peer recommendations obtained from colleagues and online communities familiar with audio-focused game design.

Each instance was assessed against the five defining criteria of ABPs described in the introduction either through direct play of the relevant game section or, where direct access was impractical, by reviewing gameplay footage from online sources. Only puzzles meeting all the definitional criteria were retained. Musical challenges that relied primarily on psychomotor skill (e.g., rhythm games) were excluded. The final sample encompasses ABPs dating from 1990 to 2025 and from a wide variety of platforms. This diversity ensures that the resulting framework addresses a broad spectrum of gaming experiences, thereby reducing the likelihood that important categories or mechanics are overlooked. However, although the search strategy was meant to be comprehensive, it is not exhaustive. Some ABPs, particularly instances embedded within otherwise non-puzzle games, have likely been missed. Additionally, the use of recorded gameplay footage for certain cases, while necessary for accessibility, may have limited experiential understanding compared to direct play. Nevertheless, I believe the resulting framework encompasses most of the gameplay experiences represented in the current landscape of ABPs.

The ABPM Framework

The Audio-Based Puzzle Mechanics (ABPM) framework is depicted in Figure 2.1. There are five categories of puzzle mechanics observed in ABPs: **Input/Output**, **Spatial**, **Environmental**, **Contextual**, and **Epistemic**. The first two categories are further divided into subcategories. The Input/Output category includes **Match** and **Transform**, while the Spatial category includes **Positive** and **Negative** forms. For each of these categories and subcategories, there are two possible versions: **Direct** and **Indirect**. The next section outlines the important distinction between direct and indirect mechanics, followed by descriptions and examples of each category. A summary is presented in Table 2.1.

Figure 2.1

Diagram of the Audio-Based Puzzle Mechanics Framework.

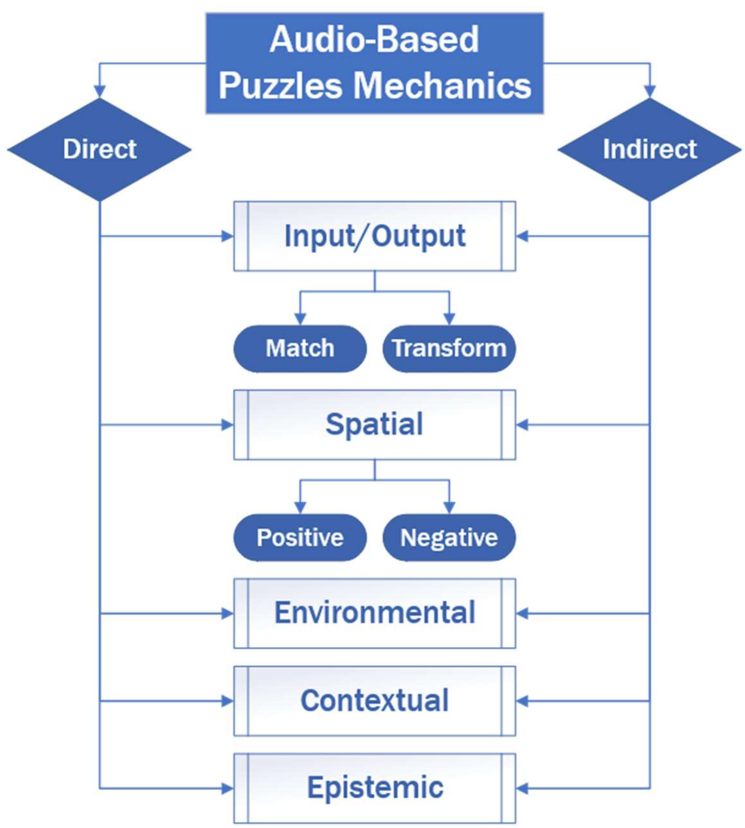


Table 2.1

Summary of each category of puzzle mechanics.

		Direct	Indirect
Input/Output	Match	The player is presented with an audio input and must select or produce an identical audio output	The player is presented with a non-audio input and must select or produce an identical non-audio output
	Transform	The player must transform an audio input into a corresponding non-audio output, or a non-audio input into a corresponding audio output	The player must transform a non-audio input into a different non-audio output
Spatial	Positive	The player's avatar must travel towards a sound source that can only be located via audio	The player's avatar must travel towards a sound source that is represented by a modality other than audio
	Negative	The player's avatar must travel away from a sound source that can only be located via audio	The player's avatar must travel away from a sound source that is represented by a modality other than audio
	Environmental	The player must manipulate the virtual environment using a property of sound	The player must manipulate the virtual environment using a property of sound that is represented by a modality other than audio
	Contextual	The player triggers an event by producing or interacting with sound	The player triggers an event by interacting with an object that produces sound as a consequence of the interaction or is themed around sound
	Epistemic	The answer to a question asked by the game can be answered by listening to the game's audio	The answer to a question asked by the game is related to audio but does not require listening to the game's audio to answer

Direct and Indirect

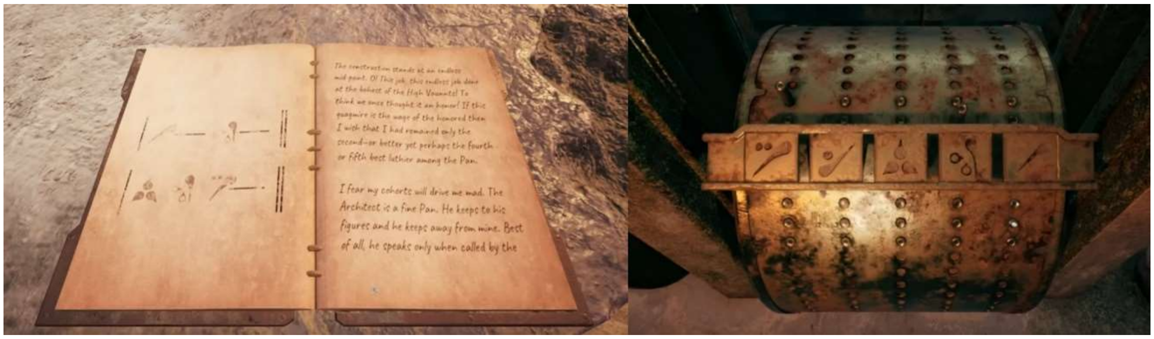
As outlined above, there are two versions of each category in the ABPM framework: direct and indirect. In direct puzzle mechanics, the player must engage with audio in some capacity, meaning auditory skills are essential for solving the puzzle. In an indirect mechanic, sound may still be present, but the puzzle can be solved without relying on audio.

A clear example of both forms can be found in the "water harp" puzzle from the game *Remnant 2* (Gunfire Games, 2023), which was briefly mentioned in the previous chapter. In one instance of the puzzle, the player finds a journal that contains a fictional notation system of five unique symbols, each corresponding to a specific peg on a large cylindrical music box (Figure 2.2). A dash in the notation indicates a pause, meaning all the pegs should be pressed down to produce silence. The notation is read from left to right, top to bottom. In this case, the player simply replicates the sequence of symbols in the journal on the music box by raising the corresponding pegs before activating the harp, which then rotates and plays the melody. Because the solution is derived entirely from a visual reference, this version of the puzzle is *indirect*, as it could be completed with the game's audio turned off.

In contrast, a bonus version of the same puzzle requires the player to listen to a melody played by a nearby Non-Player Character (NPC) on the flute and replicate it on the water harp. Here, no visual notation is provided; the player must identify the correct pitches by ear, adjusting the pegs until the produced melody matches the one they hear. This second version of the puzzle is therefore *direct*, as it requires the use of auditory skills to recreate the melody.

Figure 2.2

"Water harp" puzzle from Remnant 2. *On the left, a journal shows the fictional notation of a melody. On the right, a music box with symbols corresponding to the notation has pegs which can be adjusted to produce the melody.*



Although it may seem counterintuitive to include indirect puzzle mechanics in a framework about ABPs, they remain relevant because they are often incorporated in ABPs and co-occur with direct mechanics. In many instances, knowledge about sound or music can contribute to the player's puzzle-solving process. Furthermore, auditory elements can be used to frame a puzzle within the game's narrative, even if these elements are not necessary for the puzzle's completion. For instance, in *The Legend of Zelda: Ocarina of Time* (Nintendo, 1998), many of the game's puzzles revolve around the character's titular ocarina, asking the player to perform songs or repeat melodic sequences to progress through some areas. These puzzles include an indirect puzzle mechanic, as the required notes are visually represented on screen by symbols corresponding to buttons on the controller. Therefore, while audio is unnecessary, it is undeniable that solving these ocarina puzzles constitutes a musical experience for the player and deserves to be a part of discussions of ABPs. In all, for the reasons mentioned above, indirect puzzle mechanics are relevant to the framework and are included in it.

Input/Output Puzzle Mechanics

The first category of ABPMs is **Input/Output** (I/O). In these mechanics, players are given an input and must then select or produce a corresponding output. Inputs and outputs may take diverse forms, including audio, musical notation, instrument interaction, or text. I/O mechanics can be subdivided into **match** and **transform** varieties. In match puzzle mechanics, the player produces an output in the *same* form as the input (e.g., audio → audio, notation → notation). In transform puzzle mechanics, the output modality differs from the input (e.g., audio → notation, text → instrument). Although straightforward, this distinction is analytically useful for differentiating ABPMs and the skills they recruit, which in turn shapes player experience. For instance, transforming an audio input into notation is typically easier for players with formal music-reading proficiency. I/O mechanics are particularly rich because of the many possible input-output configurations. The following sections illustrate each subcategory and highlight the breadth of this design space.

Direct Match

A **direct match** puzzle mechanic presents an audio input and requires the player to select or produce the corresponding audio output. One common variant of this mechanic is found in puzzles that play a short sound or melody and ask the player to identify the identical sound among several options. The tombstone puzzle from *Jak II* (Naughty Dog, 2003) is a prime example of this mechanic: the player enters a room containing twelve tombstones, each of which plays a short melody when struck, and progress requires identifying the six matching pairs.

While it is the most obvious example, direct match mechanics need not be limited to identical pitch sequences. *The Lost Garden* (Smallwood et al., 2021) features intercoms that emit specific sounds (e.g., instruments, bird calls, industrial noises, etc.). Elsewhere in the area, matching consoles can be tuned with a dial to produce a range of similar sounds, only some of which correspond to the intercom sounds. The task is to find the sets of intercoms and consoles which have the same sound and adjust each console to match the intercom's sound. Rather than recognizing exact pitches or melodic contours, players must instead discriminate between timbral qualities, which can be more challenging than pitch matching. Taking this even further, in one specific instance, players are tasked with matching the sounds of footsteps. Rather than locating a literal reproduction of the original sound, players must instead identify the consoles that contain recordings of footsteps in other environments (e.g., on snow, on concrete, etc.). In this case, the goal is to identify the source action of the sounds, rather than the sound itself.

Another prevalent direct-match variant scrambles a melody for the player to reconstruct, much like assembling a jigsaw puzzle but with sound. *Circuits* (Digital Tentacle, 2014) exemplifies this approach: players receive segments of an original track and must restore the intended sequence. As levels progress, difficulty scales by increasing the number of fragments, adding decoys, and minimizing acoustic distinctiveness between pieces. Similar designs have been explored in several research studies to date (e.g., Carvalho, Guerreiro, et al., 2012; Diakopoulos et al., 2008; Hansen et al., 2012).

Direct Transform

A **direct transform** puzzle mechanic requires transforming an audio input into a corresponding non-audio output, or vice versa. For example, in the memory game *TapBeats* (Kim & Ricaurte, 2011), which takes inspiration from *Simon*, the player hears one of four drum sounds—hi-hat, snare, cymbal, or kick drum—and must then select the corresponding instrument name. This process is repeated over and over in a growing sequence of sounds that the player must memorize until they make a mistake, at which point the game is over. *Undertale* (Fox, 2015) offers a similar challenge. In one area of the game, an ambient melody can be heard; if players backtrack to an earlier piano and reproduce the melody's opening notes using five buttons mapped to specific pitches, a secret room is revealed. No explicit notation is provided, and as such, success depends on attentive listening and recall. In both of these examples, an audio input is transformed into a corresponding non-audio output; the instrument's name in *TapBeats*, and the sequence of buttons in *Undertale*.

Direct transform mechanics can also leverage player voice and asymmetric cooperation. In the co-operative game for two players *We Were Here Forever* (Total Mayhem Games, 2022), one player pilots a submarine and attempts to communicate with an aquatic creature, the Kraken, by imitating its guttural calls using buttons that produce imitations of these sounds. However, to figure out which buttons need to be pressed, they must communicate with the second player, who has access to a translation device. The first player must therefore convey the sounds verbally or by imitation, while the second identifies them using the translation device. In this instance, for the second player, the input consists of another player's voice, which showcases the innovative potential of such puzzle mechanics in multiplayer games.

Indirect Match

An **indirect match** puzzle mechanic presents a non-audio input and requires the player to select or produce a corresponding non-audio output. The canonical example of this mechanic is the gameplay loop pioneered by *Touch Me* (Atari Inc., 1974) and *Simon*, where an ever-lengthening sequence of coloured buttons light up and play corresponding tones, which players have to reproduce by pressing the same buttons. Because the input (buttons lighting up) corresponds to the output (buttons being pressed), and the tones are redundant with the visual cues, this puzzle can be solved even without audio. Several iterations on this mechanic can be observed in games that have come out since, including *The 7th Guest* (Trilobyte, 1993), *Warframe* (Digital Extremes, 2013), *Dream Chronicles: The Chosen Child* (KatGames, 2009), and *Hollow Knight: Silksong* (Team Cherry, 2025).

Another instance of an indirect match mechanic appears in *Figment* (Bedtime Digital Games, 2017). In one segment of the game, players must step on specific keys of a giant keyboard as indicated by a head-on diagram of the instrument with certain keys depressed (Figure 2.3). Although the keys produce sounds when pressed, the audio feedback is extraneous; success depends solely on reproducing the depicted configuration. Because both the input and the output are instrument-based (keyboard → keyboard), this constitutes a match mechanic.

Figure 2.3

Keyboard puzzle from Figment. *The player encounters a diagram of a keyboard with certain keys pressed down (left). The player steps on the same keys on a giant keyboard (right).*

Indirect Transform

An **indirect transform** puzzle mechanic presents players with a non-audio input that must be transformed into a non-audio output in a different form. In the second chapter of *Bendy and the Ink Machine* (Joey Drew Studios, 2017), players find a note that contains a riddle that names four instruments. To progress, the player must visit another room, which contains multiple instruments, and simply click on those four instruments in the specified order, thereby transforming a textual input into an output consisting of a sequence of simple interactions.

Spatial Puzzle Mechanics

Spatial puzzles exploit the fact that sounds have spatial origins, and that perceived loudness can vary with proximity. Often, these puzzles rely on audio being presented binaurally, although that is not always the case. They can be divided into **positive** and **negative** subcategories. In positive designs, the objective is to move *towards* a sound source, while in negative designs, the objective is to move *away* from a sound source. Movement may be performed directly by the player (e.g., in virtual reality) or indirectly through an on-screen avatar. In some instances, the goal can be to reposition an object that the player does not control directly but instead steers using other means.

Direct Positive

A **direct positive** puzzle mechanic requires navigating toward an audio source that can be localized only by listening. *The Lost Garden* offers several such puzzles. In one, stepping onto a central platform turns off the lights, making the screen go completely dark. A sound begins to play elsewhere in the now-unlit room, and the player must locate it by ear to restore the lights.

Subtler direct positive designs use audio gradients to guide players through mazes. Near the end of *Prince of Persia: The Sands of Time* (Ubisoft Montreal, 2003), players enter a circular room with multiple doors. Exiting through one door re-enters the same space from another door, effectively trapping the character. The correct route out of the maze is hinted by faint water-splashing sounds that grow louder near the appropriate door. *The Legend of Zelda*: *Ocarina of Time* implements a similar structure in the *Lost Woods* area, where players navigate a series of rooms with three exits. The path out of the maze is indicated by the soundtrack's ocarina being amplified when near the correct exit.

Direct Negative

A **direct negative** puzzle mechanic requires navigating away from an audio source localized solely through listening. *Maze Sounds* (AMIDEA Games, 2016) tasks its players with navigating mazes with invisible walls that emit noise. Moving closer to the walls will make these noises grow louder, allowing the players to avoid them, as coming in contact with a wall results in a game over. Direct negative mechanics are comparatively scarce in commercial titles, likely because they invert a common design intuition (i.e., to approach salient sounds). Where they do appear, these mechanics invite game designers and players to treat quietness as progress feedback, rather than the more familiar pull of an attractive sound source.

Indirect Positive

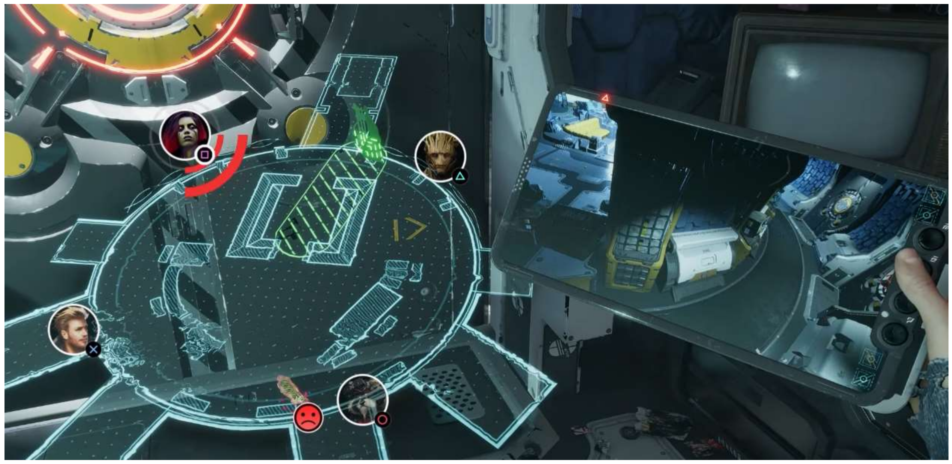
An **indirect positive** puzzle mechanic represents a sound source via a modality other than audio while still directing the player towards it. In *Outer Wilds* (Mobius Digital, 2020), players are given a tool called a signalscope. This tool allows them to tune to specific frequencies and visualize the direction of the sound's origin via two semicircles that converge into a circle when aimed correctly (Figure 2.4). The device is especially useful when players are tasked with locating a missing character, as they can follow the sound of a harmonica using both auditory feedback and the converging visual indicator to navigate the maze-like environment in which the character is trapped.

Figure 2.4

The signalscope tool from Outer Wilds. *Top: the signalscope is pointing away from the source. Bottom: the signalscope is pointing directly at the source; the two semicircles have converged into a green circle.*

Figure 2.5

Llama-herding puzzle in Marvel's Guardians of the Galaxy. *Red lines near Gamora (top-left) indicate the llama (bottom, with red emoticon) is moving away from her voice.*



Indirect Negative

An **indirect negative** puzzle mechanic, players or objects move away from a non-audio representation of sound. In *Marvel's Guardians of the Galaxy* (Eidos-Montréal, 2021), players must herd a llama in a circular room by activating speakers situated at the corners. Each speaker plays the sound of a character singing; two characters' singing voices attract the llama, while the other two repel it. When those latter speakers are active, red curved lines radiate outward, visually signalling repulsion and allowing the player to steer the animal accordingly (Figure 2.5).

Environmental Puzzle Mechanics

Environmental puzzle mechanics require manipulating the virtual environment using a property of sound: pitch, time, loudness, or timbre. These mechanics feature a direct mapping between an acoustic parameter and an environmental variable. Direct environmental mechanics map audio itself to environmental change, whereas indirect mechanics achieve change through instruments or devices whose sounds are incidental rather than directly causal.

Direct Environmental

In **direct environmental** puzzle mechanics, players manipulate aspects of the virtual world by modulating properties of sound. Puzzles in *One Hand Clapping* (Bad Dream Games, 2021) serve as great examples of these mechanics, as players complete challenges using their own voice, with the game translating live vocal input into continuous environmental change. In several areas, pitch is mapped to vertical position, such that higher pitches raise platforms or increase the height of a line as it is being drawn horizontally. The mapping is continuous rather than discrete, which encourages players to gain finer control over their voice to move objects with more accuracy and solve puzzles more efficiently.

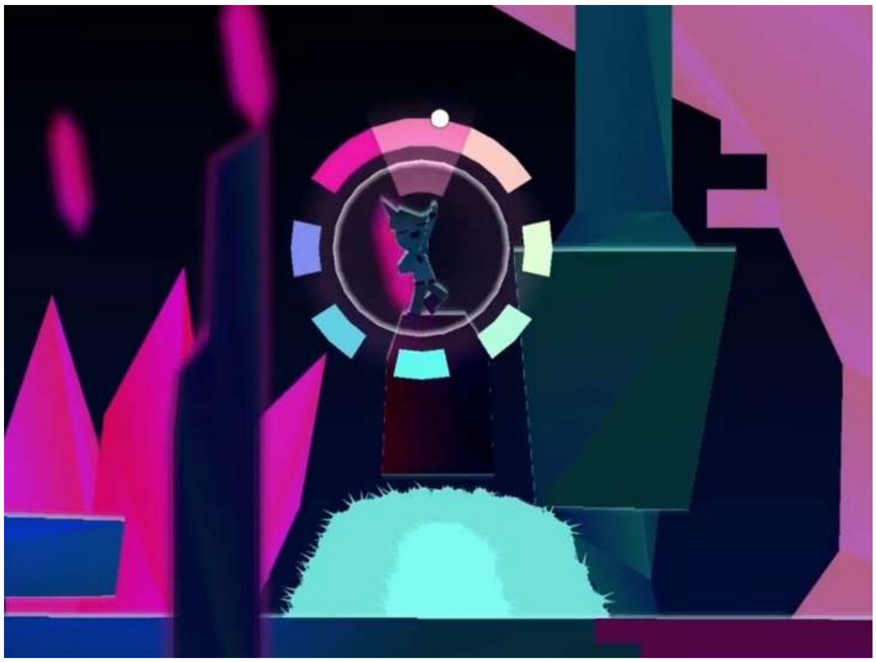
Indirect Environmental

Indirect environmental puzzle mechanics retain the mapping between acoustic properties and environmental change but mediate it through a representational layer. In *Sonority* (Hanging Gardens Interactive, 2022), players move and rotate platforms by playing sequences of pitches on an instrument. Notes appear as letter names (A-G) or numbers (1–7), and the intervallic distance between successive notes determines the magnitude of translation or rotation. Ascending intervals move objects upward or clockwise, while descending intervals move them downward or counterclockwise.

In *Wandersong* (Wishes Ultd. et al., 2018) and *Deltarune* (Fox, 2025), instruments similarly mediate directional control. *Wandersong* maps the protagonist's singing voice to a radial "direction wheel" surrounding the avatar. Objects move along the direction indicated by the currently sung pitch (Figure 2.6). Similarly, in one section of *Deltarune*, a piano maps distinct pitches to the four cardinal directions. Striking the key associated with a given pitch displaces an object in the corresponding direction.

Figure 2.6

A player moves a platform upwards in Wandersong *by singing the pitch associated with "up" on the direction wheel.*



Contextual Puzzle Mechanics

Contextual puzzle mechanics are those in which contextual cues signal that a sonic event needs to occur, whether by interacting with an acoustic object or performing a piece of music. While some contextual mechanics resemble environmental ones, the distinction is that contextual designs rely on the presence or invocation of sound to trigger game events, rather than on a direct mapping between a property of sound and an environmental object (e.g., pitch ↔ height).

Direct Contextual

Direct contextual puzzle mechanics require the player to use audio to trigger an event. For instance, in *Phasmophobia* (Kinetic Games, 2020), players can communicate with ghosts via a real-world microphone. By asking appropriate questions, the players can elicit responses that support identification of the entity, which is the game's main problem-solving activity. Similarly, *In Verbis Virtus* (Indomitus Games, 2015) casts the player as a wizard who speaks incantations into a microphone, which triggers spells that can affect the world and can be used to solve puzzles.

Indirect Contextual

Indirect contextual puzzle mechanics trigger events through interaction with objects that produce sound as a consequence of being used or are musically themed. The *Legend of Zelda* series offers numerous instances of such mechanics. In *Ocarina of Time*, *Majora's Mask* (Nintendo, 2000), and *The Wind Waker* (Nintendo, 2002), there are several situations where progress is hindered until players deduce that a specific song needs to be performed. Another example can be observed in *Paper Mario: The Thousand-Year Door* (Nintendo, 2004). In an early chapter of the game, the final enemy of the area is irritated by crickets. By exploring, players can acquire a piece of equipment that changes the sound of Mario's attacks to cricket chirps, substantially weakening the enemy.

Epistemic Puzzle Mechanics

Epistemic mechanics are those in which the solution is knowledge either provided by or related to audio. Although the following puzzle mechanics might not fit under some scholars' definition of puzzle, this category helps situate quizzes and trivia within the broader framework.

Direct Epistemic

Direct epistemic puzzle mechanics pose questions that are answerable by listening to the game's audio. An example appears in *Banjo-Kazooie* (Rare, 1998) in the late-game section entitled "Grunty's Furnace Fun," where players participate in a diegetic game show that includes questions about music and sounds heard earlier in the game, such as naming a level from its musical theme or recognizing a character's voice. Other designs include *Phonopath* (Regamey, 2011). In each level, players are presented with an audio file that must be manipulated (e.g., reversing, filtering, amplitude adjustment) using software of the player's choice to extract a password for progression. Finally, this puzzle mechanic can also be observed in games about real-world music. Apple's iPod mini-game *Music Quiz* (Apple, 2003) played a track from the user's library and asked for identification from multiple choices. A similar game, *SongPop Classic* (FreshPlanet, 2012), expanded this format across thousands of songs.

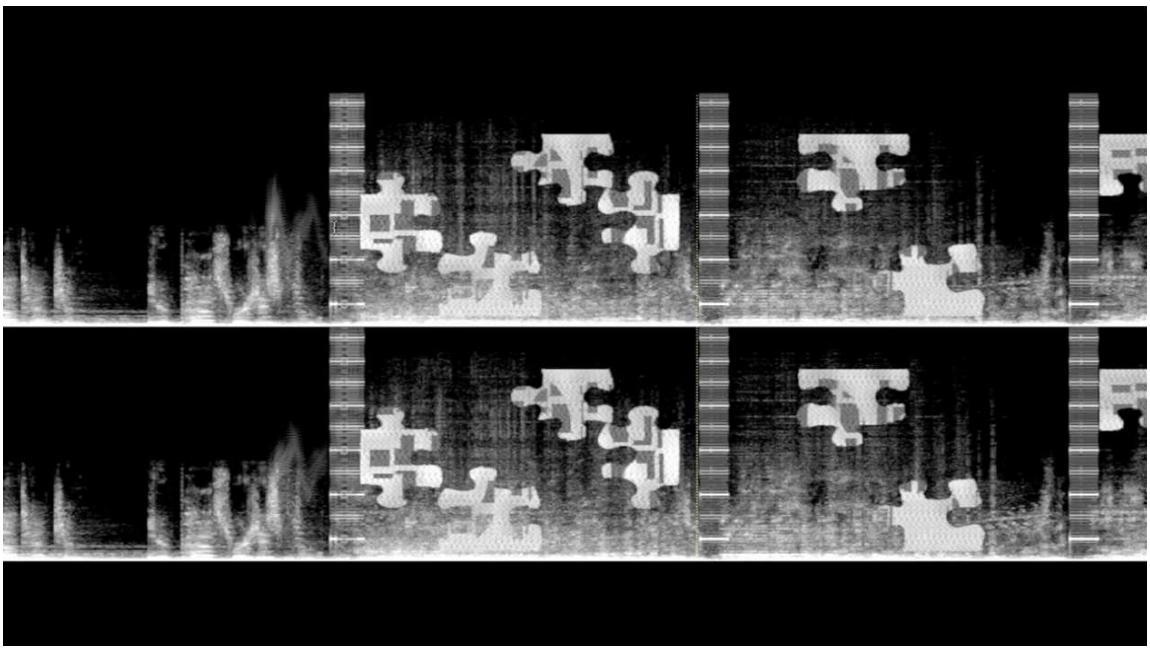
Indirect Epistemic

Indirect epistemic puzzle mechanics involve questions related to audio that do not require listening to the game's audio. Again, many titles employ this format with real-world music knowledge. Examples include completing lyrics in *Don't Forget the Lyrics* (LB ART, 2021) or *Finish the Lyrics* (Baczkowski, 2018), as well as user-generated quizzes on platforms like Sporcle (e.g., album covers, song titles). An in-game analogue of this format appears in *Final Fantasy VI* (Square, 1994), where players must select the correct continuation of lyrics for an opera after having glanced at the libretto earlier.

More unconventional instances of this mechanic can be found in *Phonopath*, where answers are occasionally hidden within specific visualization formats of audio files. For instance, in one level, players complete a jigsaw puzzle whose pieces are hidden within the spectrogram of an audio clip (Figure 2.7).

Figure 2.7

The jigsaw puzzle from Phonopath, *which can be solved by accessing the audio file's spectrogram. Screenshot from Phonopath (2014).*



Conclusion

This chapter introduced the Audio-Based Puzzle Mechanics framework, a taxonomy for organizing and describing the mechanics of Audio-Based Puzzles. The framework consists of five categories—Input/Output, Spatial, Environmental, Contextual, and Epistemic—each with direct and indirect variants that specify the mechanic's presentation and its relationship to game audio. By foregrounding mechanics rather than common themes, the framework extends prior work on ABPs and offers a tool suited to the finer analysis of player-puzzle interactions. This emphasis on puzzles' building blocks enables more precise comparisons across titles and supports systematic discussion of design choices and their effects.

There are both practical and theoretical applications of the framework. For designers, the framework can guide ideation by clarifying the available design space and highlighting underused combinations during the development process. For researchers, it offers variables that can be isolated and manipulated to study impacts on player experience (e.g., perceived difficulty, enjoyment, accessibility, cognitive load). For instance, findings such as Carvalho, Duarte, et al. (2012) on strategy differences between visual and auditory puzzles raise the opportunity for comparative studies across ABPM categories, such as whether spatial and environmental mechanics elicit distinct problem-solving strategies, or whether mechanics within the same category converge on similar approaches.

As with any taxonomy, edge cases resist neat classification, and some contemporary or future mechanics may not fit cleanly into the present scheme. These limitations, discussed further in the following chapter, should be viewed as opportunities to refine the framework whether by adjusting category definitions or adding new subtypes. Nonetheless, the ABPM framework is, in its current form, useful in its possible applications and is malleable enough to adapt to emerging design practices in the field of ABPs. I encourage designers and researchers to adopt, test, and critique the framework in practice, so that subsequent iterations can better account for accessibility considerations and the growing diversity of audio technologies.

3. The ABPM Framework: Limitations, Future Steps, and Trends

The previous chapter introduced the Audio-Based Puzzle Mechanics (ABPM) framework, a novel tool for distinguishing among the types of mechanics found in Audio-Based Puzzles (ABPs). The purpose of this chapter is to elaborate on several topics introduced earlier. First, it examines the definitional boundaries and limitations of the framework in greater depth, drawing on case studies from the initial survey of ABPs to explore and clarify the theoretical issues involved in defining what constitutes an ABP. Second, the chapter assesses potential avenues for expanding the ABPM framework, including components that were considered during development but ultimately discarded from the final version. Finally, it presents an analysis of the collected sample of puzzle mechanics, using the ABPM categories to describe and interpret general trends within the field.

Definitional Boundaries and Framework Limitations

As defined in the introduction, ABPs are challenges that:

1. Emphasize problem-solving as their core activity;
2. Aim toward finding a solution rather than achieving victory;
3. Operate within static systems that yield determinate consequences;
4. Demand nontrivial cognitive engagement, rather than psychomotor skill.
5. Depend on the perception, production, or manipulation of sound, or formal representations thereof, as a necessary condition for progress or solution.

Although this definition successfully distinguishes ABPs from other puzzles, it is not without limitations. Several examples encountered during the initial survey challenge the boundaries of the definition and reveal areas that require clarification. Four issues are of particular importance. First, the distinction between *solving* a puzzle and *executing* its solution must be reconsidered in light of certain borderline cases. Second, the process by which one distinguishes between an exercise and a puzzle needs further examination. Third, the framework must more clearly address puzzles that rely on recorded speech. Finally, one puzzle mechanic encountered during the analysis did not fit neatly within the ABPM framework, prompting reflection on how the framework might be adapted or extended.

Solving and Executing

Karhulahti (2013) argues that puzzles are non-kinesthetic challenges, in that they do not rely on significant psychomotor effort. Nonetheless, they can coincide with kinesthetic challenges due to the distinction between the cognitive process of *discovering* a solution and the physical process of *executing* it. Under this view, challenges that initially appeared too demanding in terms of psychomotor skill may be reconceptualized as involving two distinct activities: a puzzle-solving phase and a kinesthetic execution phase. An example of this can be found in *The Lost Garden* (Audio Games Lab, 2021). One puzzle requires players to listen to speakers that emit short rhythmic patterns and to reproduce these patterns by pressing a button, aided by a visual cue that moves horizontally and switches direction with each beat. Because no instructions accompany the devices, players must first solve the puzzle of inferring their purpose. Once the mechanism is understood, however, the primary challenge becomes psychomotor, as players must accurately replicate the rhythmic sequence. At this stage, the task closely resembles challenges found in rhythm music games.

This raises the question of whether such a challenge should be considered an ABP within the present framework. One could argue that the initial interpretive phase constitutes an ABP, but this portion represents only a small portion of the overall activity. For the purposes of the ABPM framework, it is likely to be more productive to consider challenges holistically. This position aligns with Costikyan (2002) and Mora-Cantallops (2018), who propose that most if not all games contain some element of puzzle-solving. Distinguishing solving from executing too rigidly risks expanding the scope of ABPs so broadly that it becomes unfruitful. Thus, the previous example from *The Lost Garden* would not be considered an ABP, as a major part of the experience requires psychomotor execution.

Puzzles and Exercises

A second limitation pertains to the relationship between puzzles and exercises. The current definition effectively differentiates puzzles from other forms of play, such as toys or games, but does not explicitly address their distinction from structured educational tasks. Many music theory exercises, such as identifying notes on a staff or determining chord qualities in a dictation, meet the definitional criteria for puzzles, even though they are not typically considered as such.

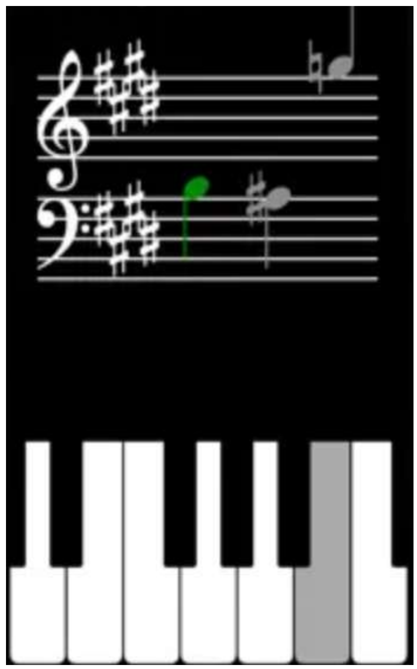
One might argue that the difference lies in intention: exercises aim to teach or improve skills, whereas puzzles aim to entertain. Yet these aims frequently overlap, particularly in serious games that combine pedagogical objectives with engaging gameplay.[6] *iClef* (Baratè & Ludovico, 2013), which requires players to identify notes on a staff under varying clefs and conditions, exemplifies this dual purpose. Similarly,

[6] A serious game is a game specifically designed with the primary objective to educate rather than to entertain.

commercial titles such as *Chords Enchanter* (XD Games Studio, 2022) and *Scale Enchanter* (XD Games Studio, 2021) operate under a similar premise. All three games rely on minimalistic designs, essentially making them digital exercises despite their presentation as games (Figure 3.1).

Figure 3.1

Screenshot of the in-game iClef *interface (source: Apple App Store).*



By contrast, *Partitas* (Lotz, 2021) implements a similar gameplay experience, yet contextualizes it within a narrative structure and provides in-game rewards for completing levels, thus more closely aligning with conventional notions of a game. The boundary between exercises and puzzles is therefore inherently unclear. Given this ambiguity, establishing a strict definitional distinction between games and exercises using similar mechanics does not appear feasible. As such, for the purposes of this framework, deferring to the creators' own marketing and framing might be more appropriate. Accordingly, the three serious games mentioned above qualify as ABP games, whereas a music theory homework assignment does not. Moreover, the creators' framing typically reflects the way players themselves approach these experiences: users engage with such titles primarily as games rather than as formal exercises, lending further justification to this approach. Lastly, a final advantage of relying on creators' framing is that it avoids excluding puzzles whose difficulty varies substantially between individuals. This is particularly relevant in the case of ABPs, which often take the form of tasks that might be trivial to musicians but may present a genuine challenge to those without musical training. While imperfect, this solution is functional and avoids generating unnecessary categorical issues.

ABPs and Speech

A further limitation of the current definition of ABPs concerns the mode of delivery of a puzzle. The definition states that ABPs depend on the perception, production, or manipulation of sound, which raises an important question: does a puzzle become an ABP simply because it is presented aurally? For instance, a riddle presented in written form does not constitute an ABP. If the same riddle is accompanied by a voiceover that reads it aloud, then an audio component has been added, muddling the waters. The issue becomes even more complex if the riddle is delivered exclusively through recorded speech. Similar ambiguities arise when puzzles that are not intrinsically audio-based are presented with audio descriptors for accessibility. In such cases, the auditory information is functionally significant for some players, yet it does not transform the underlying puzzle into one that is designed around sound.

This tension highlights a broader conceptual challenge: at what point does the inclusion of spoken language render a puzzle "audio-based" rather than merely "audio-delivered”? The question is particularly pertinent in audio-only games developed for blind and visually impaired players, where speech often serves as the primary vehicle for conveying instructions, environmental cues, and narrative information (Archambault et al., 2007; Röber & Masuch, 2004). If every puzzle presented through speech were classified as an ABP, the definition would become overinclusive and would obscure meaningful distinctions in puzzle design. A similar risk emerged in the earlier discussion on separating solving from executing a puzzle. In both cases, an overly expansive interpretation risks weakening the analytical usefulness of the framework.

A productive way of addressing this issue is offered by Rovithis et al. (2024), who investigate the potential of “speech puzzles” (“spuzzles”) that rely on recorded voice. Drawing on Chion’s (1994) three modes of listening, the authors created puzzles that challenge the players’ *semantic listening* (attending to meaning), *causal listening* (attending to the sound’s source), or *reduced listening* (attending to the sonic qualities independently of meaning). Their work demonstrates that speech can support puzzles that do not rely on semantic listening. For example, identifying a missing phoneme in a recording engages reduced listening rather than comprehension. By contrast, a riddle recited through speech remains predominantly semantic, with the auditory delivery being incidental; the puzzle could be solved equally in written form. This distinction suggests that puzzles relying on causal or reduced listening are compatible with the ABPM framework, whereas those grounded solely in semantic listening are not.

Nevertheless, the boundary is still not entirely clear-cut. Epistemic mechanics frequently rely on semantics, meaning the exclusion of semantic listening would also result in the removal of several epistemic mechanics.[7] Furthermore, even puzzles whose content is not intrinsically sonic may, when delivered exclusively through audio, require players to construct and maintain a mental representation of the puzzle using auditory cues alone. Such a cognitively demanding process aligns with the definition of an ABP, and while it falls outside the original intention of the ABPM framework, their reliance on auditory processing warrants at least some consideration. For now, one guiding principle behind the framework is that puzzles should be designed around audio, not merely delivered through it. This should not be seen as a strict exclusion criterion for anyone looking to engage with the framework, and in fact, puzzles that are aurally delivered but not inherently audio-based may still be analyzable through the ABPM categories on a case-by-case basis. Future research should explore how to integrate such puzzles within the framework.

Hidden Inputs

Soundoku (Aline et al., 2021) adapts the structure of the mobile game *Blendoku* (Lonely Few, 2013), replacing colour gradients with grids of sounds that vary systematically in pitch, loudness, timbre, or duration. Some levels require players to organize sounds simultaneously along two dimensions, such as pitch on the horizontal axis and duration on the vertical axis. At first glance, *Soundoku* appears to exemplify the Input/Output (I/O) mechanic: the game provides a set of sonic "inputs" that players must arrange into a correct "output." However, unlike other I/O puzzles, *Soundoku* does not provide an explicit reference for the player to reproduce. In typical I/O designs, the

[7] As a reminder, epistemic mechanics are those where the game asks a question, the answer to which is found in the game's audio or is related to audio. For a full definition of these mechanics, as well as the other four mechanics from the framework (Input/Output, Spatial, Environmental, and Contextual), refer to Chapter 2.

puzzle offers a clear model serving as the target against which the player compares their attempt. In *Soundoku*, this reference is absent; players must infer the correct ordering without the benefit of a reference with the correct arrangement of inputs.

One might alternatively categorize this as a contextual mechanic, insofar as the puzzle relies on the game's rules and instructions rather than a concrete reference. While this is true, contextual mechanics typically involve simple interactions or the performance of musical works taught by the game, rather than the construction of a large-scale audio structure. *Soundoku* therefore can be envisioned as a distinct variant of I/O mechanics, one that might be described as a "hidden-input" approach, in which the input is implicit rather than explicitly provided. Because *Soundoku* is currently the only identified instance of such a mechanic, it is for now classified alongside other I/O mechanics while noting its atypical structure. Should similar mechanics emerge, revisiting the categorization may be needed.

Expanding the ABPM Framework

Throughout its development, the ABPM framework underwent several iterations. While the previous chapter presented its finalized form, a number of earlier conceptual components, some implemented temporarily and others only theorized, offer insight into how the framework might be extended or refined. Although these components were ultimately excluded, they hold potential value for certain users of the framework and may inspire future additions.

Filtering by Properties of Sound

One proposed addition to the framework was an upper-level filter specifying which acoustic property or properties a puzzle employs—pitch, duration, timbre, or loudness. Such a filter could, in theory, support a more granular characterization of ABPs. First, it would offer greater descriptive precision regarding the current landscape of ABP design: which acoustic properties dominate, which are underutilized, and which combinations of mechanic type and sonic property co-occur most frequently. Second, it could enable researchers to systematically explore how different auditory properties influence player performance and experience.

However, several practical and conceptual challenges emerged. Many puzzles can be categorized straightforwardly, such as *Undertale*'s piano puzzle (Fox, 2015), which centres on pitch, or *Bendy and the Ink Machine*'s instrument puzzle (Joey Drew Studios, 2017), which relies on timbre.[8] Other puzzles, however, draw on multiple properties simultaneously. Games that mimic traditional jigsaw puzzles but with audio fragments, for example, may require attention to several acoustic properties in combination. Then, some puzzles cannot meaningfully be classified by acoustic properties at all. *Figment*'s piano diagram puzzle (Bedtime Digital Games, 2017) requires reproducing a visual layout on a keyboard; while keystrokes generate sound, the sonic output is only incidental and plays no functional role in solving the puzzle. The same applies to contextual puzzles in which players must interact with instruments or objects, such as locating sheet music and interacting with a piano in *Resident Evil* (Capcom, 1996). Finally, titles such as *Phasmophobia* (Kinetic Games, 2020) and *In Verbis Virtus* (Indomitus Games, 2015) require vocal input without regard for pitch, timbre, or any other acoustic parameter; what matters is the verbal content, not its sonic characteristics.

[8] A description of most of the puzzles mentioned in this section can be found in the previous chapter.

Because many ABPs rely on sound properties in ambiguous ways, a rigid classification system proved unworthwhile. Although such a filter could enrich certain analyses, it could not be applied consistently across all mechanics and was therefore excluded from the final framework.

Input/Output Mechanics: Clarifying Types

In the ABPM framework, Input/Output mechanics are those in which players select or produce an output that corresponds to an input given by the game. This correspondence may involve either *matching* (e.g., audio → audio, notation → notation) or *transforming* (e.g., notation → instrument, text → audio). A limitation of the current framework is that the notion of "type" remains loosely defined. Earlier iterations attempted to standardize it through discrete categories: audio, notation, instrument, text, and voice. While appealing in principle, this system introduced problems of its own, as many puzzle interactions resist neat classification. For example, in *The Legend of Zelda: Ocarina of Time* (Nintendo, 1998), players perform songs by pressing specific controller buttons, which function simultaneously as an *instrument* and as a *notation* system. The in-game notation consists of visual representations of the buttons themselves, blurring the boundary between the two categories (Figure 3.2).

A similar issue arises in *Super Mario RPG* (Square, 1996), where players recreate melodies by jumping on tadpoles that are moving up and down a staff. The inputs in two related puzzles are delivered through solfège syllables. In one instance, these are plainly stated, while in the other, they are hidden inside words. Although solfège constitutes a form of notation, its textual embedding complicates its classification. The output, meanwhile, consists of tadpoles placed along a staff in a way that resembles quarter notes. Labelling both as "notation" obscures meaningful differences in their presentation.

Given the frequency with which inputs and outputs challenge categorization, these subcategories were removed from the final framework. The practical consequence is that determining whether a mechanic belongs to the *match* or *transform* category relies on the user's interpretive judgment. While this introduces some subjectivity, identifying whether the input and output take the same general form is typically straightforward enough for applied use.

Figure 3.2

In-game screenshot from The Legend of Zelda: Ocarina of Time*, showing the notation system based on the buttons of the Nintendo 64 controller.*

A Note on Notational Subtypes

Although the "type" layer was excluded from the I/O category, there is one observation regarding three forms that notation can take within ABPs that remains theoretically valuable for analyses on notational systems, both in ABPs and in games in general.

Conventional Notation

This category comprises notational systems that correspond to established real-world musical practices, such as Western staff notation, solfège, and tablature. For example, the piano puzzle in *Resident Evil Village* (Capcom, 2021) requires players to interpret a traditionally notated melody and reproduce it on an in-game keyboard (Figure 3.3).

Fictional Notation

Fictional notation refers to symbolic systems devised specifically for a game world and meaningful only within that context. The water harp puzzle in *Remnant 2* (Gunfire Games, 2023) exemplifies this form with its symbols (Figure 2.2). *The Legend of Zelda: Ocarina of Time* adopts this form as well with its controller-based notation (Figure 3.2).

Abstract Notation

Abstract notation encompasses representational forms that depict sonic events without encoding specific musical parameters. A noteworthy example appears in *The Witness* (Thekla, Inc., 2016), where players listen to birdsongs and draw lines approximating their melodic contours (Figure 3.4). The drawings serve as representations of the acoustic events but lack the precision necessary to determine exact pitches or rhythms. Abstract notation therefore conveys the general structure of a sound but does not permit reproduction from the notation alone.

Figure 3.3

In-game screenshot of the piano puzzle from Resident Evil: Village*.*

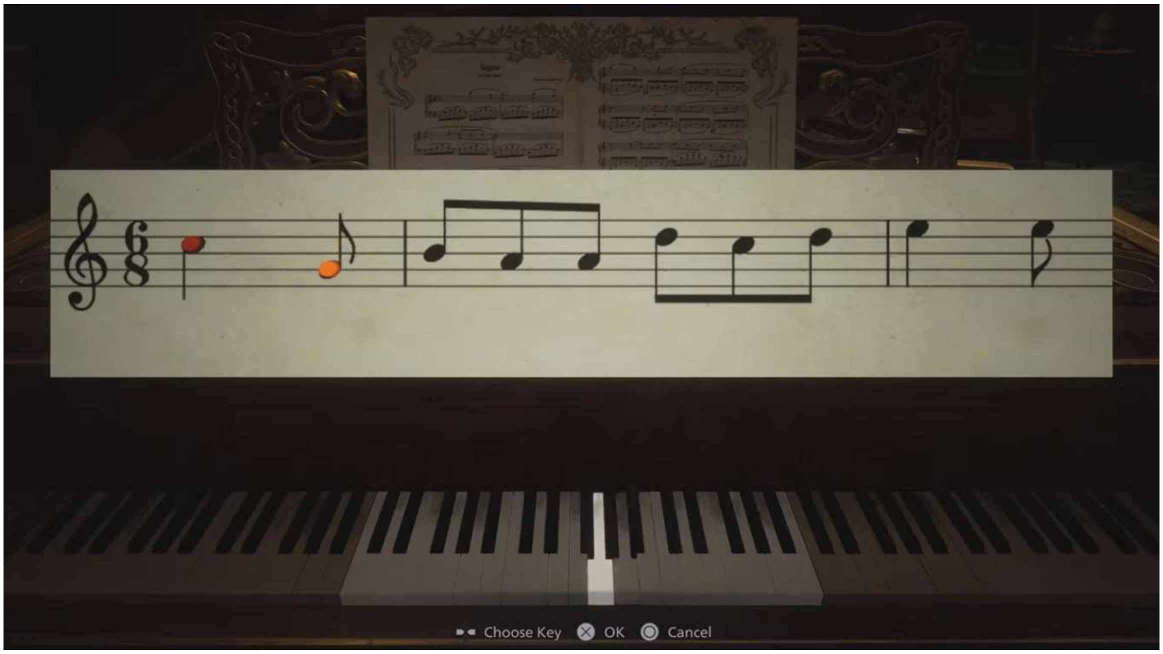


Figure 3.4

In-game screenshots from The Witness*, showcasing an instance of a birdsong puzzle in its unsolved state (left) and solved state (right). In this case, the drawn line represents a sequence of a high pitch followed by a lower pitch.*

What About Silence?

Finally, the framework must contend with puzzles centred on silence. Some games feature challenges that require players to avoid making noise or being detected by sound-sensitive enemies. These challenges often involve some amount of problem-solving, as players must determine which actions generate sound and plan routes or actions accordingly. Because such puzzles hinge on the absence of sound, it is reasonable to consider whether they warrant a distinct category within the framework. In practice, however, most silence-based puzzles can be accommodated within existing categories. Challenges involving navigation around sound sources align with spatial mechanics (notably the *negative* subcategory), whereas challenges requiring players to infer which interactions produce noise align with contextual mechanics. That being said, silence represents a unique form of sonic interaction that some might argue necessitates a distinct categorization. For the purpose of the current framework, their inclusion within existing categories feels sufficient. However, deeper research into how silence is used within puzzles could prove fruitful.

Final Note

While the ABPM framework concerned itself with video games, there is definitely room for growth, in particular to account for other multimedia works that involve audio-based problem-solving. Some instances of virtual reality (VR) games have already been included in the framework, such as *The 7th Guest VR* (Vertigo Games, 2023), but there is much to explore in the realms of Augmented Reality (AR) games, Alternate Reality Games (ARGs), tabletop role-playing games (TTRPGs), board games, arcade games, art installations, and museum exhibits. Exploring these areas could reveal if the ABPM framework needs additional components to account for other kinds of puzzle mechanics found in other contexts.

Audio-Based Puzzles: Trends and Statistics

This section traces the evolution of ABPs over time and presents descriptive statistics derived from the collected sample of 148 puzzles found within 114 games released between 1990 and 2025. It is a common design practice in puzzle games to repeat the same puzzle with incremental variation, such as increased complexity. Accordingly, repeated puzzle mechanics within the same game were counted only once when they represented variations of the same puzzle. However, mechanics that recurred in distinct contexts were treated as separate instances.

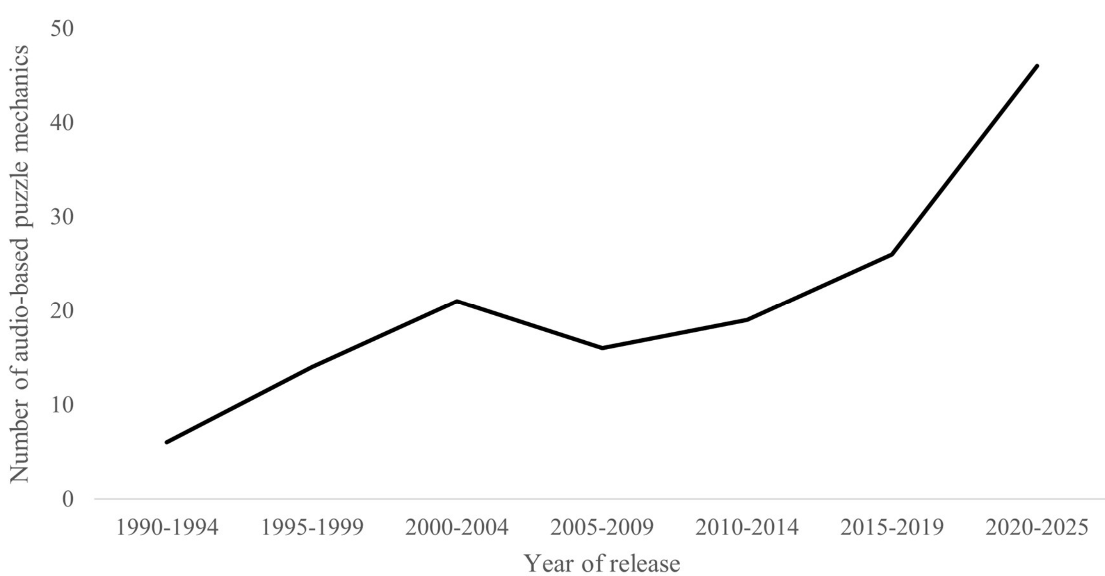
Caution should be exercised before generalizing these trends to the full population of ABPs. Although efforts were made to collect a representative and exhaustive sample, early video game titles are often less accessible and less well preserved (Gooding & Terras, 2008; James, 2012), and isolated instances of ABPs within non-puzzle genres can be difficult to locate. Some degree of sampling bias is therefore likely.

Prevalence of ABPs from 1990 to 2025

Mapping the number of puzzle mechanics over time reveals that ABPs increased in prevalence throughout the 1990s, plateaued in the 2000s, and rose sharply after 2010 (Figure 3.5). Several factors may explain this apparent stall in the early millennium. First, the trend aligns with the relative popularity of the puzzle and adventure genres, which declined in the early 2000s in favour of first-person shooters (Moss, 2011; Salter, 2014; Wolf, 2008). In the 2010s, puzzles experienced a renaissance driven by the rise of indie games, smartphones, and crowdfunding platforms that allowed players to directly support the development of games (Mora-Cantallops, 2018; Salter, 2014). Second, sound in games has historically been viewed as secondary to visuals—a belief echoing similar attitudes in film—resulting in audio design and technology being frequently overlooked in game development and game studies (Collins, 2008). As audio games are particularly reliant on technological advancement (Urbanek & Güldenpfennig, 2019), ABPs may have stalled until platforms conducive to accessible and convenient audio-based interaction, such as smartphones, became widespread. Third, while accessibility has long been a concern in game design, the 2010s saw an increase in publications offering accessibility guidelines and design recommendations for the development of audio games (Garcia & de Almeida Neris, 2013; Rovithis et al., 2014; Tong et al., 2014; Yuan et al., 2011), as well as prototypes of accessible ABP mechanics (e.g., Allain et al., 2015; Brieger, 2013; Kim & Ricaurte, 2011), which could have encouraged the development of ABPs. In any case, the number of APBs appears to be trending upward.

Figure 3.5

Number of audio-based puzzles mechanics by release year, from 1990 to 2025.



Direct, Indirect, and Mixed

Examining the modality of puzzle mechanics reveals the relative prevalence of puzzles using direct, indirect, and mixed mechanics.[9] In this analysis, the direct-only and indirect-only categories contain puzzles that use one modality exclusively, while the mixed category includes instances of puzzles that employ both simultaneously. For example, the tombstone puzzle from *Jak II* (Naughty Dog, 2003) is direct only, as it only uses a direct mechanic. In contrast, the piano puzzle from *Resident Evil*, which only requires interacting with a piano, belongs to the indirect category. In *The Legend of Zelda: Ocarina of Time*, many puzzles use mixed mechanics, as notes played on the ocarina are represented both visually and sonically.

Across the sample, mixed puzzles constitute 53.4% of cases, followed by direct only at 36.5% and indirect only at 10.1% (Figure 3.6). This distribution likely reflects the frequent use of sound-producing devices in ABP design: even when the primary mechanic is indirect, the act of interaction often produces sound, resulting in a mixed modality by default. Direct-only puzzles, although less common than mixed puzzles, appear to be growing in popularity, and between 2010 and 2014 were in fact more common than mixed puzzles (Figure 3.7). Since then, the ratio of direct only to mixed puzzles appears to have equalized. The same factors that contributed to the general rise in ABPs (i.e., indie game development and the availability of resources on auditory game mechanics) likely contributed to this shift.

[9] As a reminder, direct mechanics communicate essential information through sound, whereas indirect mechanics represent sonic information through another modality.

Figure 3.6

Distribution of puzzles making use of direct only, indirect only, or mixed mechanics.

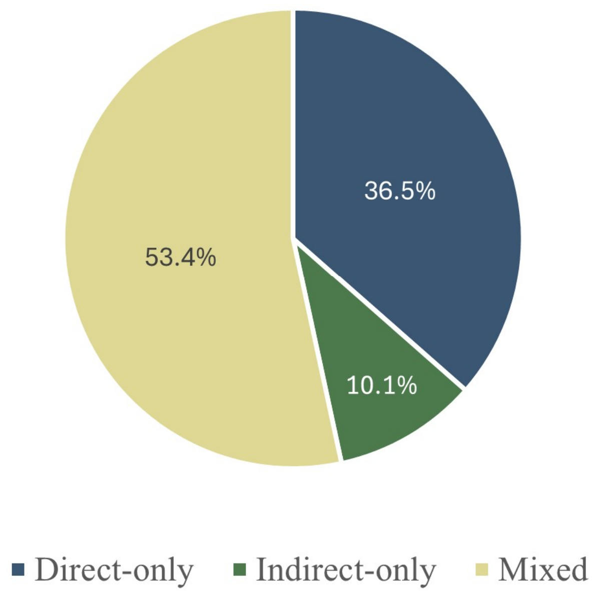
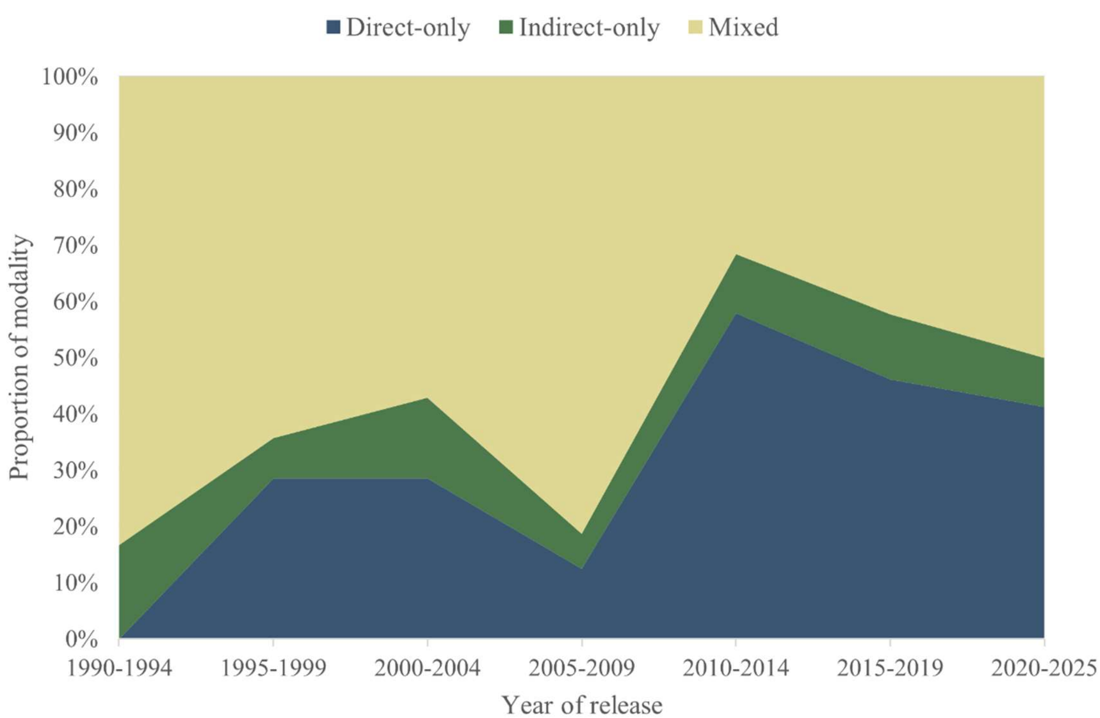


Figure 3.7

Proportion of modalities by release year.



ABPs by Category

Within the ABPM categories, I/O mechanics account for 67.6% of the sample (Figure 3.8) and have remained the most prevalent category across the surveyed time span (Figure 3.9). I/O puzzle mechanics are followed by Spatial (11.5%), Contextual (9.5%), Epistemic (7.4%), and Environmental (4.1%) mechanics. I/O puzzles are evenly divided between match (45%) and Transform (55%) subcategories. Spatial puzzles, however, show a notable imbalance: positive puzzles account for 76% and negative for 24%.

Multiple factors likely explain this distribution. I/O mechanics are versatile, relatively easy to implement, and thereby represent a low-risk and low-complexity design choice. They often involve pattern matching (e.g., matching melodies in *Jak II*) or short-term memory tasks (e.g., memorizing note sequences in *Ocarina of Time*). Because these mechanics have visual analogues in non-audio puzzles, the barrier to player understanding is low. Additionally, the various possible combinations of inputs and outputs make them a flexible option that allows for reutilization of similar mechanics without diminishing the players' perception of novelty. By contrast, environmental mechanics are typically more difficult and time-consuming to design and implement for the developers, and to understand and manipulate for the players. Given the high initial investment of these puzzles, developers are likely inclined to use it as the main mechanic of an entire game, such as *Sonority* (Hanging Gardens Interactive, 2022) or *One Hand Clapping* (Bad Dream Games, 2021), rather than as a one-off puzzle that could be difficult for the player to get adjusted to.

One also cannot ignore the impact of historical tradition and the persistence of established design tropes. I/O puzzle mechanics appear prominently in several seminal titles whose influence has extended across decades, shaping expectations for both designers and players. A particularly notable case is the gameplay loop of *Simon* (Baer & Morrison, 1978), which has been replicated in countless games and accounts for multiple entries in the collected sample. A number of influential games from the 1990s also incorporated variations of I/O mechanics, and these titles likely informed later design practices, such as *The 7th Guest* (Trilobyte, 1993), *King's Quest VII* (Sierra On-Line, 1994)*, Zork Nemesis* (Zombie LLC, 1996), *Super Mario RPG*, and of course, *The Legend of Zelda: Ocarina of Time*. By contrast, mechanics that fall outside the I/O category tend to appear in smaller productions, which may enjoy critical acclaim but typically lack the widespread visibility of major studio releases. As a result, many developers might gravitate toward I/O mechanics because they align with well-established conventions and come with a guarantee of player recognizability, unlike less common mechanics that may be perceived as experimental or niche and be more susceptible to poor reception.

Figure 3.8

Total number of puzzle mechanics in each ABPM framework category.

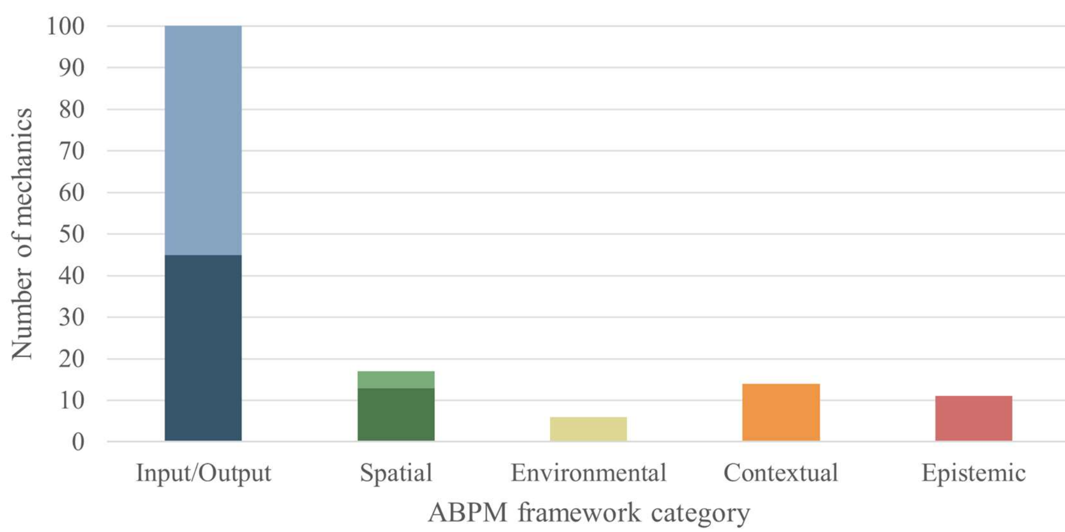
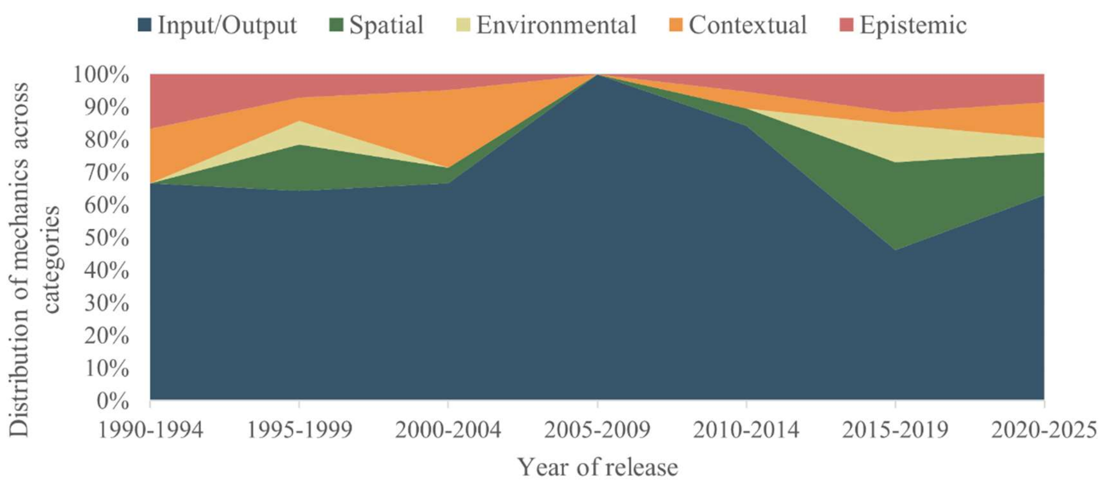


Figure 3.9

Distribution of puzzle mechanics across ABPM framework categories from 1990 to 2025.



Relationship Between Modality and Category

Examining modality within each ABPM category reveals notable divergences (Figure 3.10). Whereas mixed mechanics are predominant in the I/O, Environmental, and Contextual categories, both Spatial and Epistemic categories are dominated by direct-only puzzles. Additionally, neither the Spatial nor Environmental categories contain indirect-only puzzles.

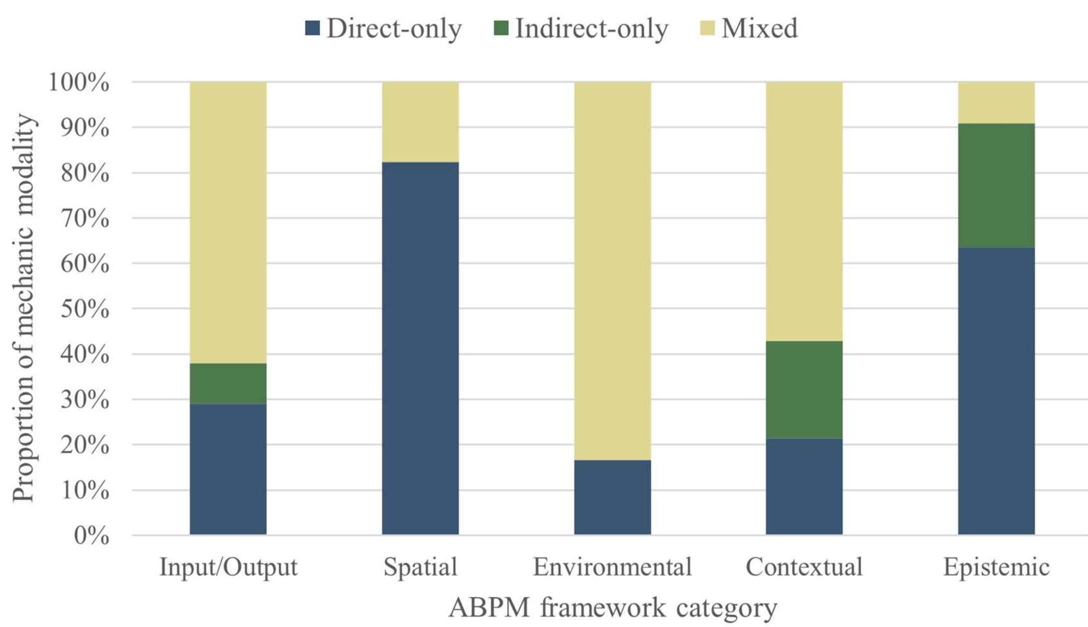
Designers working with I/O, environmental, and contextual mechanics may assume that most players possess stronger visual interpretive skills than trained auditory perception and discrimination abilities; supplementing audio with indirect mechanics therefore helps ensure players are able to solve puzzles. This decision could also reflect broader inclusivity concerns as mixed designs offer redundancy across sensory channels, which can benefit players who are hard-of-hearing or have low vision. For instance, using the ocarina to solve puzzles in *The Legend of Zelda: Ocarina of Time* brings up a visual interface that displays the buttons pressed on the controller by the player, each corresponding to a specific note. The player can therefore rely on both auditory and visual cues, making it generally easier to memorize and play melodies as each note can be mapped to a visual icon, and allowing deaf or hard-of-hearing players to complete the puzzles without having to rely on the game's audio.

The predominance of direct-only mechanics in spatial puzzles is more easily understood and can most certainly be explained by cognitive ecology. Direct spatial mechanics, by their nature, rely on loudness and/or binaural cues, and the interpretation of this type of information is naturally developed throughout life without specialized training, unlike fine pitch perception. Introducing indirect representations of spatialized sound therefore adds little value and may even be seen as counterintuitive to the auditory experience, such as in *Maze Sounds* (AMIDEA Games, 2016), where the goal of the game is to avoid walls that can only be located through audio. As a result, designers may prefer to use direct mechanics, producing puzzles that depend entirely on auditory perception.

Epistemic puzzles present a different dynamic. The relative scarcity of mixed mechanics likely stems from their quiz-like structure: these puzzles tend to hinge on the player's ability to identify or recognize sounds. Because providing additional non-audio cues could reveal too much information, designers may intentionally limit the modality to sound alone in order to preserve challenge. In this sense, the absence of mixed representations may reflect a deliberate design choice rather than a technical limitation.

Figure 3.10

Proportion of modalities in each ABPM framework category.



Conclusion

This chapter addressed several conceptual and methodological limitations of the ABPM framework and provided insight into the rationale behind its design decisions. While some definitional challenges, such as distinguishing puzzles from exercises or determining when speech constitutes an audio-based mechanic, may constitute unsolvable puzzles (ha!), articulating the framework's boundaries helps ensure transparency and analytical consistency. The chapter also outlined potential expansions to the framework and identified areas where further refinement may become necessary as new forms of audio-based interactions emerge.

The statistical survey of ABPs over 35 years highlighted several major trends, including an overall increase in ABPs and a growing use of direct-only mechanics. These developments reveal the close ties between ABPs, shifts in the video game industry, and research on audio games. The rise of indie games, advances in accessible game design, and changing attitudes toward the role of audio in interactive media, are all likely factors that play into the prevalence of ABPs in the video game market. In all, the creation and application of the ABPM framework reveal the evolving nature and the exciting potential of ABP design. The framework also provides a structured vocabulary for analyzing existing works and a foundation upon which future research and design innovation can build upon.

4. Protocol for a Mixed-Methods Study of Puzzle-Solving in Musicians and Non-musicians

Video games are increasingly being used by researchers "to aid in our understanding of skill acquisition, cognitive capacity and plasticity, development and aging, and individual differences" (Boot, 2015, p. 1). In recent years, several studies have explored the relationship between problem-solving ability and gaming. Video games, especially those from the puzzle genre, are a great candidate to study problem-solving as players must build upon previously learned rules to solve novel problems. Such games have been shown to have a positive impact on logical reasoning (Chen, 2019; Hubana, 2022; Shute et al., 2015; Weng, 2022) and creativity (Crombie et al., 2016; Roshanian Ramin et al., 2022). Puzzle games have also been used to develop measures of problem-solving ability (Foroughi et al., 2016; Sarathy et al., 2024; Shute & Wang, 2015).

Although academic attention has been dedicated to video game puzzles and cognitive abilities, as of today, very few studies have specifically investigated this relationship in the context of Audio-Based Puzzles (ABPs). As such, little data is available on players' problem-solving and perceptual abilities when confronted with gaming tasks that have a particular focus on auditory stimuli. In addition, it is currently unclear how an individual's musical expertise or listening skills might affect their puzzle-solving performance and subjective experience. Considering that ABPs present unique cognitive challenges not found in other puzzles, an investigation into the relationship between player, puzzle, and audio is warranted.

The purpose of this chapter is to document the protocol for an experiment comparing musicians and non-musicians on audio and visual puzzle-solving tasks. The study aimed to investigate the relationship between puzzle-solving and musical expertise via the following research goals:

- Explore the influence of musical expertise on puzzle-solving performance and strategies
- Explore the subjective experiences of musicians and non-musicians when solving ABPs in terms of challenge, enjoyment, and engagement with music

The protocol is presented as a standalone chapter because the subsequent chapters all draw on this experiment and its data while addressing different topics. This structure enables each chapter to provide a more focused literature review and discussion of its respective research questions. Chapter 5 examines the quantitative results of the experiment and relates them to literature on near and far transfer effects of musical instruction. Chapter 6 explores themes of game difficulty, optimal challenge, and design practices, while Chapter 7 considers the role and potential of ABPs in music education and in the gamification of music.

Method

Participants

Participants in this study (N = 30) were divided into two groups: musicians (n = 10) and non-musicians (n = 20). Participants were assigned to the musicians (M) or non-musicians (NM) group based on their self-reported musical expertise. Participants in the M group had achieved a level equivalent to grade eight of the Toronto Royal Conservatory of Music in any instrument and were currently practising at least two hours a week. Participants in the NM group had taken part in no or minimal instrumental lessons. Other inclusion criteria for both groups were normal vision and hearing, proficiency in English, and no prior familiarity with the games used in the experiment. Individuals who did not fit the criteria for either group were not eligible for the study. Descriptive statistics for both groups can be found in Table 4.1.

Recruitment was conducted through posters around the University of Ottawa campus, word-of-mouth, and the INSPIRE laboratory's Integrated System of Participation in Research (ISPR) student pool. Individuals from the ISPR student pool are students currently enrolled in an undergraduate introductory course and are rewarded with up to four percent of their course grade for participating in research studies. For this specific study, participants from the ISPR student pool were rewarded with one and a half percent of their final grade. No compensation was offered to participants outside of the ISPR student pool. The study was conducted in-person at the INSPIRE facilities on the University of Ottawa campus and received approval from the University's Office of Research Ethics and Integrity. All participants provided informed consent before participation.

Materials and Measures

Apparatus

Tasks were administered on a HP Victus 15 laptop (15.6 ″ display size, 1920 × 1080 screen resolution, 60 Hz refresh rate, Windows 11 operating system). Audio was delivered via a Logitech H390 USB headset. Output level was participant-adjusted to a comfortable loudness using the on-screen calibration for the Profile of Music Perception Skills (PROMS) and remained participant-set for the gaming tasks. In-game volume for *Circuits* was set to 100%; in-game volume for *Square Logic* was set to 100% for sound effects but 0% for music. The PROMS was completed in Firefox (versions 134–137, 64-

bit) with default audio settings. The games were presented in Fullscreen at a viewing distance of ~60 cm and participants' verbalizations during gameplay were captured through the headset microphone. Screen and audio were recorded with OBS Studio version 30.2.3. Testing took place in a quiet room with stable lighting.

Stimuli

The two games used in this study, *Circuits* (Digital Tentacle, 2014) and *Square Logic* (New Dawn Game Studio, 2023), were selected for multiple reasons:

1. They are commercially available and therefore are more representative of the current gaming market than internally developed games used only for the study's purpose, thus improving external validity.
2. They both have levels with clear beginning and end, and can be solved using extraneous actions, facilitating the capture of certain performance metrics.
3. They have no narrative, meaning the participants could not be distracted by elements external to the puzzle and could focus solely on puzzle-solving.
4. They are accessible to a non-gaming population as they only require the use of a mouse to play as opposed to a controller or a combination of mouse and keyboard.
5. They contain a tutorial which explains the premise of the game and how to operate it.
6. They are relatively less known than other puzzle games, making it unlikely that potential participants would be familiar with them.
7. There is a high contrast between the puzzle-solving process in both games in terms of modality (audio vs. visual).

Both games are presented in more detail below.

Table 4.1

Descriptive statistics for the musician and non-musician groups.

Variable	Musicians	Non-musicians	Total
Number of participants	10	20	30
Age (mean, *SD*)	23.1 (*4.9*)	19.6 (*2.4*)	20.7 (*3.7*)
Gender			
Male	4	7	11
Female	6	13	19
Weekly Gaming Hours			
None	3	5	8
1–5	4	7	11
6–10	0	4	4
11–15	2	2	4
16+	1	2	3
Proportion of Gaming Hours Spent on Puzzle Games[10]			
None	1	6	7
Very Few	1	5	6
Few	2	1	3
Half	2	1	3
Most	0	1	1
All	1	1	2
Years of Musical Training (mean, *SD*)	14.2 (*5.7*)	—	—
Weekly Practice Hours		—	—
Less than 2	0		
2–4	7		
5–7	0		
8–10	1		
11+	2		
Main Instrument		—	—
Piano	4		
Violin	2		
Voice	2		
Flute	1		
Drums	1		

[10] Including only participants who reported more than zero weekly gaming hours.

Circuits

Circuits is an audio-based puzzle game where the main puzzle-solving mechanic can be categorized as Input/Output Direct Match.[11] In each level of the game, the player has to reconstruct an instrumental piece of music that has been fragmented into several pieces. To do so, the player is presented with an interface resembling a circuit board with nodes and links which represents a piece of music that is usually 30 to 60 seconds in length (Figure 4.1). At the bottom of the screen are small circles which contain fragments of this piece of music, normally important melodic elements (such as a single note or phrase) or rhythmic patterns (such as a drumbeat). By pressing on the play button at the bottom-left corner of the screen, the player can enter "playmode" to listen to the entirety of the piece while watching the circuit board fill up from left to right, with the ability to rewind or fast forward (Figure 4.2). The goal is to place the fragments on the appropriate nodes as to reconstruct the piece. Once the player believes they have put the fragments back in the correct order, they can initiate the final playback process by pressing on the play button at the left of the circuit board. Once the piece has been played in its entirety, a screen informs the player if the solution they have provided is correct, in which case the player can proceed to the next level. Otherwise, they are asked to return to the circuit board and reorganize the fragments (Figure 4.3).

Starting at level three, the player has the option to isolate layers of the piece of music to assist them in identifying the position of the fragments inside the piece (Figures 4.4 and 4.5). This is a completely optional process as the player can opt to simply listen to the piece of music with all its layers active. However, doing so is more challenging as

[11] See Chapter 2 for details about categories of ABP mechanics.

the player must discern between the fragments they are looking for and the other layers of the song. At level five, the game introduces decoys, which are musical fragments present amongst the options at the bottom of the screen, but which do not correspond to anything in the piece of music. These red herrings are typically made to sound similar to actual fragments, making it difficult to discriminate between the decoys and the actual fragments. As the game continues and the player reaches higher levels, more gameplay elements are added to make the game more complex. However, since these were not included in the study, they will not be discussed.

Figure 4.1

Circuit*'s basic interface for the second level (annotated)*

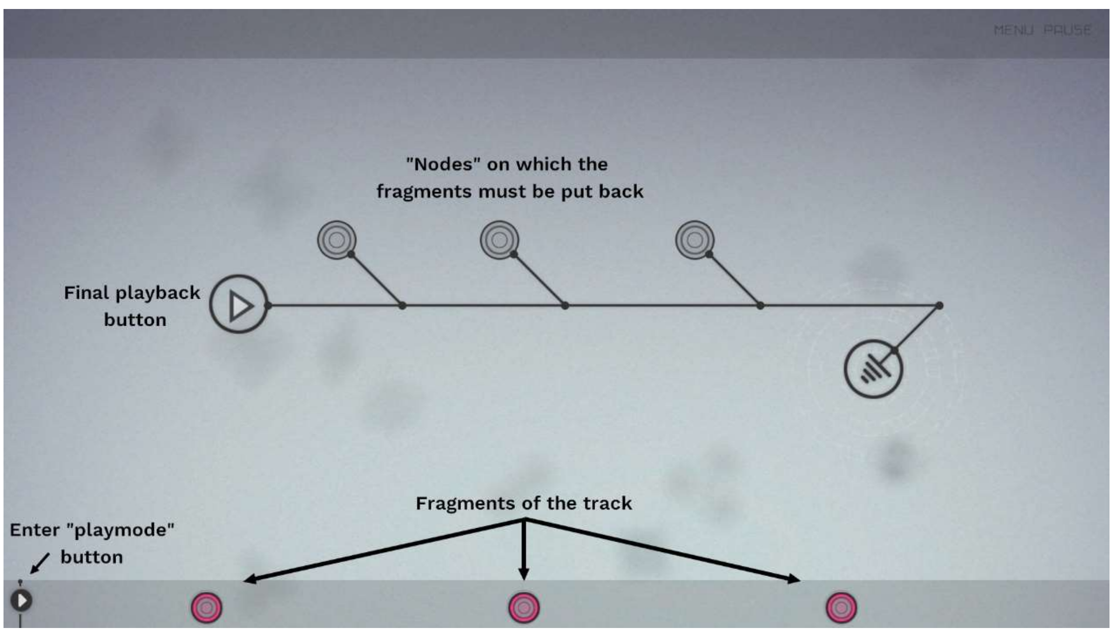


Figure 4.2

Circuit*'s "playmode" interface for the second level. The circuit board slowly fills up as the track is being played. Players can rewind/fast-forward via the bottom bar.*

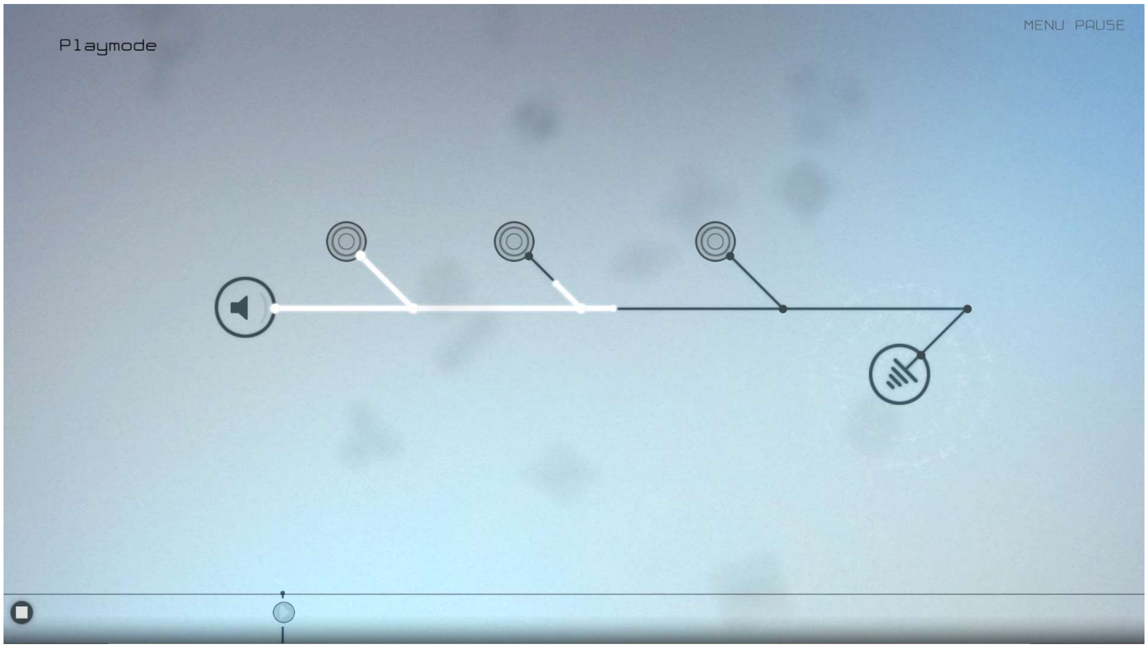


Figure 4.3

Circuit*'s error feedback screen.*

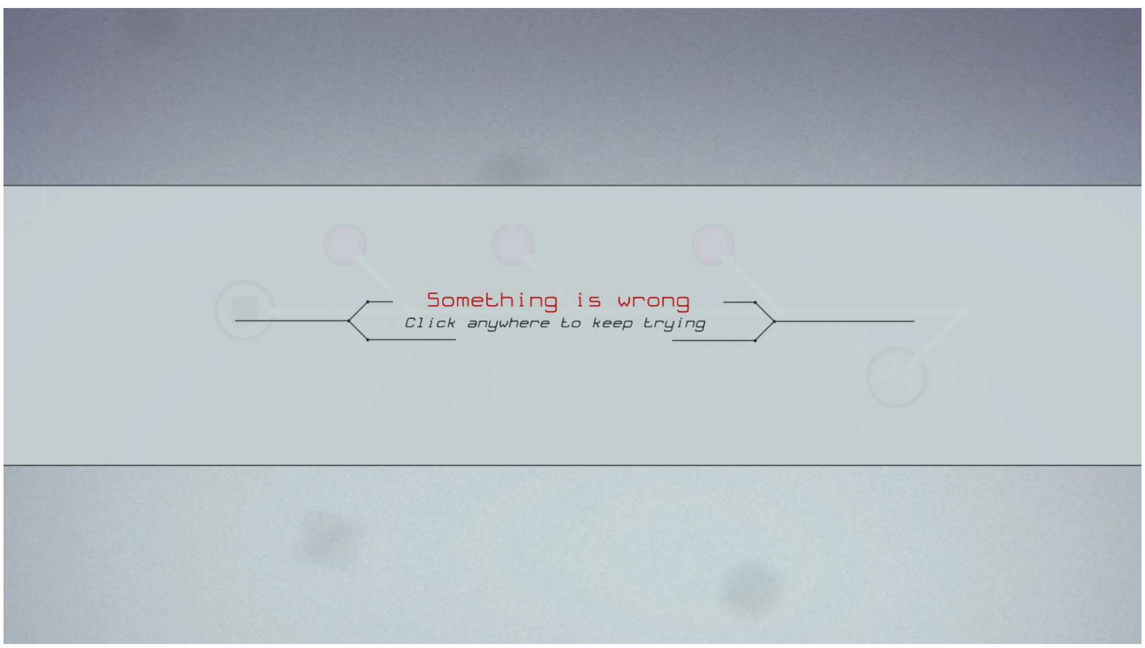


Figure 4.4

Circuit*'s tutorial messages in the third level, which introduces music layers.*

The top text reads: "Click to select a music layer. You will only hear your selected music layer. The white color is the sum of all music layers."

The bottom text reads: "The colors on the musical circles represent a group of similar instruments. We call this music layer"

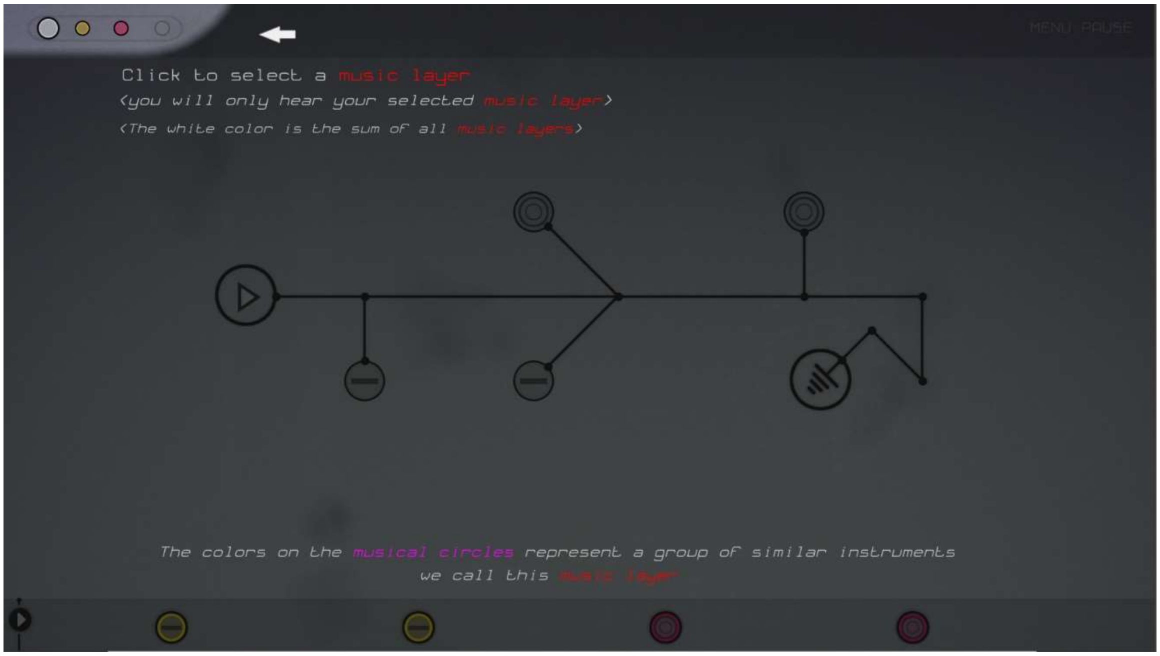
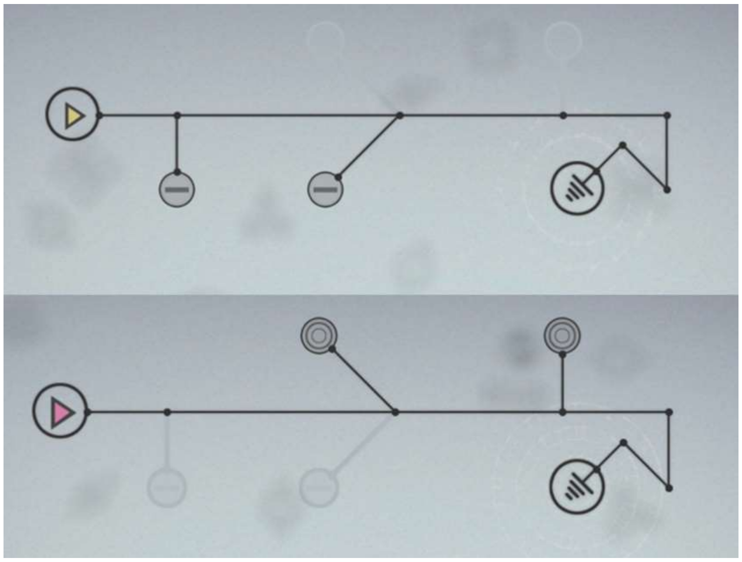


Figure 4.5

Example of Circuit*'s music layers. Top: the yellow layer (bass) is isolated. Bottom: the red layer (piano) is isolated.*



Square Logic

Square Logic is a visual puzzle game where players fill out a grid by drawing squares and rectangles. At specific spots in the grids are numbers; each drawing of a square or a rectangle must contain only one number, and the drawing must be the size of that number (Figure 4.6). As such, players are limited in the ways they can fill out the grid, as they have to find a way to satisfy all the conditions imposed by the numbers inside it. If the player draws a shape which does not fulfill the condition set by the number, the game will indicate this to the player by making the outline red (Figure 4.7). The player first begins with a tutorial explaining how the game is played and afterwards is given access to the 5 × 5 category. After a certain number of levels are completed, the player unlocks the 7 × 7 category, and so on until they reach the final category of 13 × 13. In this study, players completed the tutorial, ten levels of the 5 × 5 category, and five levels of the 7 × 7 category.

Figure 4.6

A 5 × 5 level of Square Logic *in its starting state (left) and solved state (right).*

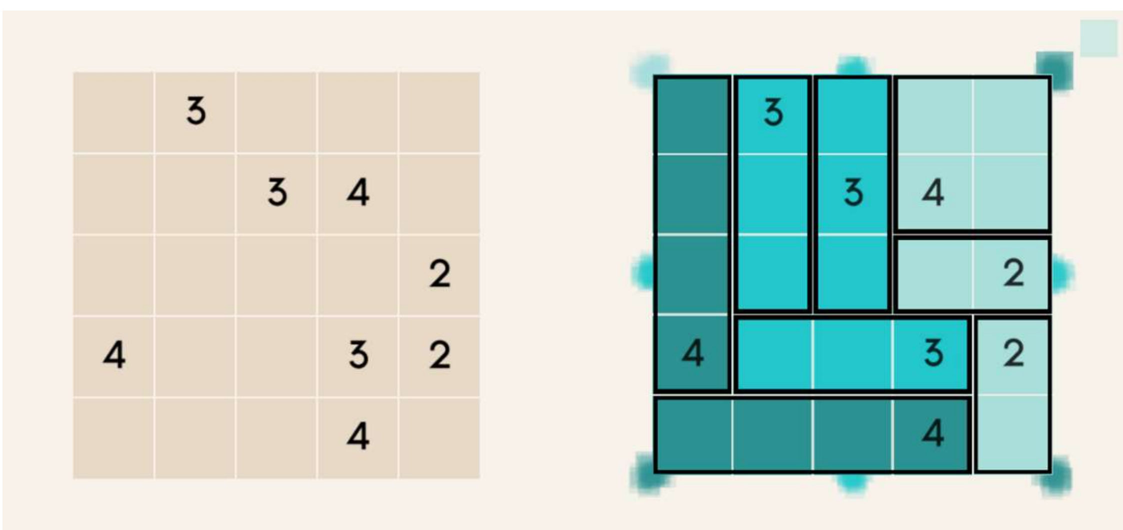
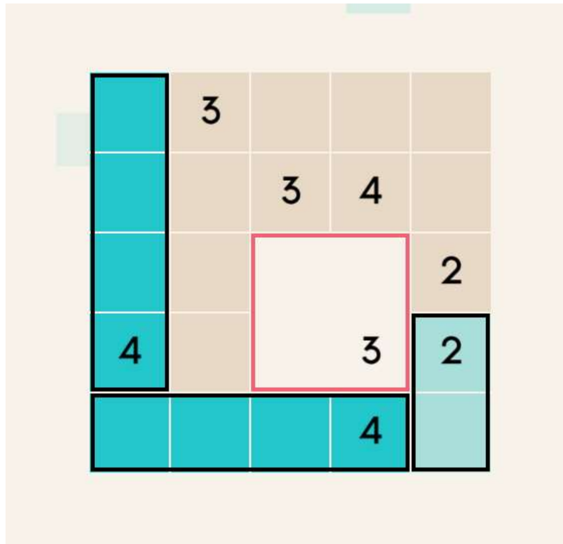


Figure 4.7

A 5 × 5 level of Square Logic *in which one of the shapes does not respect the size condition and is outlined in red.*



Measures

Demographic Questionnaire

The purpose of the demographic questionnaire was to capture basic demographic information (age, gender), gaming habits (weekly hours playing video games, proportion of gaming hours spent on puzzle games), and for the musician group, information about musical expertise (main instruments, years of musical training, weekly practice hours). A copy of the questionnaire can be found in Appendix B, and results are presented in Table 4.1.

Modular-PROMS

The Profile of Music Perception Skills (PROMS) is a battery conceived as an alternative to the use of musicianship status as a method to determine musical competence. The original test, called the Full-PROMS, consists of 9 subtests, each containing 18 trials, which assess the participants' ability to perceive small differences in various musical stimuli such as melody, rhythm, and timbre (Law & Zentner, 2012). In each trial, the participant is twice presented with a reference stimulus, followed by a probe stimulus which is either identical or different to the reference. The participant then indicates if the reference and the probe are definitely identical or definitely different (for a full point), probably identical or probably different (for half a point), or if they are unsure (no point). This results in both a score for each subtest and a total score for the whole test. The entire procedure takes around 60 minutes.

Several versions of the PROMS have since been developed to accommodate different research needs: the PROMS-S, which consists of 8 subtests of 8–10 trials each, taking around 35 minutes; the Mini-PROMS, which consists of 4 subtests of 8–10 trials each, taking around 15 minutes and; the Micro-PROMS, which consists of a single test of 18 trials, taking around 10 minutes (Strauss et al., 2024; Zentner & Strauss, 2017). Additionally, the custom modular-PROMS allows researchers to decide on which subtests to include and the number of trials for each subtest.

This study opted to use the modular-PROMS with the following four subtests, each containing 8 to 10 trials:

- **Pitch:** Determine if two tones are the same or different.
- **Timbre:** Determine if two four-note chords are played by the same instrument(s) or not.
- **Embedded Rhythm:** Determine if two rhythmic patterns are the same or different. In the reference, the pattern is played in an unpitched percussive form, while in the probe, the rhythm is embedded in a melody.
- **Melody:** Determine if two short monophonic melodies with constant rhythms are the same or different.

In total, this version of the modular-PROMS took approximately 20 minutes to complete. A score for the entire test and each subtest was calculated.

While this study used musicianship status as a recruitment criterion to divide the participants into the musician or non-musician groups, the modular-PROMS was included in the protocol to better discern individual differences among participants. Additionally, the four subtests selected for the modular-PROMS were specifically chosen as they closely mimicked the challenges presented by the audio-puzzle game retained for the study, allowing for more precise analysis of the relationship between perceptual abilities of musical stimuli and puzzle-solving performance.

Performance Measures

Two variables were measured to assess puzzle-solving performance in both games: *time taken* and *number of actions taken*. A third variable was measured for *Circuits* only: *number of incorrect solutions submitted*. All three variables were inspired by the Puzzle Challenge Analysis Tool (Pusey et al., 2021), which details possible metrics to measure the challenge level of a puzzle based on its characteristics. In both games, puzzles have a clear beginning and end, and can be solved using extraneous actions, making *time taken* and *number of actions taken* apt measures of performance. The third variable, *number of incorrect solutions submitted*, only applies to *Circuits* as incorrect solutions cannot be submitted in *Square Logic*.

All performance measures were logged manually by reviewing gameplay footage. *Time taken* was measured by calculating the time between the start and end of each level. *Number of actions taken* was calculated by counting mouse clicks which were intended as a puzzle-solving action (i.e., miscellaneous clicking around the screen was not counted). In *Circuits*, actions consisted of playback activities (e.g., initiating, rewinding, fast-forwarding) and moving fragments around. In *Square Logic*, actions consist of drawing and erasing rectangles. Finally, *number of incorrect submissions* corresponds to the number of occurrences of the "something is wrong" feedback screen appearing (Figure 4.3).

Think-aloud Transcripts

A think-aloud protocol is a data collection method where participants verbalize their thoughts while performing a task (concurrent) or after completion, often while viewing a recording of themselves performing the task (retrospective). Think-aloud protocols have been used for decades and provide rich data about the reasoning process and immediate reactions of an individual during a problem-solving task (Fonteyn et al., 1993). They have also been shown to have minimal impact on meta-awareness and cognitive load (Garg et al., 2025). This study used a concurrent think-aloud method where participants' verbalizations were recorded as they played both games.

Post-Game Questionnaire

A brief post-game questionnaire was employed to collect the participants' thoughts and subjective feelings of challenge and enjoyment following both games. Participants' perceived challenge was captured using the cognitive subscale of the Challenge Originating from Recent Gameplay Interaction Scale (CORGIS; Denisova et al., 2020). The CORGIS is a validated psychometric instrument which quantifies players' perceived challenge of playing a game and can be completed by players of diverse skill levels and playing a variety of games. It consists of four subscales, each subscale containing a set of items which are presented in random order and answered using a five- or seven-point agreement Likert scale. The sum of every response is then calculated to obtain a total score.

This study used a five-point scale ranging from “Strongly Disagree” to “Strongly Agree.” Only the items from the cognitive challenge subset were used, as the other subscales were deemed inapplicable to the challenges posed by the games used in this study. Enjoyment was measured using the item “I found the game to be enjoyable,” which was integrated into the items of the cognitive subscale and rated on the same five-point scale. Participants’ more general thoughts about the games were captured using the optional open-ended question, “If you would like to share some of your thoughts about the game, please do so below (for example, what you liked or disliked, what you found easy or difficult, etc.).” A copy of the post-game questionnaire can be found in Appendix C.

Design

This study employed a 2 × 2 mixed factorial design with Group (musicians, non-musicians) as a between-subjects factor and Game Type (audio, visual) as a within-subjects factor. To avoid sequence effects, game order was counterbalanced within each group via alternating assignment: odd-numbered enrollments played *Circuits* first, and even-numbered enrollments played *Square Logic* first.

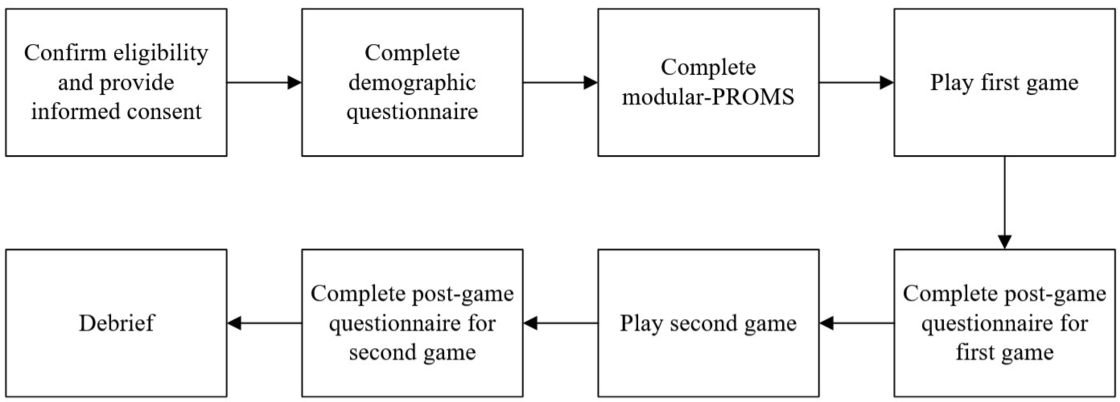
Procedure

Each participant completed a single laboratory session lasting approximately 75 minutes. Upon arrival, they confirmed eligibility and provided informed consent, then sat at a desk in front of a computer in a quiet laboratory room and followed on-screen instructions. Participants wore headphones throughout the tasks. First, they completed the demographic questionnaire followed by the modular-PROMS. Next, they played either *Circuits* or *Square Logic* while verbalizing their ongoing thoughts in either English or

French. Immediately after the game, participants completed the post-game questionnaire. They then played the other game under the same think-aloud and recording procedures, after which they completed the same post-game questionnaire. At the end of the session, participants were shown their modular-PROMS results, debriefed about the study aims, invited to ask questions or offer comments, and thanked for their time. The procedure is illustrated in Figure 4.8.

Figure 4.8

Flowchart of the study's procedure.



Conclusion

This chapter outlined the protocol for an experiment comparing the performance and experiences of musicians and non-musicians across two puzzle games, one emphasizing auditory abilities and the other visual abilities. The following chapters present and discuss the results of this experiment in relation to several key themes, including task performance, perceived challenge, and subjective experiences of enjoyment and frustration. First, the next chapter focuses on the quantitative performance results and contextualizes these findings within the literature on near and far transfer effects of musical expertise.

5. Influence of Musical Expertise on Video Game Puzzle-Solving Performance

Audio-Based Puzzles (ABPs) are challenges found in video games that involve sonic information presented in any modality. In particular, direct ABPs, as discussed in Chapter 2, require players to make use of their auditory abilities in order to solve the problem at hand. Given this defining feature, it is plausible that puzzle-solving performance in such games is related to a player's musical training. Despite this intuition, there is little empirical evidence examining this relationship, either for puzzle-solving in general and for ABPs specifically. The research presented in this chapter provides a first step toward filling this gap.

Collecting data on this topic has two primary benefits. First, uncertainty regarding how musical expertise influences puzzle-solving performance presents challenges for game designers attempting to create ABPs that are appropriately difficult for both musicians and musically untrained players. This design difficulty may partly explain why ABPs are less prevalent than visually based puzzles. Second, examining musical expertise in relation to puzzle-solving performance allows for the investigation of cognitive differences between musicians and non-musicians in a context that is more ecologically valid than a traditional laboratory task. Although such differences are well documented, it remains unclear whether they persist or change in magnitude in the context of a ludic activity such as video game puzzle-solving. Supplementary data in this area may therefore clarify existing evidence regarding the impact of musical training on cognition.

This chapter begins with a review of research on musician/non-musician differences in auditory and visual skills, followed by an overview of the limited literature linking musical training to puzzle-solving performance. It then presents and discusses the performance-related results of the experiment detailed in Chapter 4 comparing the performance of musicians and non-musicians on the APB game *Circuits* (Digital Tentacle, 2014) and the visual puzzle game *Square Logic* (New Dawn Game Studio, 2023).

Literature Review

Music Training, Auditory Perception, and Visuospatial Abilities

There is substantial evidence linking musical expertise and enhanced auditory abilities. A meta-analysis of 62 longitudinal studies reported a small positive effect of music training on auditory and linguistic processing ($g\Delta = 0.31$; Neves et al., 2022). Importantly, this effect was not moderated by domain (auditory vs. linguistic), type of music training, or type of control group, suggesting it is relatively robust across study designs. However, baseline differences between groups did moderate the effect, with larger pre-training differences associated with smaller improvements. Of the 18 studies in that review focusing specifically on auditory processing, 16 reported significant benefits, a trend observed across age groups and methodological approaches. Numerous studies over the years have shown that musicians exhibit enhanced pitch perception relative to non-musicians, including superior discrimination and memory as well as faster processing (Bianchi et al., 2016; Bidelman et al., 2013; Cohrdes et al., 2019; Micheyl et al., 2006; Schellenberg & Moreno, 2010; Tervaniemi et al., 2005).

Although less extensively studied, evidence suggests that these advantages extend to rhythm and timbre perception (Foster et al., 2021; Law & Zentner, 2012; Roden et al., 2014; Swaminathan, 2017; Zentner & Strauss, 2017). Musicians also appear better able to detect frequency changes under conditions of added artificial noise (Liang et al., 2016). Because ABPs are often encountered in environments where extraneous noise is present, both in-game (e.g., background music and sound effects) and out of game (e.g., ambient noise), and because musicians generally demonstrate finer auditory discrimination, these findings suggest that musicians might hold an advantage over non-musicians when solving ABPs.

In the visual domain, a meta-analysis of nine experiments concluded that music training led to improvements across a range of spatial tasks, including spatial memory, spatial recognition, mental rotation, and spatial visualization ($d = 0.40$; Hetland, 2000).[12] Effect sizes were consistent across studies and not significantly heterogeneous, suggesting no strong moderating variables. Subsequent studies have supported this hypothesis. For example, Pietsch & Jansen (2012) found that university music and sports students outperformed education students on a mental rotation task, and Sluming et al. (2007) reported that orchestral musicians were both faster and more accurate than non-musicians on a similar task. However, some studies have reported contrary findings: Rauscher et al. (1997) observed no effect of music training on spatial recognition, and Cohen et al. (2011) found no differences between musicians and non-musicians on

[12] These four abilities can be defined as follows: *spatial memory* is the ability to remember locations and spatial layouts; *spatial recognition* is the ability to identify previously seen spatial patterns or objects; m*ental rotation* is the ability to mentally rotate objects in space; and *spatial visualization* is the ability to manipulate and transform complex spatial information mentally. These brief definitions are provided for ease of reference, but there is substantial discussion in the literature regarding how spatial abilities should be divided and defined (see Linn & Petersen, 1985; Shepard & Metzler, 1971; Uttal et al., 2013).

measures of spatial recognition and memory. These discrepancies underscore the need for further research.

Music training appears to be particularly associated with improvements in spatial-temporal tasks, which require the mental rotation of two- or three-dimensional figures and the execution of multiple solution steps, such as the Object Assembly subtest of the *Wechsler Preschool and Primary Scale of Intelligence-Revised* (Wechsler, 1989). Performance on this measure is especially relevant to the present line of research, as it involves assembling puzzle pieces to form a recognizable object under time constraints, closely mirroring gameplay loops encountered in many puzzle video games. A second meta-analysis by Hetland (2000), synthesizing fifteen studies, reported a large, positive effect of music instruction on spatial-temporal task performance ($d = 0.84$). The effect was consistent across programs of varying length and was not moderated by socioeconomic status or parental involvement, though there was a trend toward stronger effects with one-on-one compared to group instruction. Nevertheless, positive results are not unanimous: Forgeard et al. (2008) reported no difference on the Object Assembly test between children with at least three years of music instruction and control participants.

Evidence regarding musicianship and visual working memory (i.e., the ability to temporarily store and manipulate visual information to support ongoing tasks) is mixed. Visual working memory is commonly assessed using sequence-learning tasks such as the Corsi block-tapping test (Corsi, 1973), in which participants reproduce increasingly long sequences of block locations, either tapped by an examiner on a board or illuminated on a digital screen. Several studies reported superior performance by musicians on this test or closely related tasks (Amer et al., 2013; Anaya et al., 2017; Bidelman et al., 2013; Cara, 2024; Gagnon & Nicoladis, 2021), whereas others have found no group differences (Giovagnoli & Raglio, 2011; Vassena et al., 2016).

Finally, just as musicians appear better able to filter out extraneous noise, they may also have an advantage in inhibiting irrelevant visual information. Pomini et al. (2025) compared musicians and non-musicians on the Navon test, in which participants identify either a large "global" letter or the smaller "local" letters composing it. While accuracy did not differ between groups, musicians were faster in both conditions, suggesting that long-term music training may promote more efficient processing of relevant and irrelevant visual information. Overall, although some studies report no association between music training and visuospatial abilities, a substantial body of evidence suggests a positive relationship. These findings raise the possibility that musical expertise could positively influence performance on a visual puzzle-solving task such as the one examined in the present study.

Music Training and General Puzzle-Solving Performance

Direct evidence connecting music training to puzzle-solving performance (analog or digital) is limited. Nevertheless, hypotheses can be informed by the broader literature on musical expertise and cognitive abilities. Although the causal relationship between music training and cognitive enhancement remains nebulous, considerable evidence points to a positive correlation between the two. In children, music education has been associated with improvements in IQ (Kaviani et al., 2014; Schellenberg, 2004, 2006, 2011), and two meta-analyses found small but significant benefits of music programs on children's cognitive abilities and academic results ($g = 0.28$, Cooper, 2020; $g\Delta = 0.26$, Román-Caballero et al., 2022). Both effects were robust across study designs and types of music intervention, though effect sizes were reduced when studies used active control groups or stricter methodological criteria.

Focusing on nonverbal reasoning, the domain arguably most relevant to puzzle-solving, evidence is mixed. One common measure is Raven's Advanced Progressive Matrices, which consists of increasingly difficult multiple-choice questions requiring participants to identify the missing element in a visual pattern (Raven & Raven, 2003). Hetland (2000), synthesizing five independent effect sizes, concluded that music instruction does not lead to improved performance on the test. Similarly, Schellenberg & Moreno (2010) found no difference between musically trained and untrained undergraduate students. By contrast, Thompson et al. (2004), Trimmer & Cuddy (2008), and Swaminathan (2017) reported superior performance by musically trained undergraduates. Additionally, comparable benefits have also been observed in children receiving instrumental lessons (Forgeard et al., 2008) and in musically trained adults on the Matrix Reasoning subtest of the *Wechsler Abbreviated Scale of Intelligence—Second Edition* (Wechsler, 2011), which is similar to Raven's matrices (Anaya et al., 2017).

Of particular relevance, Winston et al. (2022) administered the *Tower of London* task (Mueller & Piper, 2014), a test closely related to the classic *Tower of Hanoi* puzzle, to 88 undergraduates with varying levels of musicianship. The task assesses planning and problem-solving ability by requiring participants to pre-plan and execute a sequence of moves to transform a starting arrangement of coloured balls or rings into a goal arrangement. Participants with ten or more years of musical experience also completed a harmonization exercise, composing eight measures of alto, tenor, and bass lines to support a soprano melody in accordance with Western European harmonic conventions.

Performance on the *Tower of London* task correlated positively with both music training and harmonization scores, suggesting that greater musical expertise is associated with enhanced planning and problem-solving abilities.

Recent meta-analyses suggest that positive associations between musical training and distant cognitive domains (e.g., visuospatial skills) may be attributable to methodological weakness, as effect sizes tend to decrease with increasing study robustness (Sala & Gobet, 2017b, 2017a, 2020).[13] However, Bigand and Tillmann (2022) challenged these conclusions, arguing that methodological issues in Sala and Gobet's (2020) analysis led to an underestimation of the true effect. Reanalyzing the same data with corrected procedures, they found a small but significant effect of music training on general cognition (g = .234). Taken together, the mixed yet often positive findings regarding the impact of musical training on nonverbal reasoning and overall cognitive ability imply that musical expertise could plausibly improve video game puzzle-solving performance.

Music Training and Video Game Performance

Empirical research on the relationship between musical expertise and video game performance is scarce, although two studies are particularly informative. Pasinski et al. (2016) compared three groups playing *Rock Band 2* (Harmonix, 2008): trained musicians (>6 years of experience), music video game players with limited formal music education, and individuals with neither gaming nor musical experience. In the *Rock Band* series, players perform songs using a guitar-shaped controller in a way that mimics real guitar

[13] For example, Sala and Gobet (2017a) reported that the effect of music instruction dropped from d = 0.25 when compared against passive control groups to d = 0.03 when compared against active control groups.

play and must time inputs to a visual display, with higher accuracy yielding higher scores. Results showed that music video game players significantly outperformed musicians, who in turn outperformed non-musicians. While this task does not involve puzzle-solving, the findings suggest that musicians may possess a predisposition for music-based gameplay.

By extension, given that many ABPs involve musical material, it is plausible that this predisposition would generalize to ABPs. However, Hansen & Hiraga (2017) found no evidence supporting this assumption. Participants with high or low engagement in musical activities played an ABP game titled *Music Puzzle*, in which they reassembled a sound recording (containing music, speech, or a mix of the two) that had been segmented into fragments of equal duration. In addition, the pitch and/or equalization of a subset of fragments was altered, and players had to restore these properties before reassembling the track. No significant group differences emerged on multiple performance and subjective measures, such as the ratio of completed to abandoned game sessions, time taken to complete a puzzle, number of clicks on sound fragments, and subjective rating of game difficulty. Notably, group assignment was based on self-reported listening experiences; it therefore remains unclear whether music-playing expertise, as opposed to listening habits, affects performance or subjective ratings on an ABP task.

Summary and Hypothesis

Although there is little direct evidence on the relationship between musical expertise and ABP puzzle-solving performance, a large body of research indicates that musical training enhances auditory discrimination skills, and some evidence supports transfer to visuospatial abilities, nonverbal reasoning, and music-related video game performance. On this basis, the present study hypothesized that musicians would outperform non-musicians on performance measures for both the audio and visual puzzle games, with a larger group difference expected for the audio game. Overall, this study addresses a gap in our understanding of musicianship and puzzle-solving performance and contributes to ongoing discussions of transfer effects from music training to near and distant perceptual and cognitive domains.

Method

A detailed description of the methodology is provided in Chapter 4. As a brief recap, 30 participants (10 musicians and 20 non-musicians) completed the Melody, Embedded Rhythm, Timbre, and Pitch subtests of the *Profile of Music Perception Skills* (modular-PROMS; Law & Zentner, 2012; Zentner & Strauss, 2017), as well as a demographic questionnaire to collect information about age, gender, gaming habits, and extent of musical training. Participants then played five levels of the game *Circuits* and 15 levels of the game *Square Logic*. Performance on the games was assessed using three variables: *time taken*, *actions taken*, and *number of incorrect solutions submitted* (for *Circuits* only).

For analyses involving participants' self-reported weekly gaming hours, the median value of each response range was used (None = 0, 1–5 = 3, 6–10 = 8, 11–15 = 13, 16+ = 18). Weekly gaming hours spent on puzzle games were calculated by multiplying total weekly gaming hours by a proportion corresponding to each response category (None = 0, Very few = 0.10, Few = 0.25, Half = 0.50, Most = 0.75, All = 1.00).[14]

[14] As an example, if a participant reported playing video games 6–10 hours weekly, of which "Few" were spent on puzzle games, the values used in the analysis were 8 weekly gaming hours, and 2 weekly hours spent on puzzle games (8 × 0.25).

Statistical analyses were conducted using SPSS version 30. The primary hypotheses were tested using linear mixed-effects models (LMMs) to account for the repeated-measures design. Because each participant completed multiple levels per game, observations were not independent but nested within participants; this dependency was modelled using random intercepts. LMMs are also well suited to handling missing data, without requiring the exclusion of participants with incomplete data (i.e., uncompleted levels). Effect sizes for the group differences were estimated using Cohen's *d*, calculated from the estimated marginal means and the square root of the mean residual variance across levels. Incidence rate ratios (IRRs) were calculated by dividing the estimated marginal means of the musician group by the non-musician group. Power analyses were conducted in G*Power version 3.1.9.7 (Faul et al., 2007) using the linear regression F-test framework.

Results

Sample Characteristics and Counterbalancing

Several analyses were conducted to ensure the musician and non-musician groups were balanced on several key characteristics and procedural variables. Order of game presentation (*Circuits*-first vs. *Square Logic*-first) was independent of group assignment (Fisher's exact test, $p = 1.000$), confirming successful counterbalancing. Gender distribution also did not differ by group (Fisher's exact test, $p = 1.000$). Musicians (mean age = 23.1, standard definition = 4.9) were slightly older than non-musicians (19.6 ± 2.4). Because Levene's test indicated unequal variances, $F(1, 28) = 15.97$, $p < .001$, Welch's t-test was used. The difference in age was not significant, Welch's $t(11.23) = 2.13$, $p = .056$.

Average weekly gaming hours did not differ significantly between musicians (5.4 ± 6.1) and non-musicians (5.5 ± 5.4), $t(28) = 0.07$, $p = .946$. Similarly, weekly hours spent playing puzzle games were comparable across groups, with musicians (1.9 ± 3.0) playing slightly more than non-musicians (0.8 ± 1.2), though this difference was not significant, Welch's $t(10.50) = 1.16$, $p = .273$. In all subsequent analyses, order of game presentation, gender, age, weekly gaming hours and weekly puzzle gaming hours were included as covariates; none accounted for a significant proportion of variance in performance outcomes.

The four PROMS subtests showed good internal consistency, Cronbach's $\alpha = .78$ (95% CI [.65, .90]; average inter-item correlation = .47). Musicians scored substantially higher than non-musicians on the total PROMS score, $t(28) = 5.89$, $p < .001$, $d = 2.28$ (Figure 5.1). Total PROMS score strongly predicted group membership; ROC analysis indicated excellent discrimination, AUC = 0.973, 95% CI [0.92, 1.00], $p < .001$, confirming that both groups were successfully divided based on perceptual musical abilities.

Correlations Between Performance Metrics

Pearson correlation coefficients were calculated between each performance metric (*time taken to complete a level, number of actions taken to complete a level, number of incorrect solutions submitted*) for each successfully completed level (Table 5.1). Using Cohen's (1988) benchmarks, there were large positive correlations between *time taken* and *actions taken* ($r = 0.89$, $p < .001$), between *time taken* and *incorrect solutions* ($r = 0.64$, $p < .001$), and between *actions taken* and *incorrect solutions* ($r = 0.54$, $p < .001$).

The strength and consistency of these relationships suggest that the three metrics capture a shared underlying performance factor.

Figure 5.1

Results on the PROMS as a percentage of total possible score for each subtest and the combined score. Boxes represent the interquartile range; horizontal lines indicate medians; whiskers show data spread; dots indicate outliers.

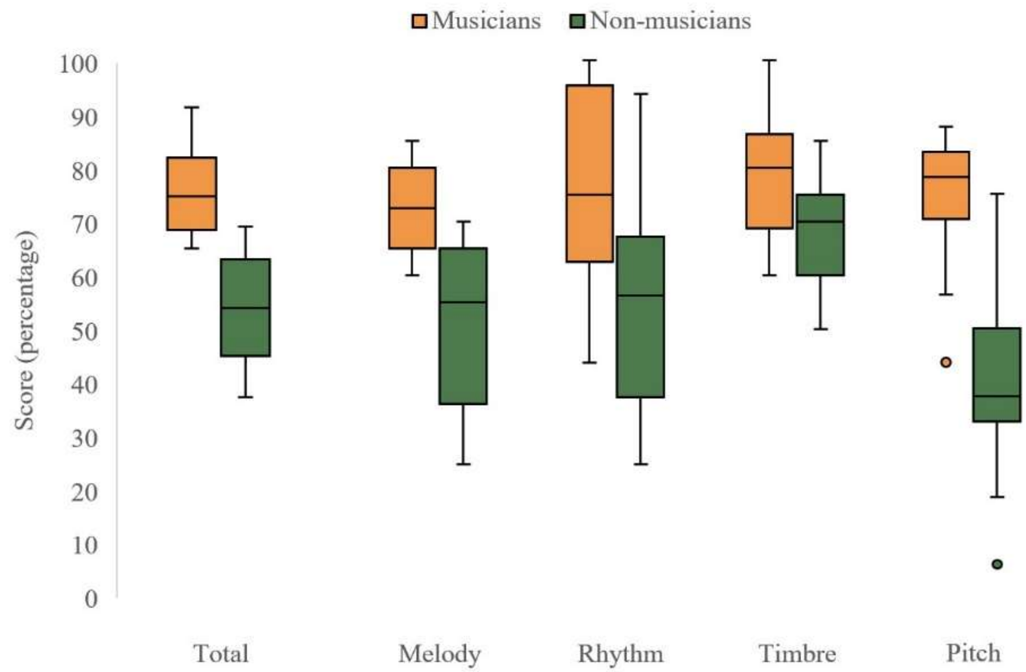


Table 5.1

Pearson correlation coefficients between performance metrics.

Variable	Time Taken	Number of Actions Taken	Number of Incorrect Solutions Submitted
Time Taken	1.00	0.89**	0.64**
Number of Actions Taken		1.00	0.54**
Number of Incorrect Solutions Submitted			1.00

† $p < .10$, * $p < .05$, ** $p < .01$

Puzzle-solving Performance in *Circuits*

Completion Rates

Completion rates across the five levels of *Circuits* differed between groups (Figure 5.2a). All musicians finished the five levels. Among non-musicians, all participants completed the first three levels; 16 of 20 completed the fourth level, and 9 of 20 completed the fifth level. Participants who did not complete Level 4 did not attempt Level 5. The following analyses were conducted on successfully completed levels only.

Group and Level Effects

Linear mixed-effects models (LMMs) were used to examine the effects of Group (M vs. NM), Level (1–5), and their interaction for all three performance metrics. Participants were included as random intercepts. Negative binomial generalized linear mixed-effects models (GLMMs) were used for both *actions taken* and *incorrect solutions* since they consist of count data and to account for overdispersion.

For *time taken*, the LMM revealed a significant main effect of Level, $F(4, 32.49) = 41.34$, $p < .001$, with completion times increasing across levels. There was also a significant main effect of Group ($\beta = -194.58$s, SE = 97.92), $F(1, 44.95) = 7.32$, $p = .010$, $d = 0.53$, with musicians completing levels faster than non-musicians. The Group × Level interaction was not significant, $F(4, 32.49) = 2.49$, $p = .063$. Completion times appeared to increase more steeply across levels for non-musicians than for musicians (Figure 5.2b).

For *number of actions taken*, the negative binomial GLMM showed a significant effect of Level, $F(4, 35.42) = 52.33$, $p < .001$, with actions increasing across levels. A significant main effect of Group was also observed, $F(1, 42.03) = 5.78$, $p = .021$, with musicians completing levels using fewer actions than non-musicians. The Group × Level interaction was also significant, $F(4, 35.42) = 3.13$, $p = .026$. Groups performed similarly at Levels 1–2 (IRR ≈ 1.00), but musicians required approximately 31–36% fewer actions on later levels (Level 3 IRR = 0.67; Level 4 IRR = 0.64; Level 5 IRR = 0.69) (Figure 5.2c).

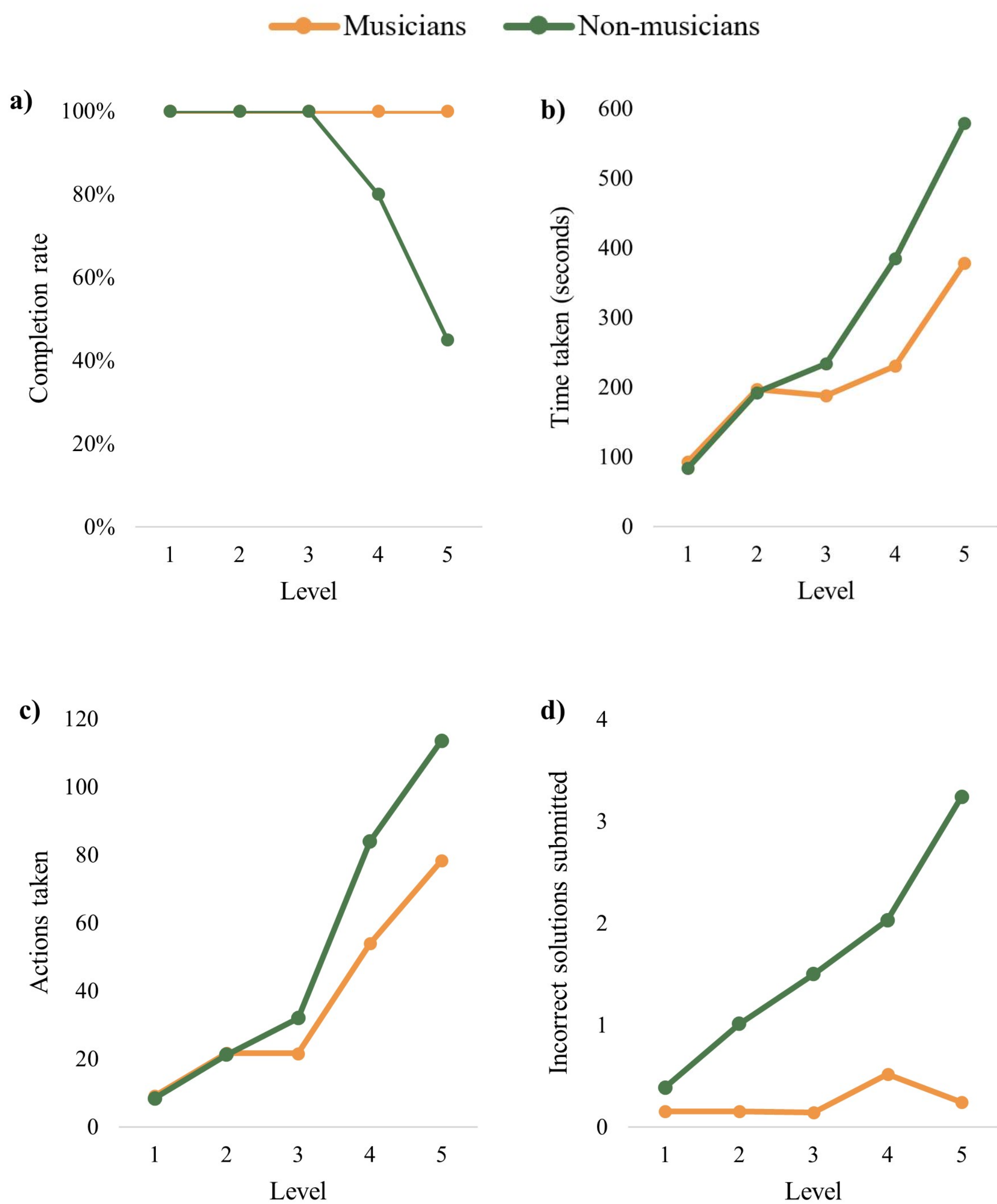
For *number of incorrect solutions submitted*, the negative binomial GLMM revealed a significant main effect of Level, $F(4, 120) = 2.54$, $p = .044$, indicating increasing error rates across levels. There was also a significant main effect of Group, $F(1, 120) = 17.47$, $p < .001$, with musicians submitting approximately 96% fewer incorrect solutions than non-musicians (IRR = 0.04, 95% CI [0.01, 0.47]). The Group × Level interaction was not significant, $F(4, 120) = .87$, $p = .483$, suggesting the group difference was consistent across levels (Figure 5.2d).

Performance and Extent of Musical Training

Within the musician group, Pearson correlation coefficients were computed between years of musical training and performance metrics on *Circuits*. There was a moderate-to-large positive correlation between *years of musical training* and *time taken*, although it did not reach significance ($r = 0.49$, $p = 0.154$), hinting at a possible tendency for participants with more extensive training to take longer to complete levels. A moderate, nonsignificant positive correlation was also observed between years of training and number of actions ($r = 0.31$, $p = .380$). No association was found between musical training and number of incorrect solutions ($r = 0.01$, $p = .983$).

Figure 5.2

Comparisons of performance metrics on Circuits*. a) completion rates b) estimated marginal means of time taken c) estimated marginal means of number of actions taken d) estimated marginal means of number of incorrect solutions submitted.*[15]



[15] Although estimated marginal means in Figure 5.2d suggest diverging error rates between groups at higher levels, the Group × Level interaction was not statistically significant ($p = .483$). This likely reflects insufficient statistical power given the small sample size, as well as the sparse data at higher levels due to few non-musician participants reaching them and submitting incorrect solutions.

PROMS Subtests and Level Differences

Pearson correlation coefficients were calculated between scores on each PROMS subtest and performance metrics for Levels 3, 4, and 5 of *Circuits* (Table 5.2). Levels 1 and 2 were excluded from the analysis due to their nature as short tutorial levels.

For Level 3, several moderate-to-large significant correlations emerged. *Number of actions taken* was negatively correlated with Melody, Embedded Rhythm, and Pitch subtest scores, indicating that participants with higher perceptual scores required fewer actions to complete the level. *Number of incorrect solutions* was negatively correlated with Timbre and Pitch scores, denoting fewer errors among participants with stronger timbre and pitch discrimination. No significant correlations were observed for *time taken*.

For Level 4, moderate negative correlations were observed between all three performance metrics and Melody scores (r = -0.35 to -0.38), reaching marginal significance ($p < .10$). Moderate-to-large significant negative correlations were also observed between Pitch scores and all performance metrics (r = -0.41 to -0.54), indicating better performance among participants with stronger pitch perception.

For Level 5, only one significant association emerged. Embedded Rhythm scores were moderately negatively correlated with number of incorrect solutions ($r = -0.50$, $p < .05$), suggesting that rhythmic perception may have supported performance in this particular level. No other correlations reached significance or exceeded $r = 0.30$. Analyses for Level 5 specifically were likely limited by the lower number of participants who reached this level.

Table 5.2

Pearson correlation coefficients between performance metrics and scores on each subtest of the PROMS.

Variable	PROMS Subtest Score			
	Melody	Embedded Rhythm	Timbre	Pitch
Level 3				
Time Taken	-0.16	-0.13	-0.29	-0.09
Number of Actions Taken	-0.59**	-0.40*	-0.31	-0.40*
Number of Incorrect Solutions Submitted	-0.24	-0.12	-0.40*	-0.41*
Level 4				
Time Taken	-0.38†	-0.32	-0.23	-0.50*
Number of Actions Taken	-0.36†	-0.21	-0.07	-0.41*
Number of Incorrect Solutions Submitted	-0.35†	-0.20	-0.33	-0.54**
Level 5				
Time Taken	-0.07	-0.18	-0.19	-0.03
Number of Actions Taken	-0.15	-0.10	-0.18	-0.02
Number of Incorrect Solutions Submitted	-0.28	-0.50*	-0.16	-0.13

† $p < .10$, * $p < .05$, ** $p < .01$

Puzzle-solving Performance in *Square Logic*

Completion Rates

All participants completed all 15 levels of *Square Logic*, with the exception of one participant who completed only the first 10 levels due to time constraints.

Group and Level Interactions

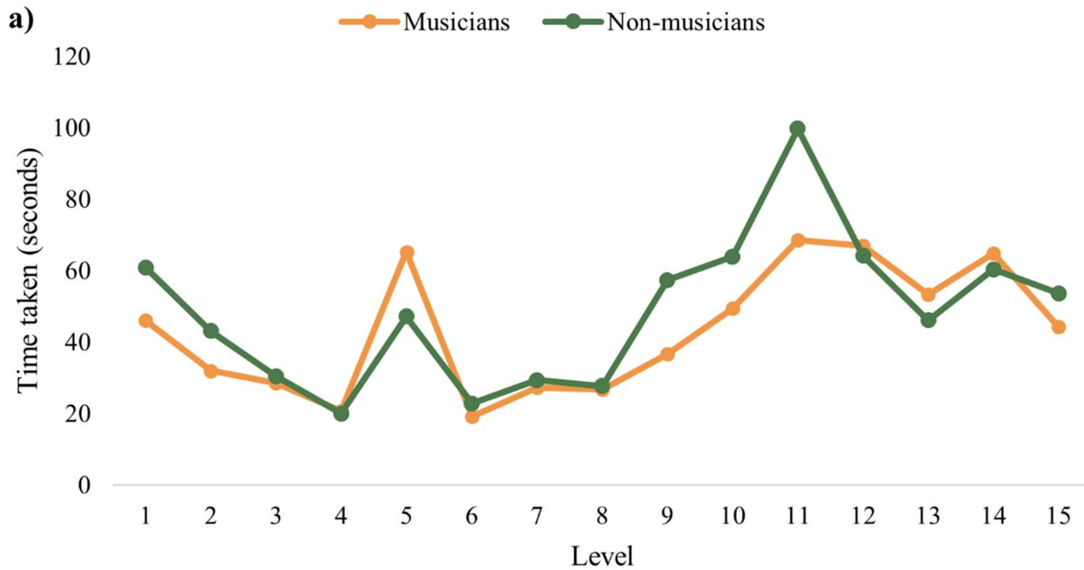
LMMs were fitted to *time taken* and *number of actions taken*, with Group, Level (1–15), and their interaction as fixed effects and participant as a random intercept. A negative binomial model was used for actions taken.

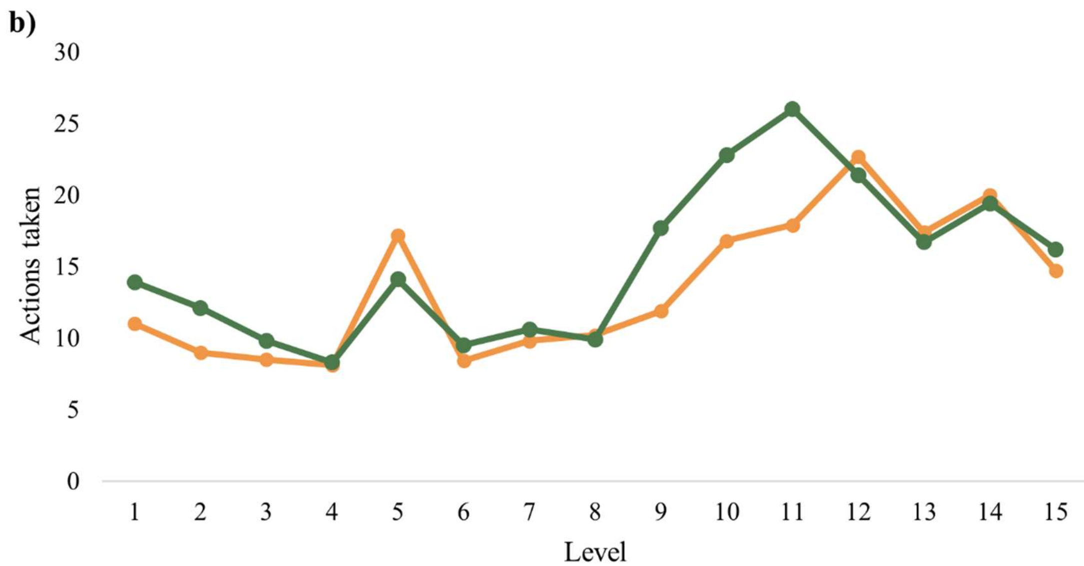
For *time taken*, there was a significant main effect of Level, $F(14, 54.04) = 14.69$, $p < .001$, indicating substantial variability in completion time across levels. The main effect of Group was not significant ($\beta = -8.93$, SE = 10.96), $F(1, 199.66) = 2.24$, $p = .136$, $d = 0.15$, nor was the Group × Level interaction, $F(14, 54.04) = 0.681$, $p = .782$, indicating comparable performance for both groups (Figure 5.3a).

A similar pattern was observed for *number of actions taken*. The negative binomial GLMM revealed a significant effect of Level, $F(14, 405.00) = 38.09$, $p < .001$. The main effect of Group approached but did not reach significance, $F(1, 405.00) = 2.864$, $p = .091$, with musicians requiring approximately 9% fewer actions (IRR = 0.91, 95% CI [0.67, 1.25]). The Group × Level interaction was also not significant, $F(14, 405.00) = 1.12$, $p = .336$. (Figure 5.3b).

Figure 5.3

Comparisons of performance metrics on Square Logic a) Estimated marginal means of time taken per level, in seconds b) Estimated marginal means of number of actions taken per level.





Enjoyment

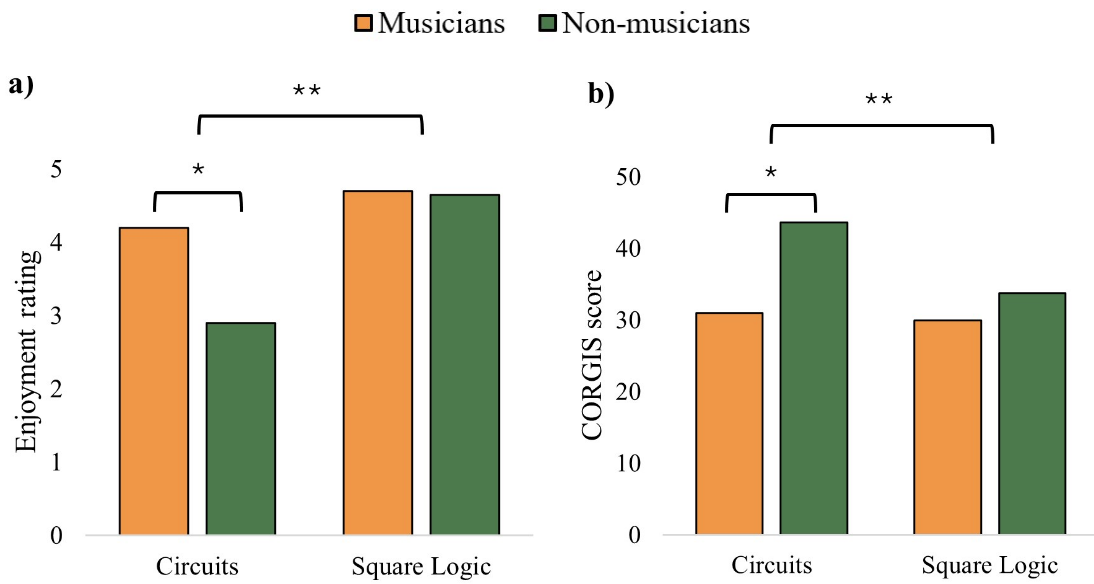
A 2 × 2 mixed ANOVA with Game (*Circuits*, *Square Logic*) as a within-subjects factor and Group (M, NM) as a between-subjects factor revealed a significant main effect of Game, $F(1, 28) = 18.81$, $p < .001$, $\eta p^2 = .402$, with *Square Logic* rated as more enjoyable overall. There was also a significant main effect of Group, $F(1, 28) = 7.42$, $p = .011$, $\eta p^2 = .209$, with musicians reporting higher enjoyment than non-musicians. Importantly, the Game × Group interaction was significant, $F(1, 28) = 5.80$, $p = .023$, $\eta p^2 = .172$. Simple-effects analyses showed that musicians enjoyed *Circuits* more than non-musicians, whereas enjoyment ratings for *Square Logic* did not differ between groups (Figure 5.4a).

Perceived Challenge

A second 2 × 2 mixed ANOVA on perceived challenge (CORGIS cognitive subscale score) revealed a significant main effect of Game, $F(1, 28) = 10.57$, $p = .003$, $\eta p^2 = .274$, with *Circuits* rated as more cognitively demanding. There was also a significant main effect of Group, $F(1, 28) = 12.84$, $p = .001$, $\eta p^2 = .314$, with non-musicians reporting higher challenge overall. The Game × Group interaction was significant, $F(1, 28) = 7.03$, $p = .013$, $\eta p^2 = .201$. Simple-effects analyses showed that non-musicians rated *Circuits* as significantly more challenging than musicians, whereas no group differences were observed for *Square Logic* (Figure 5.4b).

Figure 5.4

a) Enjoyment rating of Circuits *and* Square Logic*, from 1 to 5, for both groups.*
b) Perceived challenge of each group for both games, as measured by the cognitive subscale of the CORGIS, out of a possible total of 55.



† p < .10, * p < .05, ** p < .01

Correlation Between Key Variables

Pearson correlation coefficients between *years of musical training*, *PROMS total score*, *CORGIS scores*, and *Enjoyment ratings* are reported in Table 5.3. A large positive correlation between years of musical training and PROMS total score approached significance (r = 0.52, p = .121), likely reflecting limited statistical power within the musician group.

For *Circuits*, PROMS total score was negatively correlated with perceived challenge (r = -0.42, p = .022). There was also a positive but non-significant correlation between PROMS total score and enjoyment (r = 0.34, p = .063). Taken together, these two correlations suggest that participants with stronger perceptual musical abilities found the game less challenging and more enjoyable. CORGIS scores for *Circuits* and *Square*

Logic were also moderately correlated (r = 0.39, p = .035), indicating consistency in perceived challenge across games. Finally, perceived challenge and enjoyment were negatively correlated for *Circuits* (r = -0.41, p = .026), but not for *Square Logic*, suggesting that the relationship between difficulty and enjoyment was specific to the ABP game.

Table 5.3

Pearson correlation coefficients between years of musical training, PROMS total score, CORGIS scores, and enjoyment ratings.

Variable		Years of Musical Training	PROMS Total Score	CORGIS Score		Enjoyment Rating	
				Circuits	*Square Logic*	*Circuits*	*Square Logic*
Years of Musical Training		1.00	0.52	-0.16	-0.03	0.24	0.14
PROMS Total Score			1.00	-0.42*	-0.19	0.34†	0.22
CORGIS Score	*Circuits*			1.00	0.39*	-0.41*	0.03
	Square Logic				1.00	-0.08	0.07
Enjoyment Rating	*Circuits*					1.00	-0.03
	Square Logic						1.00

† $p < .10$, * $p < .05$, ** $p < .01$

Power Limitations

A notable limitation of the present study concerns statistical power. Post hoc power analyses were conducted to evaluate the statistical power of the Group effects across all models. Power was consistently below the conventional 80% threshold: LMM for *time taken* in *Circuits* (power = .54); GLMM for *actions count* in *Circuits* (power = .37); LMM for *time taken* in *Square Logic* (power = .06); GLMM for *actions count* in *Square Logic* (power = .26). This suggests the sample of 30 participants was insufficient to reliably detect group differences; null Group effects should therefore be interpreted with caution rather than as evidence of no difference between musicians and non-musicians. Future research should aim for a larger sample to achieve adequate power, particularly given that the observed effect sizes for Group effects ranged from small to medium.

Discussion

The purpose of this study was to investigate quantitative differences between musicians and non-musicians on video game puzzle-solving tasks. It was hypothesized that musicians would outperform non-musicians on both an Audio-Based Puzzle (ABP) task as well as on a visual puzzle task, with a larger group difference expected for the ABP game. Overall, the findings partially supported this hypothesis. Musicians significantly outperformed non-musicians on the ABP task, whereas no significant performance differences emerged on the visual puzzle task. In addition, musicians reported greater enjoyment and lower perceived challenge for the ABP game than did non-musicians.

Performance in *Circuits*

Participants completed the first five levels of the puzzle game *Circuits*. Completion rates differed substantially between groups, with more than half of the non-musician group failing to complete Level 5 (Figure 5.2a). Performance was assessed using three metrics: *time taken to complete a level*, *number of actions taken*, and *number of incorrect solutions submitted* (Figure 5.2bcd). Linear mixed-effects models revealed a significant main effect of Level for all three measures, reflecting the intended progressive increase in difficulty across levels.

Crucially, there was a significant main effect of Group for all three performance metrics, as well as significant Group × Level interactions for *number of actions taken*. These interactions indicated that although performance declined across levels for both groups, the rate of increase in number of actions was steeper for non-musicians than for musicians. Together, these findings demonstrate that musicians not only performed better overall but were also less adversely affected by increasing task difficulty. These effects were observed even though analyses were restricted to successfully completed levels. Consequently, performance in the musician group was effectively compared against the highest-performing subset of non-musicians on the fourth and fifth levels, suggesting that the actual performance gap between groups may be underestimated.

The substantial performance differences between groups are likely attributable, at least in part, to differences in auditory perceptual abilities. Musicians scored significantly higher than non-musicians on the modular-PROMS (Figure 5.1), indicating superior melody, rhythm, timbre and pitch discrimination abilities, consistent with prior research on musicians' auditory skills (Neves et al., 2022). Correlation analyses between PROMS subtests scores and performance on Levels 3, 4 and 5 of *Circuits* further suggested that each level recruited different perceptual abilities (Table 5.2). Beyond illustrating how *Circuits*' game design draws on multiple musical properties, these varied demands may have amplified group differences. Because musicians outperformed non-musicians on all PROMS subtests, they were likely better equipped to meet the diverse auditory challenges presented by the game. In contrast, non-musicians who performed well on some subtests but poorly on others may have struggled on levels that elicited the skill they were least proficient in, thereby reducing overall performance.

Participant's enjoyment of and perceived challenge in *Circuits* may also have contributed to the observed group differences. Analyses revealed that musicians found *Circuits* significantly more enjoyable and less challenging than non-musicians (Figure 5.4). According to Csikszentmihalyi's theory of *Flow*, optimal engagement arises when task demands are well matched to an individual's skills, whereas a mismatch can lead to frustration or boredom (Csikszentmihalyi, 1990). Flow theory has been previously applied to game design, emphasizing the importance of maintaining an appropriate balance between challenge and player ability at every stage of the game (Falstein, 2005; Juul et al., 2009). Given that non-musicians rated *Circuits* as substantially more challenging, it is plausible that the task exceeded their optimal challenge threshold, resulting in frustration, increased rates of abandonment, and poorer performance.

This interpretation is supported by findings from Pusey et al. (2022), who demonstrated that unmet psychological needs during gameplay (e.g., autonomy, competence, or relatedness) are associated with reduced enjoyment and motivation. Conversely, when these needs are satisfied, players exhibit greater resilience and enjoyment. In addition, McAuley et al. (2011) demonstrated that musician advantages in music perception are partly attributable to motivational factors, such that musicians perform better when confronted with tasks they find engaging and that align with their interests. Taken all together, these findings suggest musicians may have experienced greater intrinsic motivation when playing *Circuits*, which facilitated superior performance, whereas non-musicians may have felt reduced engagement thereby negatively impacting performance.

The present findings are broadly consistent with those of Pasinski et al. (2016), who reported superior performance by musicians in the rhythm game *Rock Band 2*, suggesting an affinity for music-related gameplay among individuals with musical training. However, the results do not corroborate the results of Hansen and Hiraga (2017), who found no significant differences between participants with low and high engagement in musical activities on an ABP task. One likely explanation for this discrepancy is the method by which participants were divided; Hansen and Hiraga grouped participants based on listening habits and musical engagement, whereas the present study employed musical achievement as a basis for grouping. This methodological difference may have produced groups with a larger disparity in perceptual abilities in the present study.

An alternative explanation is that participants' prior gaming experienced may have influenced results. However, post hoc analyses indicated that weekly gaming hours did not account for a significant proportion of variance in performance. Moreover, both *Circuits* and the game used by Hansen and Hiraga were designed to be intuitive and easy to learn, making it unlikely that prior gaming experience substantially influenced outcomes. Finally, it is possible that differences in game design and interface contributed to the contrasting findings. Although both ABP games involved reconstructing fragmented audio and employed direct Input/Output mechanics, differences in platform (mouse-based computer interaction versus mobile touch controls) may have altered the cognitive demands placed on players.

A more unexpected finding was the moderate, albeit non-significant, negative correlation between extent of musical training and performance within the musician group. Participants with more years of musical training tended to take longer and use more actions to complete the levels of *Circuits*. One possible interpretation is that a threshold level of musical training or auditory skill is sufficient to confer an advantage in ABP performance, beyond which additional expertise does not provide a noteworthy impact. Under this perspective, between-group differences would be driven by surpassing this threshold, whereas within-group variability would reflect individual differences or other factors such as gaming strategies. Ultimately, however, the sample size of the musician group was small ($n = 10$), and the resulting correlation was not statistically significant. Therefore, the relationship between extent of musical training and ABP performance remains unclear and would benefit from further investigations with more statistical power.

Performance in *Square Logic*

Participants completed the first ten levels of the 5 × 5 mode and the first five levels of the 7 × 7 mode of *Square Logic*. Linear mixed-effects analyses revealed a significant main effect of Level for both *time taken* and *number of actions taken*, reflecting expected variability in difficulty across levels. More importantly, there was no significant main effect of Group and Group × Level interaction for either performance metric (Figure 5.3). Thus, unlike in *Circuits*, musicians and non-musicians performed comparably on the visual puzzle task. The absence of group differences in enjoyment and perceived challenge further supports the conclusion that *Square Logic* imposed similar cognitive demands on both groups. If musical training had conferred a meaningful advantage, musicians might have reported lower perceived difficulty or greater enjoyment, as observed for *Circuits*. The lack of such differences suggests that musical expertise did not influence participants' engagement with or perception of this task.

As presented in the introduction of this chapter, evidence for the transfer of musical skills to the visual domain is mixed. While this study hypothesized that musicians would slightly outperform non-musicians on this visual puzzle-solving task, the present findings align with the broader literature suggesting that far-transfer effects of musical training are limited (Schellenberg & Lima, 2024). Together, the contrasting results for *Circuits* and *Square Logic* support the view that the benefits of musical training are more likely to reflect near-transfer effects, such as enhanced auditory perception, rather than broad improvements in distant cognitive domains. This conclusion aligns with the small positive effect ($g\Delta = 0.31$) of music training on auditory perception reported by Neves et al. (2022), but contradicts the higher pooled effect ($d = 0.40$) reported by Hetland (2000) regarding improvements across a range of spatial skills.

This conclusion is certainly not definitive, as there are alternative explanations which could explain the absence of observed differences between both groups in the visual game. As mentioned previously, the study did not have enough statistical power to detect smaller effects. It is also possible that the results reflect the specific cognitive demands of *Square Logic*. Although the task requires some visual reasoning, it might rely primarily on logical deduction rather than spatial abilities. Musical training has been associated more consistently with spatial-temporal skills such as mental rotation (Hetland, 2000; Pietsch & Jansen, 2012; Sluming et al., 2007) and increased performance on pattern recognition tests such as *Raven*'s matrices (Anaya et al., 2017; Forgeard et al., 2008; Swaminathan, 2017; Thompson et al., 2004; Trimmer & Cuddy, 2008), skills which may not be strongly engaged by this particular puzzle game. As a result, *Square Logic* may not provide a sensitive measure of the visuospatial abilities that have previously been linked to musical expertise.

Additionally, the selected levels of *Square Logic* may have been insufficiently demanding to reveal subtle group differences. Participants completed only the initial levels of both grid sizes, potentially limiting variability in performance. Moreover, participants were not instructed to optimize speed or efficiency, which may have further reduced sensitivity. Requiring explicit pre-planning and execution, such as in the *Tower of London* task (Winston et al., 2022), may have revealed differences in planning which were otherwise concealed by the game's simplistic design. However, given that significant differences emerged between both groups in the *Circuit* task under identical instructions suggests that insufficient task sensitivity is unlikely to be the sole explanation.

Conclusion

This chapter presented the quantitative findings of a study examining the influence of musical training on video game puzzle-solving performance. The results demonstrated that musicians significantly outperformed non-musicians on an audio-based puzzle task, but not on a visually based puzzle task. These findings extend prior research on musicians' enhanced auditory discrimination by showing that such advantages persist in a ludic context and translate into superior performance. In contrast, the null results for the visual task do not align with previous findings of far transfer from musical training to visuospatial and nonverbal reasoning skills, at least within the context of puzzle gaming. However, given the limited statistical power of this study, these conclusions should be interpreted cautiously.

Of course, *Circuits* and *Square Logic* represent only a tiny subset of puzzle video games. Future research should investigate whether these findings generalize to other ABPs, particularly those employing indirect mechanics (as opposed to *Circuits*' direct-only design) or mechanics from categories other than Input/Output in the ABPM framework. Further work should also explore whether visual puzzle games that more strongly engage spatial-temporal abilities reveal transfer effects associated with musical training.

As with any study, several methodological limitations should be considered. The sample was small and relatively homogenous, being primarily composed of young undergraduate students. In addition, gaming experience was not included as a recruitment criterion and was only examined in post hoc analyses. Replication studies with larger and more diverse samples, as well as designs that explicitly account for gaming expertise (and especially familiarity with puzzle games), would strengthen conclusions in this emerging area of research.

Despite these limitations, this study provides novel evidence that musicianship can influence video game puzzle-solving performance. By demonstrating significant differences in performance, enjoyment, and perceived challenge between musicians and non-musicians, this work contributes to the literature on transfer effects of musical training and sets the table for a more in-depth analysis of how players' musical backgrounds affect their experiences when solving ABPs. The following chapter builds on these findings by analyzing qualitative data from the experiment, focusing on players' spontaneous comments, strategies, and subjective experiences during puzzle-solving.

6. The Design of Audio-Based Puzzles: A Case Study and General Recommendations

The previous chapter explored performance differences between musicians and non-musicians on audio-based and visual-based puzzle games using quantitative methods. This chapter extends that investigation through qualitative analysis, examining participants' recorded gameplay and post-game questionnaire responses, as well as relevant online user reviews to identify game elements that significantly impacted player experience. Building upon relevant literature on game design and user experience, the chapter addresses a persistent challenge in Audio-Based Puzzle (ABP) design: creating experiences that are appropriately challenging and enjoyable for players across the full spectrum of musical expertise. Using the seven-step method approach developed by Visani Scozzi et al. (2017), this analysis examines participants' subjective experiences during gameplay of *Circuits* (Digital Tentacle, 2014) before offering broad design recommendations for developers seeking to create ABP experiences with wide appeal. Throughout this analysis, gaps in our current understanding of ABP design are highlighted alongside opportunities for future research.

Designing Engaging and Enjoyable Games

Game Design Research

Games User Research (GUR) and Games User Experience Research (UXR) represent interdisciplinary fields with both commercial and academic objectives, focused on understanding the relationship between game design and player experience. Research in these areas typically employs mixed methods, combining quantitative approaches (performance metrics, satisfaction scales) with qualitative techniques (interviews, case studies). Several analytical frameworks have emerged to systematize this work. The Mechanics, Dynamics, and Aesthetics (MDA) model provides a structured approach to analyzing game design through examination of these three fundamental components (Hunicke et al., 2004). The 6–11 framework extends this analysis by incorporating psychological dimensions, examining game design through six emotions and eleven instincts (Dillon, 2011, 2018). Beyond advancing theoretical understanding of factors that make gaming experiences engaging and enjoyable (Caroux, 2023), such research informs the design of educational games, enhancing their appeal to target audiences (Haworth & Sedig, 2011).

Two psychological frameworks have proven particularly influential in theorizing optimal game design: Flow (Csikszentmihalyi, 1990) and Self-Determination Theory (Deci & Ryan, 1985; Ryan & Deci, 2000). As briefly mentioned in Chapter 5, Csikszentmihalyi's theory posits that activities perceived as too easy induce boredom, while those deemed too difficult cause anxiety or frustration. When task difficulty aligns appropriately with skill level, individuals can enter a flow state characterized by complete absorption in the activity. Applied to game design, this suggests that well-designed games should provide challenges calibrated to the players' skill level, which typically increases as players develop game-specific competencies through continued play (Falstein, 2005; Juul et al., 2009).

Self-Determination Theory conceptualizes motivation along a continuum spanning from amotivation, through various forms of extrinsic motivation, to fully intrinsic motivation. The theory proposes that satisfaction of three fundamental psychological needs (autonomy, competence, and relatedness) facilitates movement toward more self-determined forms of motivation, resulting in enhanced engagement and well-being. Ryan et al. (2006) demonstrated that psychological needs satisfaction predicted both game enjoyment and preference for future play among players with varying experience levels. Notably, autonomy and competence emerged as particularly critical for single-player experiences, while all three needs contributed significantly to multiplayer engagement. Intuitive controls were found to support both in-game competence and autonomy. More recently, Pusey et al. (2022) showed that players whose psychological needs were met during challenging puzzle gameplay exhibited greater resilience and enjoyment, whereas those with unmet needs were more likely to stop playing.

Difficulty Settings

To appeal to diverse player populations, developers must design experiences that provide optimal challenge levels while meeting the psychological needs of the broadest possible audience. One fundamental approach involves offering multiple static difficulty settings (e.g., easy, medium, hard) from which players self-select. However, this solution presents limitations, as it inadequately addresses the full spectrum of player skills, learning rates, and preferences (Sepulveda et al., 2019). Consequently, some games have implemented adaptive difficulty systems that automatically adjust to match player skill levels in real time.[16] These systems commonly utilize in-game performance metrics (accuracy, completion time, survival rate), though some incorporate biometric feedback

[16] "Adaptive difficulty systems" is used here as a broad term which encompasses many approaches in which the game automatically balances its difficulty in response to external data, such as Dynamic Game Difficulty Balancing (DGDB) and Dynamic Difficulty Adjustment (DDA).

(heart rate), behavioural traits (preferences, personality characteristics), or combinations of methods for more targeted adjustments.

Research suggests that adaptive difficulty systems enhance player engagement and enjoyment (Cafri, 2025; Mi & Gao, 2025; Romero-Mendez et al., 2023; Xue et al., 2017) and can improve learning outcomes in serious games (Víteková et al., 2026). However, such systems may lead players to overestimate their abilities (Constant & Levieux, 2019) and require substantial development resources. Additionally, players may recognize and exploit these systems by deliberately performing poorly to trigger easier gameplay. While some designers advocate concealing adaptive systems to prevent such exploitation (Rollings & Adams, 2003), others raise ethical concerns about non-transparent adaptation, emphasizing the importance of disclosure (Melhart et al., 2024). Evidence also suggests that disclosing adaptive features may positively influence player immersion (Denisova & Cairns, 2019).

In-Depth Genre Research

Beyond general principles of game design and player experience, the field benefits from focused investigations of specific genres. Of particular relevance to the present study, Linehan et al. (2014) examined the rate at which challenges are presented and difficulty increases in puzzle games through analysis of four commercially successful titles. Their analysis revealed a common pacing structure: new skills were introduced through simplified challenges focused specifically on that skill, followed by progressive increases in complexity that allowed players to practise newly acquired abilities. Critically, complexity was reduced whenever new mechanics were introduced, facilitating learning. Such analyses advance understanding of effective puzzle game design, particularly regarding the creation of smooth learning curves that enhance player experience.

Given their distinctive delivery medium and role as vehicles for accessible gaming, audio games have been the subject of substantial design research. Synthesizing two decades of investigation, Urbanek and Güldenpfennig (2019) mapped key topics in audio game design research while identifying persistent gaps. Their analysis revealed strong connections between audio game design and technological advancement, with researchers increasingly exploring new platforms and input methods beyond traditional desktop gaming. The authors also noted that this line of research would benefit from greater inclusion of experienced audio game players, particularly those with low vision, to generate data more relevant to this population. Finally, Urbanek and Güldenpfennig emphasize that the actual process of audio game design remains under-researched. Although some studies document challenges and prototype iterations, most provide limited insight into development processes, a stark contrast to video game design processes, which are extensively documented through interviews and case studies.

Challenges of Designing Audio-Based Puzzles

In addition to designing for both sighted and low vision players, developers creating games centred around ABPs face the challenge of accommodating the full spectrum of listening capabilities and expertise, from individuals with limited acoustic discrimination to trained musicians. This challenge may be inconsequential for isolated ABPs embedded in otherwise non-audio-centric games (potentially even providing a moment of satisfaction for musically trained players). However, when ABPs constitute a game's primary mechanic, this design tension becomes pronounced. User reviews of

Circuits on the video game sales platform Steam illustrate experiences from both extremes of the musical expertise spectrum:

> "Nice to listen to but difficult to solve without a keen ear (which I do not have)" —darby88888888 (February 19, 2018)
>
> "Maybe I'm musically challenged, but I think it could have used some easier early levels with more obvious differences to listen for" —rpgarcher (May 2, 2015)
>
> "Circuits is a wonderful puzzle game to challenge your musical hearing abilities…if you don't already have over a decade of musical experience. […] I was able to complete all the puzzles with total ease, without using any hints or switching to individual instruments layers" —The Postminimalist (March 12, 2015)
>
> "For a musician, this game provides no real challenge or gameplay" —Sha|)ini0 (November 9, 2015)[17]

Given the substantial difficulty that music-based challenges can pose for non-musicians, some developers explicitly indicate that musical expertise is not required to play the game. For instance, *Sonority*'s (Hanging Gardens Interactive, 2022) description on *Steam* states "no musical knowledge is necessary" (*Sonority on Steam*, n.d.), an assurance echoed in multiple reviews:

> "Don't worry, you need more maths and eye than ear" —Rava (November 1, 2022)
>
> "*Sonority* is a logic-based musical puzzle game that has progressive difficulty and can be played without background or understanding of anything musical" —YQMaoski (June 24, 2022)[18]

Similarly, *Partitas* (Lotz, 2021) reassures prospective players that there is "no need to understand sheet music to play" (*Partitas on Steam*, n.d.), though reviews suggest this claim may be misleading:

---

[17] User reviews accessible at https://steamcommunity.com/app/282760/reviews/?browsefilter=toprated (Accessed 21 March 2026).
[18] https://steamcommunity.com/app/1432390/reviews/?browsefilter=toprated (Accessed 21 March 2026).

> "Unless you already are musically trained or at least have a fairly good ear for music, you will be brute-forcing this entire game" —Jawajedi (July 5, 2021)

> "I only recommend this game to someone who actually know [sic] about musical notes otherwise you have to brute force the whole game […] Partitas is more of a practice game for musicians rather than a learning game for someone with no knowledge of musical notes" —Sasan (September 14, 2021)[19]

These reviews reveal the presence of anxiety among prospective players with limited musical training when presented with ABPs, which might be warranted in cases such as *Partitas*, where claims of accessibility prove inconsistent with actual gameplay requirements. The central design question therefore emerges: How can developers create ABPs that remain enjoyable and appropriately challenging for musicians and low vision players (who are naturally drawn to auditory experiences) while maintaining accessibility for non-musicians and sighted players who may lack relevant skills? One approach to addressing this challenge can be found in Rovithis et al. (2024), who developed audio puzzles centred on speech rather than music ("spuzzles"). The premise is that these puzzles require listening skills without demanding musical knowledge, instead relying on acoustic perception and logical deduction. Participants responded positively to these experiences and generally agreed that the challenges were accessible regardless of musical expertise.

The remainder of this chapter aims to address the aforementioned design challenge of creating ABPs with broad player appeal. This analysis focuses specifically on the musician/non-musician divide rather than the sighted/blind distinction, reflecting the primary differentiating factor between participant groups in this study, though recommendations may extend to more general design considerations. Furthermore, this

[19] https://steamcommunity.com/app/1590270/reviews/?browsefilter=toprated (Accessed 21 March 2026).

research supplements existing literature by examining a commercial game rather than a prototype and making ABP design the central focus of the examination. The following section analyzes participants' experiences playing *Circuits*, with particular attention to levels which presented gameplay obstacles. Design recommendations are then provided, first for *Circuits*-specific improvements, followed by general principles for designing ABPs that appeal to a wide audience.

Analyzing the Design of *Circuits*

This section analyzes the design of *Circuits* to identify areas that could be improved, subsequently informing more general recommendations in the following section. Chapter 5 established that *Circuits*, compared to *Square Logic* (New Dawn Game Studio, 2023), was perceived as significantly more difficult and less enjoyable. Non-musicians found the game particularly challenging, resulting in some participants' inability to complete certain levels. The present analysis identifies game elements contributing to these challenges using the structured method outlined by Visani Scozzi et al. (2017). Developed to evaluate learning design in puzzle games and identify significant gameplay obstacles, this mixed-methods approach consists of seven steps, some of which were completed in previous chapters but are briefly restated here for clarity.

Step 1: Chart Learning Curve of the Game

The first step involves charting individual puzzle complexity to visualize progression across levels. Following Linehan et al. (2014), complexity is defined as the number of player actions required to solve a level, with higher numbers representing greater complexity. This visualization not only illustrates complexity progression but also provides a baseline for comparison with players' actual actions counts. Applied to

Circuits, optimal action represents a playthrough in which players listen once to the reference track for each layer, then once to each node, which is a more realistic baseline than a theoretical minimum number of actions that can be achieved through chance.[20] *Circuits*'s learning curve is shown in Figure 6.1.

As discussed previously, Linehan et al.'s analysis revealed that well-received puzzle games typically reduce complexity when introducing new concepts, allowing players to learn and practise new skills. *Circuits* appears to violate this principle in levels 3 and 5. Level 3 introduces the layer functionality, enabling players to isolate individual instrumental tracks to facilitate the task at hand. Level 5 introduces two new mechanics: decoy nodes (available nodes that do not belong in the track) and doubled nodes (identical nodes appearing twice because a specific pitch occurs in two positions). Yet both levels require more actions to solve than preceding levels and involve placing more nodes overall, increasing complexity while simultaneously introducing new elements. The impact of this design choice is most evident in player responses to level 5, discussed in detail in a later step.

Step 2: Recruit Participants, Analyze Player Expertise

The second step involves recruiting participants to play the game. The original method recommends recruiting participants with diverse gaming experience, subsequently categorizing them (e.g., novice, intermediate, advanced) to examine how gaming skills influence player experience. This separation acknowledges that players at different skill levels may have distinct learning needs. Here, participants were divided

[20] As a reminder, the goal of *Circuits* is to reassemble a track that has been fragmented into several pieces ("nodes"). A more detailed explanation of the game as well as examples of some of its levels can be found in Chapter 4.

into two groups: “gamers” (≥6 hours weekly, n = 11) and “non-gamers” (<6 hours weekly, n = 19). Since this study did not explicitly control for gaming experience during recruitment, but rather assessed expertise through self-reported weekly gaming hours, this specific analysis will be limited. As in Chapter 5, participants were also categorized as musicians (“M,” n = 10) or non-musicians (“NM,” n = 20), given its relevance to ABP performance.

Step 3: Gameplay is Recorded in a Rich, Multi-modal Manner

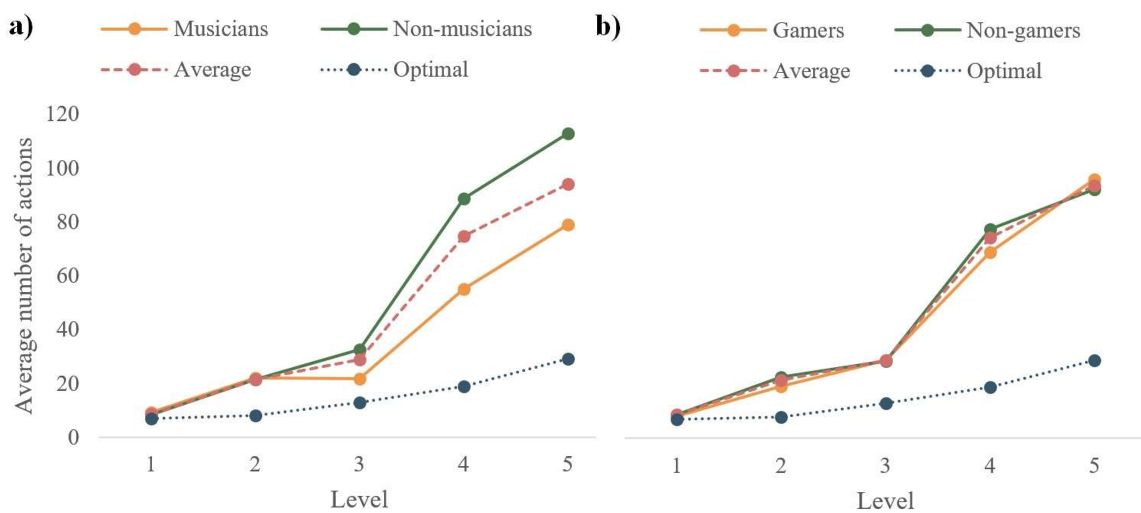
The third step involves recording gameplay data. The authors recommend capturing both quantitative data points (quantifiable player-game interactions such as number of actions) and qualitative data on player experience through methods such as concurrent or retrospective think-aloud protocols. As described in Chapter 4, this study recorded multiple data points, of which several are of relevance to this particular analysis: number of actions per level, gameplay footage, concurrent think-aloud commentary, and post-game questionnaire responses.

Step 4: Analyze Participant Performance

The fourth step involves analyzing collected quantitative data to assess player performance. Each player’s performance is compared to optimal solutions by plotting action counts on the same graph, either individually or as statistical group means. Figure 6.1a compares musician and non-musician group performance with the overall participant average and optimal solution. Figure 6.1b does the same but divides participants along gaming habits. As established in Chapter 5, a clear divide exists between musician and non-musician performance. As expected, given that gaming habits were not found to be a significant covariate in the quantitative analysis, this divide is almost absent when comparing the gamer and non-gamer groups, with a small gap appearing in level 4 but disappearing in level 5. This non-effect may reflect limitations in the methodology for recruiting and categorizing participants by gaming expertise and should be interpreted cautiously.

Figure 6.1

Complexity of the first five levels as measured by the optimal number of actions and compared to the participants' average number of actions, with a) performance of musicians (n = 10) and non-musicians (n = 20); b) performance of gamers (≥6 hours weekly, n = 11) and non-gamers (<6 hours weekly, n = 19).



Step 5: Identify Game Sections Where Problems Have Arisen

All groups followed similar trajectories and performed relatively close to optimal solutions through Level 3. At Levels 4 and 5, however, a steep increase in all players' performance relative to optimal solutions becomes evident and is particularly pronounced among non-musicians. Additionally, in the non-musician group, while the first three levels were completed by all participants, Level 4 had an 80% completion rate and Level 5 had a 45% completion rate. These data points indicate that both levels contain design elements that make them particularly challenging to solve, warranting focused analysis. The examination therefore centres on these two levels.

Step 6: Conduct Qualitative Analysis of Those Problems

The sixth step involves close analysis of sections where players' gameplay was hindered and examining related qualitative data. These gameplay segments are then coded according to the nature of the gameplay "breakdown" and strategies employed to overcome them:

> Breakdowns are categorized as relating to *action* (when a player fails to execute an action successfully), *understanding* (when the player is confused what to do next or is suffering from a misunderstanding) and *involvement* (when they appear bored or frustrated). Breakdowns are classified as '*major*' when recurring across multiple players or when they significantly obstacle player progress (e.g., the player 'dies' or is stuck for an extended time period). All other instances are '*minor*' breakdowns (Visani Scozzi et al., 2017, p. 5)

While these codes provide an effective starting point, adaptation is necessary for this game and study context. First, player action breakdowns regarding execution are essentially nonexistent, as *Circuits* was selected specifically for its ease of use (requiring only mouse control) and lacks sections with difficult manual execution challenges such as timing-based or navigational obstacles. However, this analysis expands the definition of "action breakdown" to encompass moments when players understand objectives but cannot accomplish them. This modification preserves the spirit of the original definition while better fitting the game's context.

Considering major breakdowns as moments when players could not progress and ultimately abandoned the game, such breakdowns occurred in Levels 4 and 5 and can mostly be attributed to player understanding or action. Regarding player understanding breakdowns, in Level 4, three players never listened to the reference track while attempting to reconstruct the puzzle. These players successfully completed the first three levels through trial and error due to the levels only containing a small number of nodes to be placed. Given limited verbal commentary during gameplay, determining the exact cause of these understanding breakdowns proves difficult. It is likely that they stem partly from inadequate tutorial engagement in Level 1, which goes over basic game mechanics. In their post-game questionnaire, NM19 commented: “The game was a bit confusing, I wasn’t sure where to place the high and low pitches.” It is also plausible that players may also have misconceived the game’s objectives. Indeed, some may have believed the game was about music composition, with the ultimate goal of creating pleasant-sounding arrangements, rather than replicating the reference track—a disappointment expressed in some Steam reviews:

> “I thought it was a game about making music, it is not. Is about trying to identify patterns” —OCocoMágico (November 17, 2015)
>
> “Not what I was expecting at all, and not in a good way. This isn’t a game about music making, it is an acoustic jigsaw puzzle […] If my arrangement sounds good to me, that should be enough!” —Literally Gilda (April 27, 2014)[21]

Level 5 also featured major breakdowns attributable to player understanding. While two players understood the objective of reproducing the reference track, they failed to utilize the layer function that would have allowed them to listen to isolated instruments, which significantly simplifies the task. For participant NM7, the source of this misunderstanding remains unclear, as they successfully employed the layer function when it was introduced in Level 3 but did not use it in the following levels, possible having forgotten its availability. For participant NM9, this breakdown apparently stemmed from rushing through the Level 3 tutorial explaining the layer functionality,

[21] https://steamcommunity.com/app/282760/reviews/?browsefilter=toprated (Accessed 21 March 2026).

resulting in never switching layers subsequently. Despite this additional obstacle, both participants were able to complete Levels 3 and 4, but the task proved insurmountable in Level 5, where multiple overlapping layers made identifying relevant musical information too challenging.

NM9 repeatedly expressed during gameplay that tasks were extremely difficult and that they struggled to differentiate between fragments. For instance, in Level 5: "In the melody, to me, it sounds like we're repeating the same thing eight times."[22] This can be construed as a player action breakdown, numerous instances of which occurred in Levels 4 and 5 for multiple players. These players clearly understood game mechanics yet experienced difficulty replicating the reference track due to inability to discern differences between fragments. Participant NM6 encountered several minor breakdowns throughout the first three levels, stating multiple times that differentiating specific fragments proved challenging. This became an insurmountable obstacle in Level 4 when tasked with placing five notes in the melodic layer, where they stated: "I can't see the difference between all the pink circles. I think they're all the same," a challenge echoed by participant NM17, who similarly struggled with reordering the melodic fragments.

NM3 and NM10 faced comparable challenges in Level 5, stating several times during gameplay that distinguishing between fragments proved difficult. In their post-game questionnaire, NM10 stated: "I found it very difficult to remember the exact sound I was hearing since a lot of them were very similar to me." This suggests players struggled not only with fragment differentiation but also with sound memorization for comparison purposes. From a Flow perspective, these action breakdowns indicate that

[22] "Dans la mélodie, pour moi, ça sonne comme si on répétait la même chose huit fois."

challenge level exceeded skill level to such a degree that players could not enter a flow state. The inability to perceive distinction between fragments despite understanding what was required represents a fundamental skill-challenge mismatch, causing frustration and anxiety.

These negative emotions were also witnessed in involvement breakdowns, though classifying them as "major" proves tricky, as the study context may have provided extrinsic motivation that sustained engagement (e.g., contributing to research, wanting to impress or to not disappoint). Player involvement breakdowns appeared to be far more common among non-musician participants, which could be expected given that on average, this group found the game more challenging and less enjoyable. However, player involvement breakdowns could also be observed in the musician group, particularly immediately after submitting an incorrect solution, or in the case of M7, after misunderstanding the game's instructions, which led to confusion. These expressions of frustration and discouragement also suggest compromised competence needs, as described by Self-Determination Theory. The lower enjoyment score reported by non-musician participants may reflect the absence of the positive affect and persistence that characterize intrinsically motivated play. In addition, several participants occasionally mistook the button which submits the final answer for the button that plays the reference track, which caused irritation. Since intuitive controls contribute to feelings of in-game competency and autonomy (Ryan et al., 2006), this confusing aspect of the design may have impeded the fulfillment of players' psychological needs.

While not a primary objective of the seven-step method, several puzzle-solving strategies were observed during review of gameplay footage and are worth discussing here. Indeed, players employed at least three distinct approaches to reconstructing pitch-based segments, involving comparison of a) individual nodes, b) intervals between nodes, and c) melodic lines. These strategies are illustrated in Figure 6.2. Both musicians and non-musicians were observed humming notes and tapping rhythms to assist with memorizing individual nodes or sequences. While quantitative comparison of strategies between groups would provide valuable insights, too many confounding factors exist in this study to yield robust results. Participants' preferred strategies were often ambiguous and had to be inferred from gameplay, as they rarely mentioned them explicitly in commentary. Furthermore, awareness of being recorded and the requirement to provide concurrent commentary may have influenced behaviour. This research area would benefit from studies with refined focus on player strategy, possible employing retrospective think-aloud protocols or post-game interviews with the express goal of identifying employed strategies.

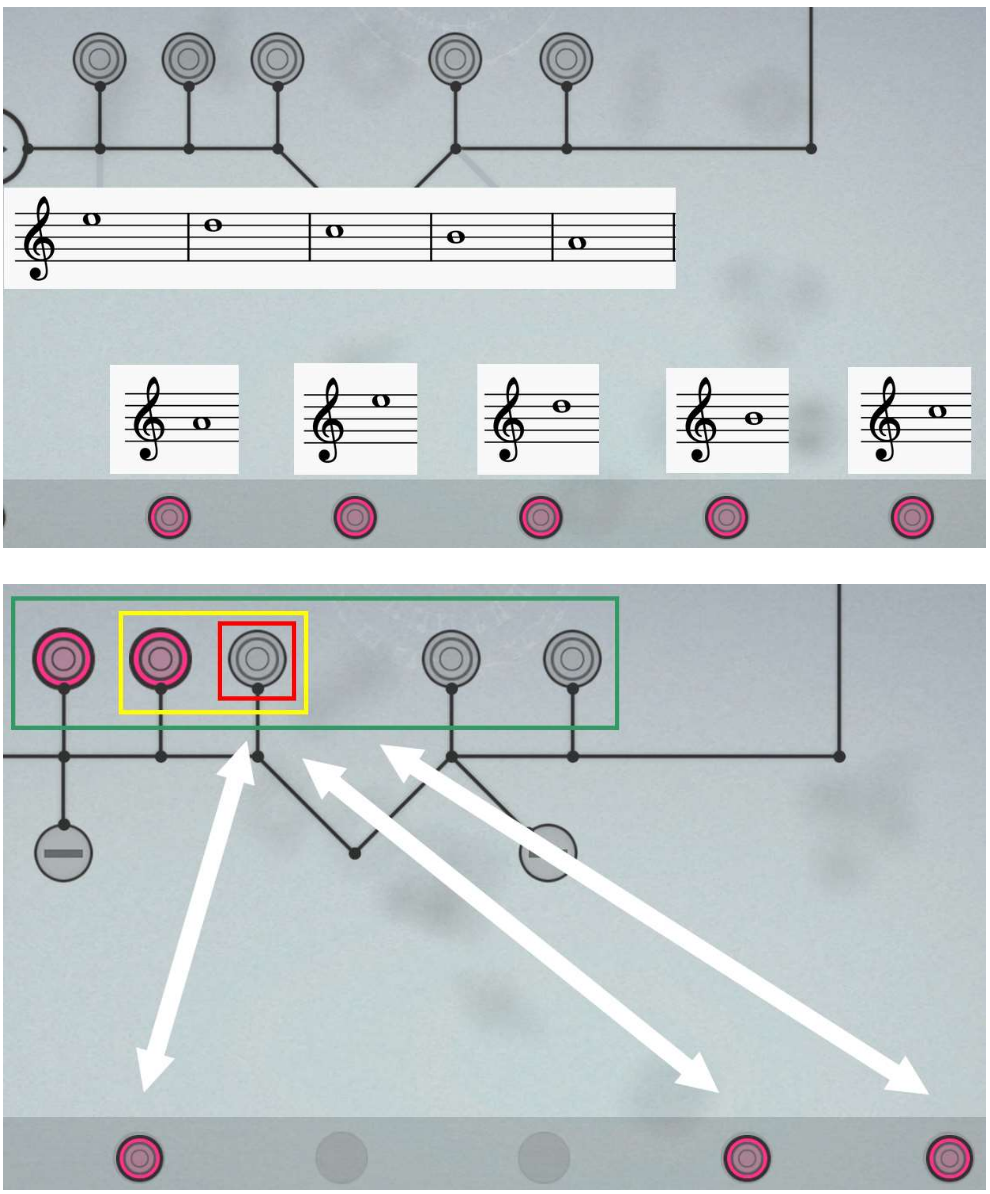
Nonetheless, one clear difference between musician and non-musician groups concerned language used to describe game events and puzzle-solving approaches. Musicians were far more likely to employ specialized terminology (e.g., "major/minor", "semitone", "suspension") to describe the musical content of the game and to rely on external preconceptions of expected chord or melodic progressions to inform their puzzle-solving. Non-musicians employed more varied language to describe musical events, including conventional descriptors like "higher" and "lower" for pitch, but also qualifiers such as "intense," "dramatic," or "warm." Participant NM6 compared musical fragments using the terms "soft" and "harsh," while NM20 used "hopeful" and "sombre." NM5 described some fragments as sounding more "horizontal" or "vertical" than others.

Both NM10 and NM20 likened melodies to staircases; NM10 referencing the stepwise melody of Level 4, and NM20 the descending arpeggios in Level 5.

Research has documented how cognitive representations of musical properties and descriptive language vary by culture. While mapping pitch onto vertical space (high/low) predominates in Western culture, some cultures map pitch onto alternative domains such as small/large, thin/thick, or young/old (Dolsheid et al., 2015; Eitan & Timmers, 2010). Studies demonstrate that people employ diverse terminology to describe music, organizable into categories such as musical features (instruments, form), movement descriptions (running, flying), emotions (happy, sad), or nature metaphors (plants, animals), which are influenced by culture and language (Cancino-Chacón et al., 2020; Kim & Belkin, 2002; Mizener et al., 2022). Lidji et al. (2007) showed that musicians were more inclined than non-musicians to map pitch onto a horizontal axis in addition to the common vertical mapping, demonstrating that musical expertise can also affect cognitive representation of music. As presented above, variety in music descriptors was observed in participants' verbalized thoughts, though determining the factors influencing these choices with certainty is not possible here. While collecting such data was not an aim of this study, participants' commentary while solving ABPs yielded intriguing insights into how musicians and non-musicians conceptualize music. This study therefore suggests that ABPs could provide fertile ground for systematic explorations of sound and music conceptualization.

Figure 6.2

Top: notational representation of the melodic layer of level four in Circuits (note lengths not to scale).
Bottom: Illustration of three observed strategies in Circuits, using how players find the third note in the melody as an example. The red box illustrates players who determine which note belongs in the node by comparing it with each individual possible option. The yellow box illustrates players who determine the note by comparing the intervals between the previous note and each possible option. The green box illustrates players who determine the note by memorizing the entire melody (or a segment of it) and trying to replicate it using the possible options.



Step 7: Offer Design Recommendations

The final step involves formulating recommendations to address the causes of identified gameplay breakdowns. As noted in Step 1, *Circuits*'s learning curve does not follow design principles of other successful puzzle games, as level complexity increased throughout the first five levels without allowing players to get accustomed to newly introduced mechanics. Applying Self-Determination Theory to these observations, it becomes clear that *Circuits* failed to satisfy several players' fundamental psychological needs. The steep difficulty increase threatened competence needs by making success unattainable for many non-musicians. The lack of adequate difficulty customization options limited autonomy, as players could not adjust the experience to their skill level. In their post-game questionnaire, participant NM9 stated:

> The first few levels seemed easy to be whilst maintaining the perfect challenge ratio. The 4th and 5th levels were extremely hard for me as they had many layers in the melodies and many very similar notes to choose from. I felt genuinely discouraged and unintelligent when trying to complete the two last levels since I thought I was right all the times I submitted my answers, just to be proven wrong repeatedly. The 3 first levels that I deemed fun, easy yet fairly challenging made me feel smart and gave me confidence for the next rounds.

This testimony provides direct evidence of unmet competence needs in the last two levels. Notably, NM9 contrasts these negative feelings with levels 1–3, which made them feel smart and provided confidence. This shift from met to unmet competence needs coincides precisely with the complexity spike identified in Step 1, supporting the interpretation that poor difficulty calibration disrupted the psychological conditions necessary for enjoyable gameplay. Similarly, from a Flow perspective, levels 1–3 maintained the challenge-skill balance necessary for engagement, while levels 4–5 pushed beyond some players' capabilities into the anxiety zone.

However, does this necessarily indicate that the game was designed with inappropriate difficulty? Several musician participants expressed in post-game questionnaires that they did not find the game particularly challenging:

> "I would have found the task more complex (and realistic) if I had had to find the circles only from listening to the full track; that would have added a challenge. Being able to do it separately made the task easier and only required paying attention, without too much analysis"[23] —M4

> "I did not find it particularly challenging however it was only the first 5 levels" —M5

> "I found it easy especially since I have perfect pitch, so I was able to match notes to the correct ones based on that" —M8

> "Easy because I could split it up by instrument; it may have been more challenging if I was forced to listen to all at once" —M9

These responses suggest that *Circuits* faces the inverse problem with expert musicians: challenge levels that fail to reach the threshold necessary for flow and falling below the necessary complexity to fully engage the players' capabilities. The dichotomy between the responses and experiences of both groups exemplifies the core tension in ABP design discussed in the introduction: developers must create experiences that engage people with varying levels of musical expertise while providing an appropriate challenge for all.

For *Circuits*, there are several solutions that could ease this tension. First and foremost, adjusting the difficulty curve to ensure early levels introduce game mechanics and provide sufficient time for learning, integration, and effective application. This would entail reducing necessary actions and node counts in early levels, specifically those introducing new mechanics, following design principles of successful puzzle games

---

[23] "J'aurais trouvé la tâche plus complexe (et réaliste) s'il avait fallu trouver les cercles qu'en entendant le résultat total, ça aurait ajouté un défi. En pouvant y aller séparément, la tâche était plus facile et demandait seulement de porter attention, sans trop d'analyse requise."

(Linehan et al., 2014). While this might render early levels too easy for individuals with greater auditory discrimination skills, they would simply progress to more complex levels more quickly, while still allocating adequate levels and time for those requiring more practice. In ABP contexts, reducing early-level complexity also involves not only reducing action counts but also facilitating stimulus differentiation (e.g., increasing pitch distances or making rhythmic patterns and timbres more distinctive).

In Level 6, *Circuits* introduces an option to identify misplaced nodes. In Level 7, an option to automatically place one node correctly is also made available. Introducing these options earlier, ideally at the game's outset, would allow players to progress past levels presenting insurmountable challenges. This simple change would enable players to customize difficulty and adjust it to their needs and preferences. Interestingly, *Circuits* does provide a form of difficulty customization by allowing players to choose whether or not to isolate layers when solving puzzles. Some musician participants noted this made the game too easy; however, participant M6 deliberately avoided isolating layers to self-impose an additional challenge. *Circuits* rewards players who complete all 25 levels without switching to isolated layers with the "Golden Ear" achievement. Perhaps, by making it explicit from the beginning that players can choose to play on this "hard mode," more musicians would have opted to avoid using the layer functionality, rendering the game more appropriately challenging. In any case, difficulty customization options from the outset would grant players greater agency over the game's difficulty, improving overall experience.

Despite the design flaws described here, the game clearly resonated with many musician and non-musician participants. Participant NM9, who described Levels 4 and 5 as extremely difficult in their post-game questionnaire, also stated they "found the tunes and melodies to be extremely entertaining and good" and that "the purely musical part of it was super enjoyable." Participant NM7 similarly stated, "I found the game hard, but fun when you understand it." On Steam, *Circuits* boasts a "very positive" rating, with 89% positive reviews, most praising its minimalistic design, relaxing atmosphere, and musical experience. Even among negative reviews, many players praise the soundtrack and/or concept despite finding the execution or gameplay lacking:

> "Interesting idea, visually well done, nice songs. However, even when it gets complex and more and more elements get introduced to the game, it remains a matter of diligence to solve the songs and I got bored"
> —n4bb12 (November 29, 2015)
>
> "Interesting gameplay mechanics, but ultimately wasn't fun enough to keep me playing until the end" —BinarySplit (July 11, 2014)[24]

A reasonable conclusion from these comments is that ABPs can be enjoyable for a broad spectrum of individuals with varying musical training, even despite excessive challenge levels, and that demand exists for audio-based gameplay experiences beyond traditional rhythm games and electronic instrument games, which ABPs can provide. However, just like other games, ABPs require careful design to provide captivating experiences that avoid causing frustration or boredom. Using conclusions from the analysis of *Circuits* and insights from relevant literature, the following section formulates design recommendations for creating appealing ABPs.

Design Recommendations for Audio-Based Puzzle Games

ABPs remain relatively unpopular compared to other puzzle and music-based game experiences, both as a primary element of commercial games and as academic

[24] https://steamcommunity.com/app/282760/reviews/?browsefilter=toprated (Accessed 21 March 2026).

research subjects. As such, limited data exist to guide ABP design. Game designers seeking to create ABP experiences operate with little guidance and may be more likely to produce games with design flaws. This, in turn, risks diminishing player experience. Developers observing poor reception of such games may then avoid creating them. This chapter (and thesis) aims to disrupt this feedback loop by contributing to ABP research and facilitating the creation of engaging puzzles. The following paragraphs explore design considerations for making ABP games more appealing to diverse prospective players with varied listening abilities and accessibility needs, supplementing more general design recommendations for audio games (Garcia & de Almeida Neris, 2013) and accessible games (Oliva-Zamora & Mangiron, 2025).

Adjusting the Difficulty

As discussed in the introduction, appealing to broad audiences requires designing games with appropriate challenge levels for people with diverse skills levels and, in ABP contexts, musical expertise. A solution implemented in many games involves offering preset, static difficulty modes for player selection. This is certainly feasible in ABPs, as *Circuits* demonstrated with its unofficial "hard mode," and its option to offer hints and assistance for players who are stuck.

While static difficulty modes represent an easily implementable approach to addressing player skill ranges, adaptive difficulty systems ensure more appropriately customized challenge levels matched to player expertise, providing more engaging experiences. For ABPs, implementation methods will depend on puzzle nature and game mechanics. One broadly applicable approach involves adjusting properties of sound parameters most prevalent in the puzzle's design. For instance, in pitch-based puzzles, distance (i.e., frequency) between notes can be decreased or increased to make differentiation easier or harder. Rhythmic patterns can be made more distinct by increasing pause lengths or using different time signatures. Spatial puzzles can adjust stimulus loudness, and so forth.

For more specific examples of difficulty settings, consider the serious game *iClef* (Baratè & Ludovico, 2013), in which players identify notes on a staff by pressing the corresponding key on a virtual keyboard. The game gradually increases complexity by introducing wider intervals between consecutive notes, accidentals and double accidentals, ledger lines, and less common clefs. Melodies are procedurally generated at each level and complexity is adjusted based on player performance. Additionally, players are only allowed to progress to the next level if the number of mistakes remains below a certain threshold.

Another potential avenue for difficulty adjustment involves modulating use of direct and indirect mechanics within puzzles. Employing mixed mechanics (i.e., presenting auditory stimuli both in audio form and in another modality) would likely lower puzzle challenge levels by providing multiple interpretive cues. Conversely, puzzles employing only direct mechanics could prove more challenging for most players with undeveloped auditory perceptual skills. However, this hypothesis requires formal investigation; introducing redundant indirect mechanics across modalities might impose unnecessary cognitive load, adversely affecting comprehension and performance (Reeves et al., 2004; Wuertz et al., 2018). Direct and indirect mechanic configurations could be adjusted automatically through adaptive difficulty systems or toggled manually by players. By making information available through multiple modalities, mixed mechanics also enhance accessibility for players with disabilities.

Relating these solutions to player psychology, adaptive difficulty systems directly address the theoretical imperatives identified by both Flow Theory and Self-Determination Theory. By continuously adjusting challenge to match developing skill levels, such systems can maintain the challenge-skill balance that Flow Theory identifies as essential for optimal experience. Simultaneously, by ensuring tasks remain achievable, adaptive systems support the competence needs Self-Determination Theory identify as critical for intrinsic motivation and sustained engagement. However, potential trade-offs of implementing such systems include likely increased development costs, as well as the potential for a reduction in players' sense of achievement, highlighting the need for further investigation.

Stimuli Considerations

Another crucial consideration when designing puzzles with broad appeal concerns the nature of the employed stimuli. A recurring theme in ABPs involves using real-world notation or other music-related topics and objects, which adds authenticity and, in serious game contexts, offers potential to teach music principles. However, for broad appeal, using real-world musical material exacerbates the tension of designing for musically trained and untrained individuals, as those with music education will likely be familiar with the employed stimuli and therefore have an advantage over other players, complicating puzzle challenge level adjustment. Many non-musicians encounter difficulty with these puzzles and rely on other familiar systems rather than engaging with puzzles in presumable intended ways. For instance, some user-created walkthrough videos meant to assist struggling players translate musical notation or stimuli to numbers for easier comprehension by non-musicians.[25] *Sonority* even allows users to switch between letter-based notations or numbers to represent notes, though the use of letter-based notation is encouraged by the developers when starting the game.

One approach to addressing this concern involves using stimuli unfamiliar to both musicians and non-musicians, lessening any advantage one community might possess. This is exemplified in the "spuzzles" from Rovithis et al. (2024) mentioned earlier; by using speech as stimuli, these puzzles likely become more accessible to non-musician audiences while proving more challenging to musicians than music-based puzzles. However, many ABPs naturally revolve around musical stimuli. For such puzzles, particularly those employing notation systems for solving, a possible solution consists of using fictional or abstract notation rather than conventional notation.[26] Fictional notation systems have the advantage of being foreign to both musicians and non-musicians; nevertheless, musicians might retain advantages by applying common rules and principles from real-world notation systems potentially present in fictional systems. Abstract notation retains the advantage of being foreign to both communities while being sufficiently different from conventional notation systems that any residual musician advantage is lessened or eliminated.

When designing ABPs, another important consideration involves relationships between puzzle stimuli and well-researched cognitive phenomena. One obvious example

[25] See, for instance, walkthroughs for the melody bay puzzles in *Super Mario RPG* (Gamerpillar, 2023) or the water harp puzzle from *Remnant 2* (Legacy Gaming, 2023).
[26] *Conventional notation* refers to real-world established practices. *Fictional notation* refers to systems that only exist within a game world. *Abstract notation* refers to methods of depicting sonic events that do not represent specific musical parameters. A discussion regarding these notational systems, as well as examples of each, can be found in Chapter 3.

concerns the connection between loudness and proximity, where sounds originating from closer sources are louder, and which is ubiquitous in puzzles employing direct spatial mechanics. Another example involves pitch, typically mapped onto vertical or size continuums. For instance, in *Sonority*, platforms rise when players perform ascending intervals and lower with descending intervals. In *The Witness* (Thekla, Inc., 2016), players reproduce sound sequences by ordering dots of varying sizes, with larger dots representing lower-pitched sounds.

While these design decisions appear intuitive, research examining how such decisions affect puzzle-solving experiences is rare. As discussed earlier, cognitive representations of music vary by culture; it remains unclear whether these design decisions impact the performance and approach to puzzle-solving of players whose cognitive mappings of sound differ from how the game presents those sounds and allows the player to manipulate them. While the listed examples represent simple mechanics to which players might quickly adapt, more complex designs making multiple assumptions about intuitive design could impose longer adaptation periods that might negatively impact some players' experiences.

Moreover, less-researched cognitive mappings present additional challenges for informed design decisions. Again, in *Sonority*, ascending intervals also have the ability to rotate platforms clockwise, and descending intervals can rotate them counterclockwise. While evidence exists that individuals embed pitch in multi-dimensional spatial structures (Deutsch et al., 2007), with some theoretical models using spirals or helixes to represent pitch chroma, pitch height, and key relationships (Chew, 2014; Shepard, 1964), apparent evidence of cognitive association between pitch and clockwise/counterclockwise movement is arduous to find. Consequently, whether this particular *Sonority* design choice impacts puzzle-solving is unknown. Perhaps, when possible, a safer design decision would be to allow the player to customize directionality of pitch-based game mechanics.

There are additional design decisions that can be informed by music cognition research. For instance, puzzle designers should consider the average number of tones individuals can maintain in working memory when performing tasks, which varies according to stimulus-based factors such as tonal context and musical structure (Ding et al., 2018; Dowling, 1978; Li et al., 2013). In addition, given that auditory stimuli are temporal, designers must make abundantly clear when looping sound sequences begin and end to prevent confusion.[27]

While a case was made here for intuitive design, it is entirely possible that games employing counterintuitive design could prove more interesting and enjoyable to players precisely because they violate fundamental assumptions about game operation or sound representation. For instance, Chapter 2 provided examples of negative spatial mechanics, in which the goal is to move away from sound sources, contradicting the typical video game instinct to investigate new sounds in favour of seeking silence. Further investigations into this topic would prove considerably valuable to make informed recommendations. Once again, the point warrants emphasis: numerous unexplored design opportunities exist in the realm of ABPs, offering rich veins of innovative gaming mechanics and experiences.

[27] See, for instance, user *daynthelife*'s Reddit post regarding some of the puzzles from *The Witness*: "My issue with the shipwreck puzzle (and to an extent all the audio puzzles) is this: when does the audio loop start? If the audio goes 'high-low-middle', how am I to tell that from 'low-middle-high'?" (daynthelife, 2022).

Music Selection

Another crucial aspect of ABPs, specifically music-based puzzles, is that players seek games with music they find enjoyable. As participant NM9's earlier testimony suggests, players who find gameplay difficult but enjoy musical elements might persevere through levels or games they do not find particularly gratifying if they connect to the soundtrack. Conversely, players might enjoy puzzle aspects of a game but dislike the music so much that they decide to abandon it, as some of *Circuit*'s Steam reviews suggest:

> "If the music was better, it might have been worth it. It wasn't bad, but it wasn't good enough to justify the rest of the game" —SeaWyrm (June 30, 2014)
>
> "The art style and gameplay is good, but the music is just not up to scratch in a game that relies on it so heavily" —Square of Downsides (June 22, 2014)
>
> "This game has GREAT potential, if you could select the music. But right now you're locked into someone else's favorite, which is heavy on the screechy autotune and electronica. Sorry, but that get [sic] so obnoxious after a while" —H Kane (April 19, 2014)[28]

This last review by user *H Kane* points to one solution: designing puzzles around various music genres and allowing players to choose according to their personal preference. Other music-based games such as *Just Dance*, *Guitar Hero*, or *Beat Saber* typically offer wide song varieties. Additionally, some games allow players to create and share custom levels using tracks from within or outside the game, benefitting both developers and communities by promoting engagement, extending the game's lifespan, increasing player enjoyment, and providing avenues for social affiliation (Bilińska-Reformat et al., 2020). Of course, providing music variety and the possibility to create levels is not always feasible. Such endeavours incur development costs and can contradict

[28] https://steamcommunity.com/app/282760/reviews/?browsefilter=toprated (Accessed 21 March 2026).

developer-artist vision if puzzles are designed specifically around particular genres or compositions. In short, like other design recommendations in this section, this advice is not meant as a requirement but rather as a consideration.

Conclusion

This chapter explored how subjective experiences of musicians and non-musicians differ when playing ABPs and how this creates tough expectations for game developers seeking to appeal to a broad audience. Through a case study of gameplay breakdowns and strategies in *Circuits* using the seven-step method presented in Visani Scozzi et al. (2017), supplemented by online reviews and established psychological theories (Flow, Self-Determination Theory), this chapter offered several design recommendations intended to ease this design tension and improve ABP gameplay experiences for most players. It is important to remember, however, that these recommendations are limited by the scarcity of research in the area and the design of this study; there would be tremendous insights from additional research with more varied participant pools (e.g., age, cultural background, gaming profiles) and experimental conditions (e.g., natural gameplay rather than in-lab, comparisons between different ABP mechanics).

The experimental results presented in recent chapters also illustrate how ABPs could be used as experimental paradigms to gather insights into how individuals with different training or demographic backgrounds conceptualize and manipulate sound, from cognitive representations and semantic descriptions of music to problem-solving strategies. Furthermore, while this chapter focused specifically on ABPs, the findings have broader implications for game design across domains. The core challenge identified here, that is designing for audiences with vastly different skill levels, applies to any game that draws on specialized knowledge or abilities. The recommendation to provide early difficulty customization, reduce complexity when introducing new mechanics, and use stimuli that level the playing field between novices and experts, could inform design across genres. More fundamentally, this analysis demonstrates the value of examining not just whether players succeed, but also the reasons behind why they don't, and how different populations experience the same game differently.

Ultimately, this chapter reveals a paradox at the heart of ABP design: the very factor that makes these games appealing to some players creates barriers for others. The recommendations offered here represent steps towards creating games that do not need to reassure players that musical knowledge is not required, but rather invite all players to develop their listening skills through play. At the very core of ABPs are tragically untapped opportunities for teaching principles of music, engaging with music in a novel fashion, and promoting creativity in players, topics which the next chapter delves into more deeply.

7. From Music to Problem-Space: Puzzlification and its Dimensions

This chapter examines the relationship between Audio-Based Puzzles (ABPs), in particular music-based puzzles, and the broader landscape of sound and music within gameplay experiences. ABPs have consistently fallen outside the boundaries of scholarship on music video games as well as on the application of game design elements to musical activities. As such, the musical gameplay experiences that emerge from ABPs remain undertheorized. To address this gap, this chapter introduces the concept of *puzzlification*, the process by which music is transformed from a fixed work to be listened to or performed into a puzzle to be solved. The chapter specifically focuses on what puzzlification does, describing two dimensions that emerge from the dynamic it creates between player, game, and music: the induction of new listening experiences, and the creation of new opportunities for musical participation. The chapter concludes by speculating on the implications of these dimensions for music education, proposing ABPs as a promising medium for developing musical skills and fostering musical creativity.

Music as Gameplay

Music Video Games

In *Man, Play and Games* (1958/2001), Roger Caillois proposed a typology of play organized along a continuum between two poles. At one end, *paidia* describes play that is spontaneous, improvisatory, and freeform, consisting of unstructured fun without defined goals or rules. At the other, *ludus* describes play that is rule-governed, structured and goal-directed. Rather than representing distinct categories, Caillois conceived of these as ideal types, with most play activities falling somewhere between them.

Discussions and classifications of music video games traditionally position these games as occupying spaces near one of these poles. For instance, Pilchmair and Kayali (2007) analyzed several interactive features in music video games and came to the conclusion that the genre can be split in two: rhythm games (*ludus*-based play in which the goal is to reproduce rhythms accurately) and instrument games (*paidia*-based play in which there is no goal). Similarly, Kassabian and Jarman's (2016) exploration of interactive music experiences showcases how music video games typically present the player with a goal that needs to be achieved (such as performing a song successfully) or offer unstructured play which allows for experimentation, and note that music in virtual worlds seem to only be able to appear in one of these modes, but not both.

A common thread between both articles is that music-based puzzle games are not considered as music video games. Pilchmair and Kayali explicitly exclude these from their definition of music video games, stating that "games that only include musical puzzles, or use music for certain interactive tasks cannot qualify as music video games" (p. 429). Kassabian and Jarman opt to define music video games as "those where the majority of gameplay or activity, and if applicable, winning or losing, are predicated on the ability to make good sound and/or musical choices" (p. 124), and explicitly exclude any indirect epistemic puzzle mechanics, considering them "games *about* music, but not *music games*" (p. 125), thus making *engaging* with music a core tenet of music video games.

As was showcased in Chapter 2, ABPs come in a wide variety of forms and interactive possibilities, many of which involve engaging with music. Even so, music-based puzzles appear to fall outside the boundaries established by previous scholars of music video games, which are typically represented at either end of the *paidia-ludus* continuum, while still retaining enough characteristics of music-based gameplay to arguably be considered part of the music video games family. To exemplify this distinct positioning on the continuum, one can look at *Sonority* (Hanging Gardens Interactive, 2022). In *Sonority*, players must play series of pitches in order to move platforms in the virtual environment. The direction (ascending/descending) and the distance between notes of each interval in the sequence dictate how the corresponding platforms move (Figure 7.1). At first, the player only has access to a limited number of notes and instruments, forcing them into the same solution as everyone else. As they acquire pitches and instruments throughout the game, they are given more freedom over the melodic sequences they use to solve puzzles, making it entirely possible for two players to come up with different solutions. Thus, while *Sonority* offers a clear purpose to be achieved by the player, and sets up rules in the form of the puzzle mechanics used in the game, the agency that the player is given in coming up with a working solution also means that the player is free to explore their own musical ideas and preferences within the confines of puzzle-solving. In short, *Sonority* and other similar ABP games occupy a unique position distinct from rhythm and instrument games along the *paidia-ludus* continuum, which requires further investigation.

The Gamification of Music

Another area in which ABPs fall outside the current scope of research into music as gameplay is gamification. Gamification is commonly defined as "the use of game design elements in non-game contexts" (Deterding et al., 2011). It is believed that by adding game elements (e.g., leaderboards, levels, resources, challenges, or goals), gamification can render non-game products and activities more engaging. In music, gamification has often been discussed in the context of its possible applications for learning an instrument or music theory (see, for instance, Birch & Woodruff, 2017; Candel & Colmenero, 2022; Figueiredo & Bidarra, 2014; Frazier-Roberts & Peoples, 2019; Samat et al., 2022).

Figure 7.1

An early puzzle from Sonority*, where the player has access to three pitches (C, D, E). Using one three-note sequence, the player must lower a block by one unit, then raise a second block by two units. This can be achieved by inputting the sequence D-C-E, which consists of a descending one-step interval, followed by an ascending two-step interval.*

Gibbons (2018) explores the gamification of classical music in two mobile applications. The first one, *The Young Person's Guide to the Orchestra*, revolves around Benjamin Britten's eponymous work and consists of four activities that seek to introduce children to the symphony orchestra. Gibbons notes that while the app aims at "instilling an abiding appreciation for classical music" (p. 149), it also "runs the risk of perpetuating elitist narratives of artistic superiority" (p. 149) with regards to classical music, situating itself in a movement of music appreciation that has historically put on a pedestal the works of white men while ignoring artworks by women and minorities. The second app, *Steve Reich's Clapping Music*, challenges users to perform the piece by tapping on the screen to generate a clapping sound. Given the challenging nature of the work, getting better at the game results in the development of musical skills such as rhythmic accuracy, but also leads to the player getting better at performing the piece themselves. As Gibbons puts it: "the line between developing musical skills and actually performing a musical work becomes blurred into unrecognizability. At some point, for both players and listeners, playing a game becomes indistinguishable from a musical performance" (p. 155).

While Gibbons' examples illustrate what the gamification of classical music can look like, it does not fully capture the use of classical music in some music-based puzzle games such as *Musaic Box* (KranX Productions, 2008). *Musaic Box* is a point-and-click game that asks the player to reassemble excerpts of famous classical pieces of music that have been fragmented into Tetris-like blocks (Figure 7.2). Each block contains a musical fragment, which can be listened to by interacting with it. The player is also able to listen to the reference piece at any time to help them with their task.

In both *The Young Person's Guide to the Orchestra* and *Steve Reich's Clapping Music*, the game is a vehicle for the music; it exists as a cultural object or performance practice to which game elements have been added in order to increase engagement, foster music appreciation, and teach musical skills. In *Musaic Box*, however, this relationship is inverted: the music has been restructured into the puzzle itself, thus making the music a vehicle for the game. Unlike gamified music, where engaging with the music's structure is the explicit goal, in *Musaic Box*, that structural engagement is a consequence of puzzle-solving rather than its purpose. The means by which the player engages with the music is by its very nature different, as the music has been reframed into a puzzle with a problem and a solution. Thus, gamification, or the addition of game design elements to a non-game activity, does not adequately describe what occurs in music-based puzzles. In the latter, music is not embellished with game elements but fundamentally transformed into a problem-space: a structured or semi-structured environment defined by a problem to be solved and bounded by constraints necessary to solve it.

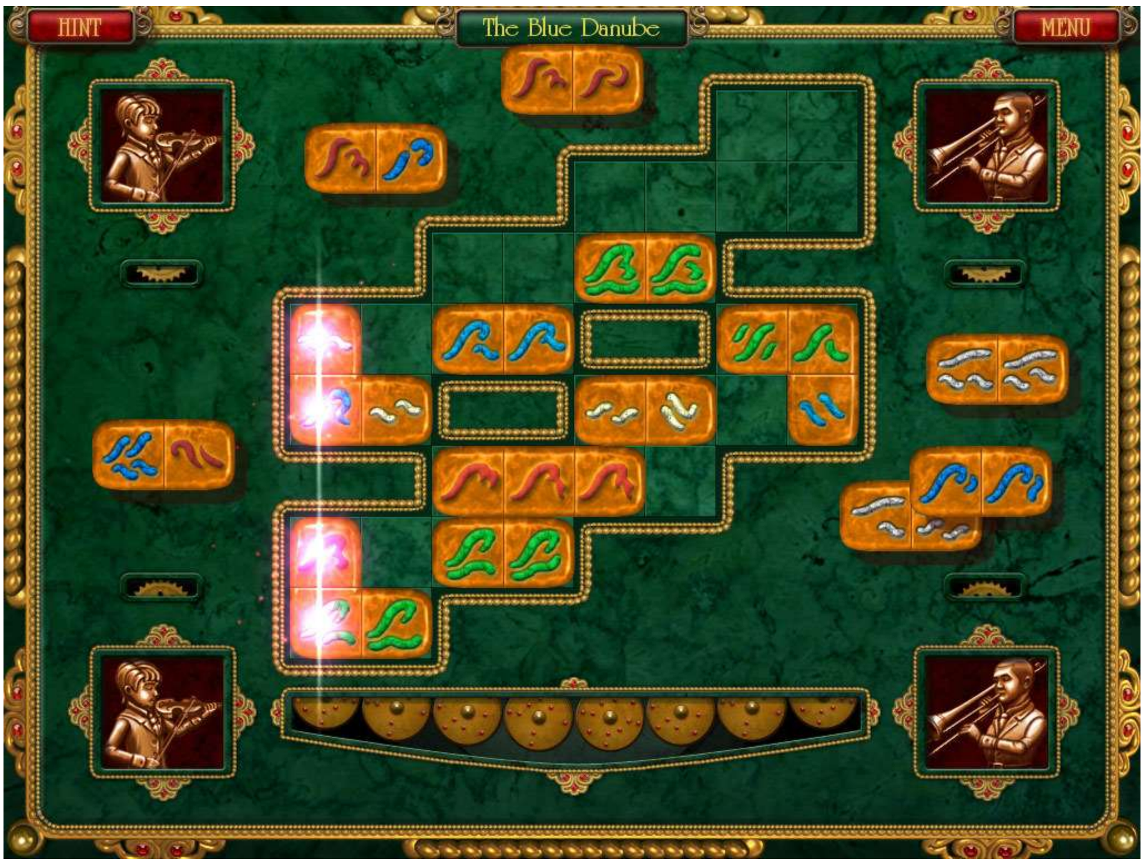
This chapter therefore suggests the concept of "puzzlification" to describe this phenomenon in which music is restructured as a puzzle to be solved. Just as puzzles themselves are difficult to rigidly define, the many forms ABP mechanics can assume make a formal definition of puzzlification elusive. Consequently, the next section takes a more productive approach by focusing on what puzzlification does rather than what it is.

The Puzzlification of Music: Two Dimensions

The previous section described how ABPs, and especially music-based puzzles, create musical gameplay experiences that have so far been underexamined in academic literature. This section aims to dive deeper into these experiences and the dynamic that they create between player, game, and music, by describing two phenomena which emerge from their particular configuration. These two phenomena are described as dimensions to account for the possible opportunities that they allow, but do not necessarily dictate. The next paragraphs focus on how puzzlification can influence *the* *listening experience* of the player, as well as create space to *engage in music-making activities*.

Figure 7.2

The "Blue Danube" puzzle in Musaic Box. *The shape of the line on each block depicts the melodic and rhythmic contour of that audio fragment, while the colour represents the instrument. Each block must be placed in the correct position as to recreate the excerpt when played from left to right.*



Puzzle-Solving and Listening

The first dimension of puzzlification is that it can induce new listening experiences in the player. This can be illustrated through the main gameplay loop of *Musaic Box* and *Circuits* (Digital Tentacle, 2014), in which a reference track is decomposed into fragments that the player must reassemble.[29] When listening to the individual fragments, the player not only listens to that sound or series of sounds, but must also think about how it relates to other fragments as well as where it fits within the reference track. When players try to decipher the fragment's relative position, they will attend to some of the fragment's sonic properties in order to compare them to that of the other fragments', such as determining that their pitch is higher or lower than a comparison tone. As such, when music has been recontextualized as a puzzled to be solved, the player cannot simply experience a passive reception of the music being presented to them. By virtue of trying to find a solution, this puzzle-solving space induces an active and analytical listening style.

We can better understand this specific dimension of puzzlification by situating it within previous literature on *modes of listening*. In *Introduction to the Sociology of Music* (1962/1976), Theodor Adorno proposed a typology of listeners ranging from the fully competent "expert listener" at one end to the most passive and inattentive listener at the other. Central to this typology is the concept of *structural hearing*, or the capacity to perceive music not as a succession of affective moments or memorable themes but as a coherent whole in which every element is understood in relation to every other. For Adorno, structural hearing is the ideal mode of musical engagement: the listener who hears structurally grasps how a motif develops, how a harmonic progression functions within the logic of the piece, and how the parts contribute to the whole.

Adorno considered structural hearing rare, being the product of sustained musical education and immersion rather than casual exposure. While puzzlification certainly

[29] For a more detailed description of *Circuits*' gameplay, see Chapter 4.

shifts the player away from a passive mode of listening to a more active one, not every player will attend to or even grasp every piece of structural information presented to them in the puzzle stimuli. Perhaps, then, it would be better to analyze the mode of listening induced by puzzlification through the lens of Sheila Williams' *synthetic* and *analytic* perception:

> Synthetic perception takes place when the information presented is interpreted as generally as possible; for example, hearing a room full of voices or listening to the overall effect of a piece of music. Analytic perception takes place when the information is used to identify the components of the scene to finer levels; for instance, listening to a particular utterance in the crowded room or tracking one instrument in an orchestral piece or identifying the components of a particular musical chord. (Williams, 1994, p. 98)

Puzzlification encourages analytic perception by making players focus on specific aspects of sound in the puzzle's acoustic scene to solve the puzzle. As was shown in the previous chapter, both musicians and non-musicians reported relying on distinct musical or sonic information to assist them in reordering the musical fragments in *Circuits*; players distinguished between fragments through salient sonic properties such as pitch, musical properties such as harmony, or other subjective and qualitative attributes of the sounds such as perceived intensity. Analytic perception is also encouraged when players must block out superfluous acoustic information, such as in some of the ABPs from *The Witness* (Thekla, Inc., 2016), in which the pitches that the player is tasked with transforming into abstract notation are hidden among distracting noises.

Another way in which we can view players' focused attention on specific sonic properties is through the lens of Pierre Schaeffer's *reduced listening* ("écoute réduite"). In his *Traité des objets musicaux* (1966), Schaeffer describes a mode of attention in which the listener suspends both the source of a sound and its referential or semantic meaning, attending instead to the intrinsic properties of the sound itself (e.g., its texture, timbre, or duration) rather than what it represents or where it comes from. Schaeffer considered reduced listening a deliberate and effortful practice, one that runs counter to our natural tendency to listen through sound toward its cause or meaning. By suspending these habitual modes of auditory perception, reduced listening transforms sound into what Schaeffer called a sound object ("objet sonore"), an autonomous entity available for purely sonic contemplation. By deconstructing a piece of music into fragments or otherwise recontextualizing sound as pieces of a puzzle, puzzlification can detach music from its referential meaning and transform it into a sound object, of which its most salient sonic properties are what the player engages with during the puzzle-solving process.

Another important factor which shapes the player's overall listening experience is the *context* created by the puzzles' problem-space. In *Ways of Listening* (2005), Eric Clarke uses ecological theory to describe how the meaning of a sound or a piece of music varies depending on the situation in which it is heard. Similarly, Stockfelt (1997) suggests that the ways in which a listener decides to engage and listen to a piece of music is shaped by the context in which the music is encountered. When players encounter music in the problem-space created by puzzlification, the ludic, puzzle-solving context directs the player's perception towards structural and relational properties while suppressing affective or referential ones; not because these latter properties necessarily disappear, but because actively trying to solve a music-based puzzle makes it functionally difficult to maintain modes of listening which aren't analytical, thus heavily constraining the player's choice of listening mode.

That is not to say, however, that puzzles lock the player into a single mode of listening throughout the entire experience. While puzzle-solving certainly shifts players into paying closer attention to the properties of what they are listening to, different moments during gameplay can call for different modes of listening. For instance, players of *Musaic Box* or *Circuits* might adopt a more holistic and affective mode of listening when listening to the reference track for the first time, after having successfully solved the puzzle, or even in the middle of puzzle-solving given that they are free to listen to the track at any moment. When players decide to adopt a more relaxed or synthetic listening mode, their experience might be enhanced due to having had to pay attention to the structural components of the piece. Puzzlification therefore allows for an oscillation between modes of listening which may produce a richer listening experience than pure structural analysis or passive reception alone.

Lastly, puzzlification of pre-existing music also constitutes a form of defamiliarization as theorized by the Russian Formalism movement. Defamiliarization ("*ostranenie*"), coined by Viktor Shklovsky in *Art as Technique* (1917/1965), describes the artistic strategy of presenting familiar objects or experiences in unfamiliar ways to disrupt automatic perception and restore the sensation of encountering something for the first time (Margolin, 2005). Puzzlification of familiar pieces of music, such as well-known pieces from the classical repertoire in *Musaic Box,* incites players to re-engage with these works through analytical listening. It encourages them to deconstruct the music into its structural components or pay attention to specific sonic properties, which completely disrupts and subverts the typical, "automatic" listening experience and induces a more conscious and attentive engagement with what they might otherwise pass over without noticing.

In sum, ABPs generate a problem-space that reshapes listening on two levels: it cultivates an analytical mode of listening the player can sustain or set aside throughout play, and it reframes the music itself, altering how a familiar or unfamiliar piece is presented and, by extension, experienced.

Puzzle-solving and Music-making

Whereas the first dimension of puzzlification concerns how the player listens to music, the second concerns how they participate in its creation. In *Musicking: The Meanings of Performing and Listening* (1998), Christopher Small proposed a fundamental reconceptualization of music as an activity rather than a thing, an object, or a fixed work. Small argued that music is something people do rather than something that exists independently of human participation, and that it encompasses far more than composing or performing, including activities such as listening, dancing, or rehearsing. Under this perspective, musical meaning is not inherent in a work but arises from the relationships between all those involved. Small's framework has since been widely taken up across music education and the social sciences (see Crossley, 2022).

Drawing on Small's definition, we can frame audio-based puzzle-solving as a musical activity in which musical meaning emerges from the meaningful interactions between player and puzzle (and/or game). Indeed, through having the player manipulate sounds, puzzlification makes the player an active participant in shaping the game's soundtrack, and this regardless of whether or not they are conscious of it. Two examples of this dimension of puzzlification can be found in *Sonority* and *The Way In* (Audio

Games Lab, 2021). As was previously mentioned, in later areas of *Sonority*, the player has agency over the pitches and instruments that will constitute the solution to any given puzzle, allowing individual players to create different pieces of music that each achieve the same goal. Importantly, once a player solves a puzzle, the solution will resonate throughout the area and loop for the rest of the game, becoming a permanent fixture of the game's soundscape. In doing so, the player restores music to a land that had lapsed into silence, contributing to the creation of the soundtrack in a way that is woven into the game's narrative.

The Way In is a maze game designed to have players experience a musical composition through navigation. The correct path through the maze contains the composition, which unfolds as the player progresses. Straying from this path causes the music to subside, leaving the player with only the ambient resonance of the surrounding space. The underlying composition consists of dense, ambient recordings processed in real time and dependent on the player's location within the maze, producing a constantly shifting soundscape of overlapping sonic textures without clear melodic or rhythmic anchors. Because the soundtrack is generated in response to the player's agency, no two playthroughs will produce the same musical experience. As such, the player is actively authoring a unique sonic event through their movement and choices rather than navigating a space that has a fixed soundtrack (Moore & Smallwood, 2022).

Just as puzzlification transforms the player from a passive listener to an active one, in *Sonority* and *The Way In*, it also transforms them from a passive receiver of the soundtrack to an active contributor. This collapse of the distinction between gameplay and sonic authorship can be related to Roland Barthes's concept of the writerly text

(Barthes, 1970). Barthes makes the distinction between two types of texts: the readerly ("*lisible*") and the writerly ("*scriptible*"). According to him, the readerly text positions the reader as a passive consumer, as it presents a fixed, coherent meaning to be received and decoded. The writerly text, by contrast, demands active participation from the reader in the production of meaning. It is open and resistant to becoming a fixed object, requiring the reader to become a producer of the text rather than merely its consumer. The writerly text fundamentally reconceives the relationship between text and reader, collapsing the distinction between production and reception.

One can see how a similar collapse occurs in *Sonority* and *The Way In* regarding their soundtracks. The player can't experience the music that resonates throughout the game as something that was composed for and delivered to them as a fixed object of which they are a passive receiver. They are actively implicated in its realization through their actions, as the very act of puzzle-solving *is* to create the soundtrack. Whether players consciously register their aesthetic agency is an open question, but the structural possibility of that agency is built into the puzzle design. While opportunities for players to influence a game's soundtrack through their actions are not unique to ABPs,[30] what distinguishes ABPs is their singular blend of structured and unstructured play and the impossibility of separating soundtrack from gameplay. The problem-space created by the puzzlification of music presents an ultimate purpose to be achieved, but affords the player agency in how they decide to reach it, and the path that the player ultimately decides to

[30] Collins (2008), for example, defines any instance of game sound responding directly to the player's inputs as "interactive audio", such as when the player jumps in *Super Mario Bros* or makes their character sing in *Asheron's Call 2: The Fallen Kings*.

take gives them authorship over the soundtrack positions them as an essential and active contributor to the game's sonic experience.

Another way in which puzzlification allows for musicking is by providing the player with an instrument that they can use to perform or compose music within and outside of the puzzle-solving space. It is not uncommon for games to provide opportunities for players to create or play music within a virtual world. Collins (2013) and Guzmán Anaya (2025), for instance, demonstrate how virtual instruments within games function not merely as ludic mechanics but as sites of musical participation, enabling players to engage with music culture both within the game environment and in the communities that emerge around it. In the context of puzzle-solving, such virtual instruments can be found in *Sonority, The Legend of Zelda: Ocarina of Time* and *Majora's Mask* (Nintendo, 1998, 2000), where the player is provided with musical instruments that are required to solve puzzles. These instruments therefore serve both a goal-oriented purpose as well as an unstructured music-making activity throughout the game. In both *Legend of Zelda* games, this latter use is further encouraged by allowing players to play pitches beyond the five notes that are required by the game, giving the player more freedom over what they can play.

Beyond virtual instruments, Collins (2013) also showcases examples of creators who have transformed games themselves (and even the consoles on which these games are played) into musical instruments, often in unconventional and sometimes unintended ways. One such example is Julian Oliver's *Quilted Thought Organ* (2001–2003), in which a modified game environment is repurposed as a three-dimensional music synthesizer. The player produces sound by moving through a virtual space and triggering audio samples through collision with objects, effectively turning navigation into musical performance. Puzzlification similarly can allow for the transformation of the game (or more precisely, the puzzle infrastructure) into a musical instrument, which the player learns and masters how to "play" throughout their playthrough. Puzzle games are typically designed with an incremental increase in complexity, allowing the player time and opportunities to practise and become familiar with the game's systems. By the end of the game, through having solved the puzzles and learning how elements interact, the player will have developed tacit knowledge of the system, the musical information embedded within it, and the interactions between the two. This knowledge consists of an embodied, practical understanding of how the puzzle infrastructure's components behave and is acquired through doing rather than instruction and as such, the player may or may not be able to articulate it explicitly.

In *Sonority*, *Musaic Box*, and *Circuits*, completion of the game is rewarded with a creative mode, allowing the player to make use of that acquired tacit knowledge and utilize the puzzle infrastructure to create and play music as they wish. The games effectively remove any remaining restrictions of goal-oriented gameplay and fully embrace unstructured musical play, giving the player space to explore their musicality through the sonic possibilities offered by the puzzle infrastructure. Even when a game doesn't make a creative or sandbox mode available, players may still manipulate these structures to create music in perhaps unintended ways, paralleling the musical practices of creators who modify software and hardware such as those presented by Collins (2013). While traditional instrument learning requires years of deliberate practice, often within formal institutional structures, puzzlification offers an alternative and accessible pathway into musical production that bypasses these requirements.

In sum, puzzlification's two main dimensions open the door to a world of novel musical and sonic experiences. Through its recontextualizing and possible defamiliarization of music within a problem space, puzzlification reshapes the listening experience of the player and offers new perspectives into the musical works they encounter. By making gameplay indissociable from the soundtrack, puzzlification also positions the player as an essential contributor to the game's soundscapes, creating accessible opportunities to explore musical creativity through composition and performance using puzzle infrastructure.

Puzzlification and Music Education

Kassabian and Jarman (2016) argue that music games carry an inherent didactic function; not necessarily as intentional educational projects, but in the sense that gameplay necessarily involves the acquisition or improvement of some skill, offering "a path to the eventual mastery of some aspect of music-making, whether or not that aspect is traditionally recognized as music-making" (p. 123). Lending support to this argument, Pasinski et al. (2016) found that frequent music video game players outperformed non-musicians on a battery of music perception tests, suggesting that music games may enhance musical aptitude in individuals without formal musical training. The listening dimension discussed in the preceding section suggests that ABPs could produce similar benefits. Analytical listening and musical participation both emerge as consequences of puzzle-solving rather than as explicit learning objectives, but what happens when these incidental outcomes are made intentional? This section aims to explore the ways in which ABPs could be designed and employed in the context of music education.

The use of games in music education is not a new proposition. With the growing popularity of music games over the past two and a half decades, several researchers have investigated their classroom potential. Gower & McDowall (2012) examined the perspectives of children and music teachers regarding rhythm and pitch games such as *Guitar Hero* and *SingStar*, finding that while teachers expressed reservations about the transferability of skills, students reported broader musical tastes and improved musical abilities. Relating their findings to a constructivist view of education, which argues that teachers ought to integrate students' interests and deliver knowledge in manners that resonate with the learner's understanding, the authors concluded that interactive music games are of high interest to students and should be taken into consideration when developing curricula. Since then, multiple studies have reported positive results from using music games to teach principles of music and develop perceptual abilities (Auerbach, 2010; Cassidy & Paisley, 2013; Jenson et al., 2016; Lesser, 2020; Paney, 2015; Peppler et al., 2011). However, these studies have focused primarily on performance-based games (i.e., those that sit closer to the *ludus* end of the *paidia*-*ludus* continuum), leaving the educational potential of ABPs largely unexplored.

This is a notable gap, especially given that puzzle games have shown promising uses as vehicles for structured, incremental learning (Hill et al., 2003; Ross, 2002). Several scholars have gestured toward the educational potential of ABPs specifically: Targett & Fernström (2003) suggested that "prolonged use of audio games could bring about the acquisition of various skills […], for example memory or pitch relationship perception” (p. 219), while Rovithis (2012) argued that musical puzzles could help develop recognition of intervals, chords, and scales, though he noted that this potential remains underexplored. There are, nonetheless, instances of ABP games that have been developed with explicit learning goals, such as *iClef* (Baratè & Ludovico, 2013), which uses a puzzle format to teach sight-reading skills, reporting high student engagement despite the task being typically considered tedious. Similarly, Lim et al. (2023) reported that their escape room puzzle game *Melody Mystery*, designed to teach basic music theory, produced encouraging results, with participants showing positive attitudes toward the subject. ABPs have also been developed to address accessibility needs. Allain et al.’s (2015) *Legend of Iris*, for instance, is an audio game featuring spatial puzzle mechanics designed to train spatial navigation skills in blind children, using an immersive soundscape to simulate real-world navigation in a safe and controlled environment.

Considered through the framework developed in this chapter, the educational possibilities of puzzlification extend across both dimensions. The first, puzzle-solving as analytical listening, maps most naturally onto existing music education goals. If puzzlification induces structural engagement with musical relationships as a condition of play, then ABPs are already functioning as informal ear training environments whether or not they are designed as such. These puzzle mechanics could be deliberately harnessed to teach musical concepts through ludic puzzle-solving rather than rote exercises, from identifying intervals and recognizing harmonic relationships to more advanced tasks such as building twelve-tone matrices or discerning complex polyrhythms.

The second dimension, puzzle-solving as music-making, opens possibilities for teaching students about musicality by allowing them to experiment with combinations of sounds and hear the results in real time. Games with multiple valid solutions, or that allow for individual authorship through puzzle-solving, are well suited to this aim as they allow students to make genuine aesthetic decisions and reflect on their consequences. Furthermore, as was briefly discussed in the previous section, if puzzlification fosters a tacit understanding of musical systems through play, and if creative modes allow that understanding to be expressed musically, then ABPs could offer an accessible pathway into music composition and performance that does not require formal theoretical training. This particular dimension could be beneficial in the case of non-musicians and younger learners who might otherwise find composition intimidating or inaccessible.

Overall, there remains a considerable gap in empirical knowledge regarding ABPs as tools for music education. The theoretical case, however, is compelling: puzzlification's dimensions align naturally with various music education goals, and the puzzle game's format of structured and incremental practice makes it an apt medium for their delivery. Comparing ABP-supported learning to conventional instruction and exploring the transferability of the musical skills that puzzlification develops remain open and interesting areas for future research.

Conclusion

This chapter has proposed puzzlification as a concept for understanding and exploring a phenomenon that has thus far been overlooked in scholarship on music as gameplay. By demonstrating that ABPs occupy a theoretically meaningful position along the *paidia-ludus* continuum, the chapter has shown the need for a more precise framework to describe the transformation of music into a problem-space. The two dimensions examined here suggest that this transformation has significant consequences for how players engage with music, both fundamentally reshaping the listening experience and opening pathways into musical creativity. The theorized applications of these dimensions into educational contexts further underscore the potential of puzzlification as a pedagogical tool. Taken together, the arguments presented in this chapter suggest that ABPs, and especially music-based puzzles, represent a distinct mode of musical engagement that merits further scholarly attention.

8. Putting the Pieces Together: A Conclusion

Prior to this thesis, Audio-Based Puzzles (ABPs) occupied an understudied space at the intersection of game studies, musicology, and psychology. The field was characterized by inconsistent terminology and a lack of shared analytical tools (a natural consequence of an area still finding its footing). Without a common framework for describing and analyzing ABPs, progress in this highly interdisciplinary field remained fragmented. To address this gap, this thesis undertook two research endeavours: an analysis of ABPs from 1990 to 2025, and a mixed-methods study of musicians and non-musicians playing puzzle games. Combined, the results lay a theoretical and empirical foundation for the study of ABPs, with repercussions that extend into how these puzzles are designed, and how we can use them to teach and better understand music. The main contributions and findings of each chapter are recapped in the paragraphs that follow; a summary, along with avenues for future research, is also provided in Table 8.1.

In the first chapter of this thesis, I laid out the pieces upon which the remainder of this thesis is built. I first established a definition of Audio-Based Puzzles, which proved more difficult task than anticipated. Indeed, the very concept of puzzle is itself nebulous, with its definitional borders remaining elusive given the many edge cases and competing aspects that must be considered. Nevertheless, by the end of the chapter, we arrived at a definition whose utility extends beyond the scope of this thesis. I also briefly discussed audio games, as most prior research on ABPs was concentrated in that field.

In Chapter 2, I introduced the Audio-Based Puzzle Mechanics (ABPM) framework, which consists of five categories (Input/Output, Spatial, Environmental, Contextual, and Epistemic) that each have a direct and indirect variant. Each category describes the player's goal or method of interaction within the puzzle, while the two variants describe the way audio is presented. Previous literature on ABPs lacked consistent terminology and definitions; the creation of this framework fills this particular gap by offering the necessary tools to systematically analyze and compare ABPs. In particular, the distinction between direct and indirect mechanics allows us to better understand the levels of integration of audio in puzzles and how that might affect their accessibility, complexity, and reception.

The ABPM framework, however, is not without its limitations, which are discussed in Chapter 3. Issues related to the use of speech or silence in puzzles, for instance, complicate the definitional boundaries of ABPs I set out in the introduction to the thesis. Nevertheless, the ABPM framework's utility as an analytical tool has been demonstrated in the survey of ABPs from 1990 to 2025 presented at the end of the chapter, which showcased the evolution and growth of the medium through the years. Early feedback from my presentations of the framework also showed promise regarding its use as a creative tool; game developers at the 2024 *Game Music and Sound Design Conference* reported new perspectives on the design of ABPs, and attendees at a lecture I gave in early 2025 at the University of Alberta's Sound Studies Institute were able to successfully use the framework to create novel puzzle concepts. While certainly informal, these initial responses to the ABPM framework speak to its potential and I look forward to any reported future uses.

Chapters 4, 5, and 6 all pertained to an experiment I conducted where musicians and non-musicians each played two puzzle games; the ABP game *Circuits* (Digital Tentacle, 2014), and the visual puzzle game *Square Logic* (New Dawn Game Studio, 2023). Chapter 4 detailed the experimental protocol of the study, which used mixed methods to analyze the participants' performance and subjective experiences when playing the two games. Quantitative measures were used to assess performance and to compare participant's enjoyment and subjective feeling of difficulty. The participants' verbal commentary during gameplay, as well as written post-gameplay feedback, were both collected to offer a more qualitative perspective into the participants' gameplay experiences and to gather insights into their thought process while solving puzzles. The mix of quantitative and qualitative data provided a rich well of information to draw from in the investigations that followed.

Chapter 5 focused exclusively on the collected quantitative data to contrast the two groups' puzzle-solving performance, enjoyment of the games, and perceived difficulty of the puzzles. Statistical analyses using linear mixed models revealed that musicians, on average, were faster and more efficient when solving *Circuits*' ABPs, while such between-group differences did not appear in the analysis of *Square Logic*. Analyses of variance also showed that musicians tended to rate *Circuits* as more enjoyable and less difficult than non-musicians. These results overall have interesting implications regarding the near and far transfer of musical skills. On one hand, the difference between both groups in *Circuits*, as well as on the *Profile of Music Perception Skills* test which was used evaluate perceptual ability, suggests that musicians' advantages in auditory perception persist when solving ABPs. On the other hand, the lack of difference between groups in *Square Logic* implies that previously recorded advantages of musical expertise on visual task might not translate into better performance on a visual puzzle-solving task. These results, however, should be interpreted cautiously, as the study had a small sample size and low statistical power, and did not control for gaming experience *a priori*. Further research using a variety of ABPs and visual puzzles would provide a more complete picture of the phenomenon.

The quantitative results of Chapter 5 also revealed stark differences between both groups in their enjoyment and perceived difficulty of *Circuits*. Chapter 6 aimed to obtain a better understanding of the root causes of this difference by shifting the analysis to the interactions between the participants and the game's design choices. Using a seven-step approach developed by Visani Scozzi et al. (2017) to examine these interactions, supplemented by public reviews of the game, the analysis highlighted one of the biggest design tension that is intrinsic to the creation of ABPs; developing an experience that is appropriately difficult to an audience with a wide variety of musicianship. Following this analysis, the chapter presented general design recommendations for ABPs specifically aimed at easing this tension and making these gameplay experiences accessible to a wider audience.

Finally, in Chapter 7, I introduced the concept of *puzzlification* to describe the musical gameplay experiences that emerge from ABPs. I argued that, by contextualizing music and/or sound within a problem-space as a puzzle to be solved, puzzlification fundamentally alters the musical experience of the player. First, the process of solving the puzzle strongly encourages the player to shift from a passive mode of listening to an active one, requiring them to pay close attention to structural rather than affective properties of sound. Puzzlification also allows for a structured oscillation between these modes, potentially enhancing the listening experience in the player. Second, I posited that ABP solving is a music-making activity, in which the player becomes a co-creator of the game's soundtrack, whether they are consciously aware of it or not. By coupling the process of puzzle-solving with the one of creating or playing sounds, puzzlification also allows for an accessible method of exploring one's musical creativity. I ended the chapter by speculating on the potential of puzzlification in the realm of music education as a structured by ludic activity to learn more about music.

Technology is ever evolving, and as such, so are ABPs. It is impossible to predict how ABPs will look in the near and far future; nevertheless, the tools presented in this thesis provide a way to design and analyze ABPs as they are currently and are flexible enough to accommodate future developments in the field. Given the rise in popularity of generative artificial intelligence, it's likely that such tools will play a role in shaping how ABPs will look. Some of its most obvious uses are related to the topic of adaptive difficulty systems, which were discussed in Chapter 6. Another way it could be incorporated is through the generation of endless puzzles; as puzzle games are usually finite and offer little to no replay value, automating the creation of new puzzles could extend the lifespan of such games.

I have been playing video games ever since I was a child. Over the past couple of years, I have had the opportunity and the pleasure to play a wide variety of ABPs, and if there is one thing I can take away from this thesis, it is that I (perhaps selfishly) want *more*. Of course, because I find them to be entertaining and enjoyable gaming experiences; but one also cannot deny the incredible potential that ABPs possess in the realms of accessibility and music education that, piece by piece, I have tried to highlight throughout this document. It is my sincere hope that the tools, results, and recommendations provided in this thesis will be of assistance to both game developers and scholars involved in the wonderful, entertaining, and puzzling realm of ABPs.

Table 8.1

Summary of each chapter's main contributions and directions for future research.

	Main contributions	Directions for future research
1	• Providing a definition of ABPs	
2	• Introducing the ABPM framework: a taxonomy of ABP mechanics and level of audio integration within them	• Using the framework to compare, analyze, or design ABPs • Validating the framework against new ABPs • Expanding to other media beyond video games (e.g. board games, tabletop role-playing games, escape rooms)
3	• Discussing the definitional boundaries surrounding ABPs • Analyzing ABP presentations/trends from 1990 to 2025	• Diving deeper into the position of silence- and speech-based puzzles within the ABPM framework • Analyzing new ABP trends beyond 2025
4	• Providing a protocol/template for comparative, mixed-methods studies of ABPs	• Testing how the protocol can be improved or altered for future investigations, such as using a retrospective think-aloud protocol or semi-structured interviews

5	• Demonstrating that musicians perform better than non-musicians on an ABP task but not a visual puzzle task
	• Demonstrating that musicians find an ABP game more enjoyable and less challenging than non-musicians, but that these differences do not emerge in a visual puzzle task
		• Replicating the study with a wider variety of games to see if the results change
		    o Comparing ABPs that use mechanics from different categories (e.g. Input/Output vs Spatial, Direct vs Indirect)
		    o Comparing with various types of non-ABPs (e.g. other visual puzzles, logic puzzles)

6	• Illustrating the application of a seven-step mixed-methods approach (Visani Scozzi et al., 2017) to study ABP design.
	• Offering general design recommendations for ABPs (specifically targeted at making them appealing to a wide audience)
	• Highlighting the potential of ABPs as a method to study cognitive representations of music and audio-based problem-solving
		• Analyzing the design of other ABPs to see if similar trends emerge
		• Conducting studies on the general design recommendations that were offered to provide empirical evidence for (or against) them
		    o Analyzing the impact of adaptive difficulty systems on players' experiences
		    o Comparing the use of different puzzle stimuli on players' experiences
		• Conducting studies on ABP solving strategies and factors that influence these strategies

7	• Introducing p*uzzlification*: a framework to study the undertheorized relationship between player, music, and game in the context of puzzles
	    o Arguing that puzzlification induces novel listening experiences
	    o Arguing that puzzlification positions the player as a co-author of the game's soundtrack
		• Conducting studies on the listening and music-making experiences of ABP players
		• Exploring and testing the uses of ABPs in music education

Appendix A

List of games surveyed to create the Audio-Based Puzzle Mechanics Framework.

Title	Year of Release	Developer	Publisher
999: Nine Hours, Nine Persons, Nine Doors	2009	Chunsoft	Spike Chunsoft
ALIDA	2004	Dejavu Worlds	Got Game Entertainment
Animal Well	2024	Shared Memory	Bigmode
Aquaria	2007	Bit Blot, LLC	
Audio Puzzler	2008	Diakopoulos, N., Luther, K., & Essa, I. (2008). Audio Puzzler: Piecing together time-stamped speech transcripts with a puzzle game. *Proceedings of the 16th ACM International Conference on Multimedia, MM '08*, 865–868. https://doi.org/10.1145/1459359.1459507	
Audio-Based Puzzle Game	2012	Carvalho, J., Guerreiro, T., Duarte, L., & Carriço, L. (2012). *Audio-Based Puzzle Gaming for Blind People*.	
Banjo-Kazooie	1998	Rare	Nintendo
Bendy and the Ink Machine	2017	Joey Drew Studios	
Binaural Navigation Game	2015	Bălan, O., Moldoveanu, F., Moldoveanu, A., & Butean, A. (2015, September 24). *Developing a navigational 3D audio game with hierarchical levels of difficulty for the visually impaired players*. Romanian Conference on Human Computer Interaction.	
Call of Duty: Black Ops	2010	Treyarch	Activision
Chords Enchanter	2022	XD Games Studio	
Chrono Cross	1999	Square	
Chronos: Before the Ashes	2020	Gunfire Games and THQ Nordic	THQ Nordic
Circuits	2014	Digital Tentacle	
Classical Music Minesweeper	2022	Kiki Castella	

CrossSong	2015	Smith, J. B. L., Kato, J., Fukayama, S., Percival, G., & Goto, M. (2017). The CrossSong Puzzle: Developing a Logic Puzzle for Musical Thinking. *Journal of New Music Research*, *46*(3), 213–228. https://doi.org/10.1080/09298215.2017.1303519.	
Deltarune	2018	Toby Fox	
Don't Forget the Lyrics	2021	Music Games & Trivia Games & Trombone - LB ART	
Dream Chronicles	2007	KatGames	PlayFirst
Dream Chronicles 2	2008	KatGames	PlayFirst
Dream Chronicles 3	2009	KatGames	PlayFirst
Eternal Darkness: Sanity's Requiem	2002	Silicon Knights	Nintendo
Eternal Sonata	2007	tri-Crescendo	Bandai Namco Games
Fallout 3	2008	Bethesda Game Studios	Bethesda Softworks
Figment	2017	Bedtime Digital Games	
Final Fantasy VI	1994	Square	
Final Fantasy VII	1997	Square	
Final Fantasy VIII	1999	Square	
Final Fantasy X-2	2003	Square	
Finish the Lyrics	2018	Daniel Baczkowski	
Genshin Impact	2020	miHoYo	HoYoverse
Goat Simulator 3	2022	Coffee Stain North	Coffee Stain Publishing
Google Doodle	2015	Google	
Hear Your Path	2022	Solo Philosopher	
Hollow Knight: Silksong	2025	Team Cherry	

Honkai: Star Rail	2023	miHoYo	HoYoverse
iClef	2013	Baratè, A., & Ludovico, L. A. (2013). Serious Games for Music Education—A Mobile Application to Learn Clef Placement on the Stave: *Proceedings of the 5th International Conference on Computer Supported Education*, 234–237. https://doi.org/10.5220/0004364902340237	
In Verbis Virtus	2015	Indomitus Games	
JAK II	2003	Naughty Dog	Sony Computer Entertainment
King's Quest VII: The Princeless Bride	1994	Sierra On-Line	
Kronos		Rovithis, E. (2012). A classification of audio-based games in terms of sonic gameplay and the introduction of the audio-role-playing-game: Kronos. *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound, AM '12*, 160–164. https://doi.org/10.1145/2371456.2371483	
Locus Sono	2015	Audio Games Lab	
Loom	1990	Lucasfilm Games	
Lorelei and the Laser Eyes	2024	Simogo	Annapurna Interactive
Loud and Clear	2019	Baas, B., van Peer, D., Gerling, J., Tavasszy, M., Buskulic, N., Salamon, N. Z., Balint, J. T., & Bidarra, R. (2019). Loud and clear: The VR game without visuals. *Games and Learning Alliance- 8th International Conference, GALA 2019, Proceedings, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11899*, 180–190. https://doi.org/10.1007/978-3-030-34350-7_18	
Lufia II: Rise of the Sinistrals	1995	Neverland	Taito, Natsume and Nintendo
Luigi's Mansion	2001	Nintendo EAD	Nintendo
Luigi's Mansion 3	2019	Next Level Games	Nintendo
Marvel's Guardians of the Galaxy	2021	Eidos-Montréal	Squareff Enix
MazeSounds	2016	AMIDEA Games	

Murder House	2020	Puppet Combo	
Musaic Box	2008	KranX Productions	
Music Puzzle	2013	Hansen, K. F., Hiraga, R., Li, Z., & Wang, H. (2013). Music Puzzle: An Audio-Based Computer Game That Inspires to Train Listening Abilities. In D. Reidsma, H. Katayose, & A. Nijholt (Eds), *Advances in Computer Entertainment* (pp. 540–543). Springer International Publishing. https://doi.org/10.1007/978-3-319-03161-3_48	
Music Puzzle	2012	Shard Workshop	
Nancy Drew: Captive Curse	2011	HeR Interactive	
Nancy Drew: Curse of Blackmoor Manor	2004	HeR Interactive	DreamCatcher
Nancy Drew: Last Train to Blue Moon Canyon	2005	HeR Interactive	DreamCatcher
Nancy Drew: Message in a Haunted Mansion	2000	HeR Interactive	DreamCatcher
Nancy Drew: Sea of Darkness	2015	HeR Interactive	
Nancy Drew: The Haunted Carousel	2003	HeR Interactive	DreamCatcher
Nancy Drew: The Silent Spy	2013	HeR Interactive	
Nancy Drew: Warnings at Waverly Academy	2009	HeR Interactive	
Nine Sols	2024	Red Candle Games	
One Hand Clapping	2021	Bad Dream Games	HandyGames
Orchestra	2007	7-128 Software	
Outcast: The Second Contact	2017	Appeal Studios	Nacon
Outer Wilds	2019	Mobius Digital	Annapurna Interactive

Paper Mario: The Thousand-Year Door	2004	Intelligent Systems	Nintendo
Partitas	2021	Lotz	
Phasmophobia	2020	Kinetic Games	
Phonopath	2011	Kevin Regamey	
Pontes Sound Puzzle	2013	Pontes Games	
Prince of Persia: The Sands of Time	2003	Ubisoft Montréal	Ubisoft
Remnant 2	2023	Gunfire Games	Gearbox Publishing
Resident Evil	2002	Capcom Production Studio 4	Capcom
Resident Evil: Village	2021	Capcom	
Resonance of the Ocean	2022	Image Labo and Heaviside Creations	Image Labo
Scale Enchanter	2021	XD Games Studio	
Seiklus	2003	cly5m	Independent
Self-Discovery	2024	Audio Games Lab	
Silent Hill	1999	Team Silent	Konami
SongPop Classic	2015	FreshPlanet	
Sonority	2022	Hanging Gardens interactive	Application Systems Heidelberg
Sound Hunter	2013	Brieger, S. (2013). *Sound Hunter: Developing an HRTF-Based Audio Game for People with Visual Impairments*. https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A836283&dswid=5744	
Sound Puzzle	2004	VIPgameszone	
Soundoku	2021	Aline, H., Dombrowsky, A., & Sporka, A. J. (2021). Soundoku: A sound puzzle game. *2021 IEEE Conference on Games (CoG)*, 01–04. https://doi.org/10.1109/CoG52621.2021.9618984	

Spuzzles	2024	Rovithis, E., Papadopoulou, A., Komianos, V., Garneli, V., & Floros, A. (2024). Speech Puzzles (Spuzzles): Engaging the Reduced, Causal, and Semantic Listening Modes for Puzzle Design in Audio Games. *Applied Sciences*, *14*(9), 3858.	
StarTropics	1990	Nintendo R&D3 and Locomotive Corporation	Nintendo
Super Mario RPG	1996	Square	Nintendo
Sync or Swim	2012	Audio Games Lab	
SynthSync	2018	Jordan-Douglass, A., Kumar, V., & Woods, P. J. (2018). Exploring computational thinking through collaborative problem solving and audio puzzles. *Proceedings of the 17th ACM Conference on Interaction Design and Children, IDC '18*, 513–516. https://doi.org/10.1145/3202185.3210766	
TapBeats	2011	Kim, J., & Ricaurte, J. (2011). TapBeats: Accessible and mobile casual gaming. *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 285–286. ASSETS '11: The 13th International ACM SIGACCESS Conference on Computers and Accessibility. https://doi.org/10.1145/2049536.2049609	
The 7th Guest	1993	Trilobyte	Virgin Interactive Entertainment
The 7th Guest VR	2023	Vertigo Games	
The Elder Scrolls: Morrowind	2002	Bethesda Game Studios	Bethesda Softworks
The Legend of Iris	2015	Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., Gaubitch, N. D., Hendriks, R. C., & Kybartas, B. (2015). An audio game for training navigation skills of blind children. *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, 1–4. https://doi.org/10.1109/SIVE.2015.7361292	
The Legend of Zelda: Breath of the Wild	2017	Nintendo	
The Legend of Zelda: Majora's Mask	2000	Nintendo	
The Legend of Zelda: Ocarina of Time	1998	Nintendo	
The Legend of Zelda: Skyward Sword	2011	Nintendo	

The Legend of Zelda: The Wind Waker	2002	Nintendo	
The Lost Crown	2008	Darkling Room	Iceberg Interactive
The Lost Garden	2021	Audio Games Lab	
The Neverhood	1996	The Neverhood, Inc.	DreamWorks Interactive
The Noize Games	2014	Rovithis, E., Floros, A., Mniestris, A., & Grigoriou, N. (2014). Audio games as educational tools: Design principles and examples. *2014 IEEE Games Media Entertainment*, 1–8. https://doi.org/10.1109/GEM.2014.7048083	
Transference	2018	SpectreVision and Ubisoft Montréal	Ubisoft
Tunic	2022	TUNIC Team	Finji
Undertale	2015	Toby Fox	
Unheard - Voices of Crime	2019	NEXT Studios	NEXT Studios and 505 Games
Wandersong	2018	Greg Lobanov	Humble Bundle
Warframe	2013	Digital Extremes	
We Were Here Forever	2022	Total Mayhem Games	
Young People's Guide to the Orchestra	2021	Carnegie Hall	
Zork Nemesis: The Forbidden Lands	1996	Zombie LLC	Activision

Appendix B

Demographic Questionnaire

1. Age: ______________

2. Gender:

☐ Female ☐ Male ☐ Nonbinary

☐ You don't have an option that applies to me. I identify as (please specify):

____

☐ Prefer not to say

3. On average, how many hours per week do you play video games?

☐I do not play video games ☐ 1-5 ☐ 6-10 ☐ 11-15 ☐ 16+

4. On average, how many of these hours are spent playing games from the puzzle genre? (For example: Baba Is You, Portal, The Talos Principle, The Witness, Escape Simulator)

☐None ☐ Very few ☐ Few ☐ Half ☐ Most ☐ All

Complete the following section if you are a musician:

1. List up to three instruments that you play, in order of proficiency (1 being the most proficient, 3 being the less proficient)
   1. _____________________
   2. _____________________
   3. _____________________

2. Years of musical training (regardless of instrument): _____________________

3. On average, how many hours per week do you practice playing an instrument (regardless of instrument)?

☐Less than 2 ☐ 2-4 ☐ 5-7 ☐ 8-10 ☐ 11+

Appendix C

Post-Game Questionnaire

Please indicate to what extent you agree/disagree with each of the following statements:

	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree Nor Disagree	Slightly Agree	Agree	Strongly Agree
Succeeding in the game required much planning	☐	☐	☐	☐	☐	☐	☐
I had to memorise a lot of different things when playing the game	☐	☐	☐	☐	☐	☐	☐
I had to think several steps ahead when playing the game	☐	☐	☐	☐	☐	☐	☐
I had to prepare for the things that the game threw at me	☐	☐	☐	☐	☐	☐	☐
Playing the game requires great effort	☐	☐	☐	☐	☐	☐	☐
I felt challenged when playing the game	☐	☐	☐	☐	☐	☐	☐
I had lots of different things to think about at once in the game	☐	☐	☐	☐	☐	☐	☐
The game made me manage several tasks at the same time	☐	☐	☐	☐	☐	☐	☐
I had to constantly keep track of what was going on in the game	☐	☐	☐	☐	☐	☐	☐
I had to think actively when playing the game	☐	☐	☐	☐	☐	☐	☐
Playing the game required me to do my best	☐	☐	☐	☐	☐	☐	☐
I found the game to be enjoyable	☐	☐	☐	☐	☐	☐	☐

If you would like to share some of your thoughts about the game, please do so below (for example, what you liked or disliked, what you found easy or difficult, etc.):

Bibliography

Adams, E. (2009). *Fundamentals of Game Design* (2nd edn). New Riders. http://gen.lib.rus.ec/book/index.php?md5=b3cd92a69cf4997b7beb271b5338611e

Adorno, T. W. (1976). *Introduction to the sociology of music*. New York : Seabury Press. (Original work published 1962)

Aline, H., Dombrowsky, A., & Sporka, A. J. (2021). Soundoku: A sound puzzle game. *2021 IEEE Conference on Games (CoG)*, 01–04. https://doi.org/10.1109/CoG52621.2021.9618984

Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., Gaubitch, N. D., Hendriks, R. C., & Kybartas, B. (2015). An audio game for training navigation skills of blind children. *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, 1–4. https://doi.org/10.1109/SIVE.2015.7361292

Amer, T., Kalender, B., Hasher, L., Trehub, S. E., & Wong, Y. (2013). Do Older Professional Musicians Have Cognitive Advantages? *PLOS ONE*, *8*(8), e71630. https://doi.org/10.1371/journal.pone.0071630

AMIDEA Games. (2016). *Maze Sounds* [Video game].

Anaya, E. M., Pisoni, D. B., & Kronenberger, W. G. (2017). Visual-spatial sequence learning and memory in trained musicians. *Psychology of Music*, *45*(1), 5–21. https://doi.org/10.1177/0305735616638942

Apple. (2003). *Music Quiz* [Video game].

Archambault, D., Ossmann, R., Gaudy, T., & Miesenberger, K. (2007). Computer games and visually impaired people. *Upgrade*.

Atari Inc. (1974). *Touch Me* [Arcade game].

Atkinson, M. T., Gucukoglu, S., Machin, C. H. C., & Lawrence, A. E. (2006). Making the mainstream accessible: Redefining the game. *Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames, Sandbox '06*, 21–28. https://doi.org/10.1145/1183316.1183321

Audio Games Lab. (2021a). *The Lost Garden* [Video game].

Audio Games Lab. (2021b). *The Way In* [Video game].

Auerbach, B. (2010). Pedagogical Applications of the Video Game Dance Dance Revolution to Aural Skills Instruction. *Music Theory Online*, *16*(1). https://mtosmt.org/issues/mto.10.16.1/mto.10.16.1.auerbach.html

Baas, B., van Peer, D., Gerling, J., Tavasszy, M., Buskulic, N., Salamon, N. Z., Balint, J. T., & Bidarra, R. (2019). Loud and clear: The VR game without visuals. *Games and Learning Alliance- 8th International Conference, GALA 2019, Proceedings, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11899*, 180–190. https://doi.org/10.1007/978-3-030-34350-7_18

Baczkowski, D. (2018). *Finish the Lyrics* [Video game].

Bad Dream Games. (2021). *One Hand Clapping* [Video game]. HandyGames.

Baer, R. H., & Morrison, H. J. (1978). *Simon* [Electronic game]. Hasbro.

Bălan, O., Moldoveanu, F., Moldoveanu, A., & Butean, A. (2015, September 24). *Developing a navigational 3D audio game with hierarchical levels of difficulty for the visually impaired players*. Romanian Conference on Human Computer Interaction.

Baldan, S., de Götzen, A., & Serafin, S. (2013). Mobile rhythmic interaction in a sonic tennis game. *CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13*, 2903–2906. https://doi.org/10.1145/2468356.2479570

Baratè, A., & Ludovico, L. A. (2013). Serious Games for Music Education—A Mobile Application to Learn Clef Placement on the Stave: *Proceedings of the 5th International Conference on Computer Supported Education*, 234–237. https://doi.org/10.5220/0004364902340237

Barthes, R. (1970). *S/Z*. Éditions du Seuil.

Bedtime Digital Games. (2017). *Figment* [Video game].

Bernard, H. R. (with Wutich, A., & Ryan, G. W.). (2017). *Analyzing qualitative data: Systematic approaches* (2nd ed.). SAGE.

Bethesda Game Studios. (2002). *The Elder Scrolls III: Morrowind* [Video game]. Bethesda Softworks.

Bianchi, F., Santurette, S., Wendt, D., & Dau, T. (2016). Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort. *JARO:*

Journal of the Association for Research in Otolaryngology, *17*(1), 69–79. https://doi.org/10.1007/s10162-015-0548-2

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, *50*(2), 339–347. https://doi.org/10.3758/s13421-021-01226-6

Bilińska-Reformat, K., Dewalska-Opitek, A., & Hofman-Kohlmeyer, M. (2020). To Mod or Not to Mod—An Empirical Study on Game Modding as Customer Value Co-Creation. *Sustainability*, *12*(21). https://doi.org/10.3390/su12219014

Birch, H., & Woodruff, E. (2017). Technical exercise practice: Can piano students be motivated through gamification? *Journal of Music, Technology and Education*, *10*, 31–50. https://doi.org/10.1386/jmte.10.1.31_1

Bit Blot, LLC. (2007). *Aquaria* [Video game].

Brieger, S. (2013). *Sound Hunter: Developing an HRTF-Based Audio Game for People with Visual Impairments*. https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A836283&dswid=5744

Brieger, S. (2015). *Developing a navigational 3D audio game with hierarchical levels of difficulty for the visually impaired players*. Romanian Conference on Human Computer Interaction.

Cafri, N. (2025). *Dynamic Difficulty Adjustment With Brain Waves as a Tool for Optimizing Engagement* (arXiv:2504.13965). arXiv. https://doi.org/10.48550/arXiv.2504.13965

Caillois, R. (2001). *Man, Play and Games*. University of Illinois Press. (Original work published 1958)

Cancino-Chacón, C., Peter, S., Chowdhury, S., Aljanaki, A., & Widmer, G. (2020). *On the Characterization of Expressive Performance in Classical Music: First Results of the Con Espressione Game* (arXiv:2008.02194). arXiv. https://doi.org/10.48550/arXiv.2008.02194

Candel, E., & Colmenero, M. (2022). Gamification and mobile learning: Innovative experiences to motivate and optimise music content within university contexts. *Music Education Research*, *24*, 1–16. https://doi.org/10.1080/14613808.2022.2042500

Capcom. (1996). *Resident Evil* [Video game].

Capcom. (2021). *Resident Evil Village* [Video game].

Cara, M. A. (2024). The Influence of Music Reading on Spatial Working Memory and Self-Assessment Accuracy. *Brain Sciences*, *14*(11), 1152. https://doi.org/10.3390/brainsci14111152

Caroux, L. (2023). Presence in video games: A systematic review and meta-analysis of the effects of game design choices. *Applied Ergonomics*, *107*, 103936. https://doi.org/10.1016/j.apergo.2022.103936

Carvalho, J., Duarte, L., & Carriço, L. (2012). Puzzle games: Player strategies across different interaction modalities. *Proceedings of the 4th International Conference on Fun and Games, FnG '12*, 64–72. https://doi.org/10.1145/2367616.2367624

Carvalho, J., Guerreiro, T., Duarte, L., & Carriço, L. (2012). *Audio-Based Puzzle Gaming for Blind People*.

Cassidy, G. G., & Paisley, A. M. J. M. (2013). Music-games: A case study of their impact. *Research Studies in Music Education*, *35*(1), 119–138. https://doi.org/10.1177/1321103X13488032

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds), *The Handbook of Qualitative Research* (2nd edn, pp. 507–535). Sage.

Chen, X. (2019). *The relationship between video games, problem-solving skills, and academic performance from IT students' perspective* [Master's Thesis]. University of Oulu.

Chew, E. (2014). The Spiral Array. In E. Chew (Ed.), *Mathematical and Computational Modeling of Tonality: Theory and Applications* (pp. 41–60). Springer US. https://doi.org/10.1007/978-1-4614-9475-1_3

Chion, M. (1994). *Audio-Vision: Sound on Screen* (C. Gorbman, Ed.). Columbia University Press.

Clarke, E. F. (2005). *Ways of Listening*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195151947.001.0001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Lawrence Erlbaum Associates.

Cohrdes, C., Grolig, L., & Schroeder, S. (2019). The development of music competencies in preschool children: Effects of a training program and the role of environmental factors. *Psychology of Music*, *47*(3), 358–375. https://doi.org/10.1177/0305735618756764

Collins, K. (2008). *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. The MIT Press. https://doi.org/10.7551/mitpress/7909.001.0001

Collins, K. (2013). Embodying Game Sound in Performance. In *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (MIT Press, pp. 89–120).

Constant, T., & Levieux, G. (2019). Dynamic Difficulty Adjustment Impact on Players' Confidence. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300693

Cooper, P. K. (2020). It's All in Your Head: A Meta-Analysis on the Effects of Music Training on Cognitive Measures in Schoolchildren. *International Journal of Music Education*, *38*(3), 321–336. https://doi.org/10.1177/0255761419881495

Corsi, P. M. (1973). *Corsi Test* [Data set]. https://doi.org/10.1037/t63810-000

Costikyan, G. (2002). I have no words & I must design: Toward a Critical Vocabulary for Games. In F. Mäyrä (Ed.), *Proceedings of Computer Games and Digital Cultures Conference*. Tampere University Press. https://www.academia.edu/1270811/I_have_no_words_and_I_must_design

Crawford, C. (1984). *The Art of Computer Game Design*. McGraw-Hill.

Crombie, W., Moffat, D., & Shabalina, O. (2016). Video games can temporarily increase creativity; especially puzzle games. *Proceedings of the 10th European Conference on Games Based Learning (ECGBL 2016)*, 152–158.

Crossley, N. (2022). Musicking to Music Worlds: On Christopher Small's Important Innovation. *Music Research Annual*, *3*, 1–23.

Croteam. (2014). *The Talos Principle* [Video game].

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York : Harper & Row.

daynthelife. (2022, November 21). *Question about the shipwreck puzzle* [Reddit Post]. R/TheWitness. https://www.reddit.com/r/TheWitness/comments/z17g8h/question_about_the_shipwreck_puzzle/

Deci, E., & Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US. https://doi.org/10.1007/978-1-4899-2271-7

Denisova, A., & Cairns, P. (2019). Player experience and deceptive expectations of difficulty adaptation in digital games. *Entertainment Computing*, *29*, 56–68. https://doi.org/10.1016/j.entcom.2018.12.001

Denisova, A., Cairns, P., Guckelsberger, C., & Zendle, D. (2020). Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies*, *137*, 102383. https://doi.org/10.1016/j.ijhcs.2019.102383

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining 'gamification'. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek '11*, 9–15. https://doi.org/10.1145/2181037.2181040

Deutsch, D., Hamaoui, K., & Henthorn, T. (2007). The glissando illusion and handedness. *Neuropsychologia*, *45*(13), 2981–2988. https://doi.org/10.1016/j.neuropsychologia.2007.05.015

Diakopoulos, N., Luther, K., & Essa, I. (2008). Audio Puzzler: Piecing together time-stamped speech transcripts with a puzzle game. *Proceedings of the 16th ACM International Conference on Multimedia, MM '08*, 865–868. https://doi.org/10.1145/1459359.1459507

Digital Extremes. (2013). *Warframe* [Video game].

Digital Tentacle. (2014). *Circuits* [Video game].

Dillon, R. (2011). The 6-11 Framework: A new approach to video game analysis and design. In T. C. Tien (Ed.), *GAMEON-ASIA 2011 Proceedings* (Vol. 2011, pp. 25–29). Eurosis. http://www.eurosis.org/cms/?q=taxonomy/term/263

Dillon, R. (2018). *On the Way to Fun: An Emotion-Based Approach to Successful Game Design*. A K Peters/CRC Press.

Dindar, M. (2018). An empirical study on gender, video game play, academic success and complex problem solving skills. *Computers & Education*, *125*, 39–52. https://doi.org/10.1016/j.compedu.2018.05.018

Ding, Y., Gray, K., Forrence, A., Wang, X., & Huang, J. (2018). A behavioral study on tonal working memory in musicians and non-musicians. *PLOS ONE*, *13*(8), e0201765. https://doi.org/10.1371/journal.pone.0201765

Dolsheid, S., Hunnius, S., & Majid, A. (2015). When high pitches sound low: Children's acquisition of space-pitch metaphors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *37*(0). https://escholarship.org/uc/item/1v08v2x9

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354. https://doi.org/10.1037/0033-295X.85.4.341

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422. https://doi.org/10.1016/j.cognition.2009.10.013

Ekman, I. (2007, January 1). *Sound-based Gaming for Sighted Audiences–Experiences from a Mobile Multiplayer Location Aware Game*. AudioMostly Conference.

Falstein, N. (2005). Understanding Fun—The Theory of Natural Funativity. In S. Rabin (Ed.), *Introduction to Game Development* (pp. 71–98). Charles River Media.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fernández-Vara, C. (2014). *Introduction to Game Analysis*. Routledge.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, *3*(4), 430–441. https://doi.org/10.1177/104973239300300403

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a Musical Instrument in Childhood is Associated with Enhanced Verbal Ability and Nonverbal Reasoning. *PLOS ONE*, *3*(10), e3566. https://doi.org/10.1371/journal.pone.0003566

Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence*, *56*, 58–64. https://doi.org/10.1016/j.intell.2016.02.011

Foster, N. E. V., Beffa, L., & Lehmann, A. (2021). Accuracy of Tempo Judgments in Disk Jockeys Compared to Musicians and Untrained Individuals. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.709979

Fox, T. (2015). *Undertale* [Video game].

Fox, T. (2025). *Deltarune* [Video game].

Frazier-Roberts, S., & Peoples, C. (2019). The Design of a Gamification Algorithm in a Music Practice Application. *Open Journal of Web Technologies (OJWT)*, *6*(1), 16–30.

FreshPlanet. (2012). *SongPop Classic* [Video game].

Gagnon, R., & Nicoladis, E. (2021). Musicians show greater cross-modal integration, intermodal integration, and specialization in working memory than non-musicians. *Psychology of Music*, *49*(4), 718–734. https://doi.org/10.1177/0305735619896088

Gamerpillar. (2023, November 18). *Tadpole Pond Musical Puzzles—All 3 Songs | Super Mario RPG* [Video recording]. https://www.youtube.com/watch?v=E1jgJ9rcYkk

Garcia, F. E., & de Almeida Neris, V. P. (2013). Design Guidelines for Audio Games. In M. Kurosu (Ed.), *Human-Computer Interaction. Applications and Services* (pp. 229–238). Springer. https://doi.org/10.1007/978-3-642-39262-7_26

Garg, A., Shelat, S., Gross, M. E., Smallwood, J., Seli, P., Taxali, A., Sripada, C. S., & Schooler, J. W. (2025). Opening the black box: Think Aloud as a method to study the spontaneous stream of consciousness. *Consciousness and Cognition*, *128*, 103815. https://doi.org/10.1016/j.concog.2025.103815

Gibbons, W. (2018). Gamifying Classical Music. In W. Gibbons (Ed.), *Unlimited Replays: Video Games and Classical Music* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780190265250.003.0011

Giovagnoli, A. R., & Raglio, A. (2011). Cognitive abilities of musicians. *Perceptual and Motor Skills*, *113*(2), 563–569. https://doi.org/10.2466/04.11.22.23.PMS.113.5.563-569

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.

Gomes, C., Figueiredo, M., & Bidarra, J. (2014, March 13). *Gamification in teaching music: Case study*. EduRe'14.

Gooding, P., & Terras, M. (2008). 'Grand Theft Archive': A quantitative analysis of the state of computer game preservation. *International Journal of Digital Curation*, *3*(2), 19–41. https://doi.org/10.2218/ijdc.v3i2.56

Gower, L., & McDowall, J. (2012). Interactive music video games and children's musical development. *British Journal of Music Education*, *29*(1), 91–105. https://doi.org/10.1017/S0265051711000398

Gunfire Games. (2023). *Remnant 2* [Video game]. Gearbox Publishing.

Guzmán Anaya, C. (2025). Fan Creativity, Collaboration, and Resignification in Sonic Participatory Cultures Within, Through, and Around Genshin Impact (HoYoverse, 2020). *Journal of Sound and Music in Games*, *6*(2), 28–52. https://doi.org/10.1525/jsmg.2025.6.2.28

Hanging Gardens Interactive. (2022). *Sonority* [Video game]. Application Systems Heidelberg.

Hansen, K. F., & Hiraga, R. (2017). The Effects of Musical Experience and Hearing Loss on Solving an Audio-Based Gaming Task. *Applied Sciences*, *7*(12), Article 12. https://doi.org/10.3390/app7121278

Hansen, K. F., Li, Z., & Wang, H. (2012). *A music puzzle game application for engaging in active listening*. 97th Information Science and Music (SIGMUS) Research Conference. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-109398

Harmonix. (2008). *Rock Band 2* [Video game]. MTV Games.

Haworth, R., & Sedig, K. (2011). The Importance of Design for Educational Games. In A. Méndez-Vilas (Ed.), *Education in a Technological World: Communicating Current and Emerging Research and Technological Efforts* (pp. 518–522). Formatex Research Center.

Hetland, L. (2000). Learning To Make Music Enhances Spatial Reasoning. *Journal of Aesthetic Education*, *34*, 179–238.

Hill, J. M. D., Ray, C. K., Blair, J. R. S., & Carver, C. A. (2003). Puzzles and games: Addressing different learning styles in teaching operating systems concepts. *ACM SIGCSE Bulletin*, *35*(1), 182–186. https://doi.org/10.1145/792548.611964

Hubana, R. (2022). Influences of Puzzle Videogames on Logical Reasoning. In N. Ademović, E. Mujčić, Z. Akšamija, J. Kevrić, S. Avdaković, & I. Volić (Eds), *Advanced Technologies, Systems, and Applications VI* (pp. 237–247). Springer International Publishing. https://doi.org/10.1007/978-3-030-90055-7_18

Hufkens, L. V., & Browne, C. (2019). A Functional Taxonomy of Logic Puzzles. *2019 IEEE Conference on Games (CoG)*, 1–4. https://doi.org/10.1109/CIG.2019.8848107

Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research. *AAAI Workshop - Technical Report*, *1*.

Iacovides, I., Cox, A. L., McAndrew, P., Aczel, J., & Scanlon, E. (2015). Game-Play Breakdowns and Breakthroughs: Exploring the Relationship Between Action, Understanding, and Involvement. *Human–Computer Interaction*, *30*(3–4), 202–231. https://doi.org/10.1080/07370024.2014.987347

Indomitus Games. (2015). *In Verbis Virtus* [Video game].

James, N. (2012). Videogames are Disappearing. In *Best Before*. Routledge.

Jenson, J., de Castell, S., Muehrer, R., & Droumeva, M. (2016). So you think you can play: An exploratory study of music video games. *Journal of Music, Technology and Education*, *9*, 273–288. https://doi.org/10.1386/jmte.9.3.273_1

Joey Drew Studios. (2017). *Bendy and the Ink Machine* [Video game].

Juul, J., Perron, B., & Wolf, M. J. P. (2009). Fear of Failing? The Many Meanings of Difficulty in Video Games. In *The Video Game Theory Reader 2* (pp. 237–252). Routledge.

Karhulahti, V.-M. (2013). Puzzle Is Not a Game! Basic Structures of Challenge. *Proceedings of DiGRA 2013: DeFragging Game Studies.* https://www.academia.edu/4432576/Puzzle_Is_Not_a_Game_Basic_Structures_of_Challenge

Kassabian, A., & Jarman, F. (2016). Game and Play in Music Video Games. In M. Kamp, T. Summers, & M. Sweeney (Eds), *Ludomusicology: Approches to Video Game Music* (pp. 116–132). Equinox Publishing.

KatGames. (2009). *Dream Chronicles: The Chosen Child* [Video game]. PlayFirst.

Kaviani, H., Mirbaha, H., Pournaseh, M., & Sagan, O. (2014). Can music lessons increase the performance of preschool children in IQ tests? *Cognitive Processing*, *15*(1), 77–84. https://doi.org/10.1007/s10339-013-0574-0

Kim, J., & Ricaurte, J. (2011). TapBeats: Accessible and mobile casual gaming. *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 285–286. ASSETS '11: The 13th International ACM SIGACCESS Conference on Computers and Accessibility. https://doi.org/10.1145/2049536.2049609

Kim, J.-Y., & Belkin, N. (2002). *Categories of Music Description and Search Terms and Phrases Used by Non-Music Experts*. International Society for Music Information Retrieval Conference. https://www.semanticscholar.org/paper/Categories-of-Music-Description-and-Search-Terms-by-Kim-Belkin/1187d5b34fda59fc1e65b281d37e6a03aa886043

Kim, S. (2004). What is a Puzzle? *The Games Cafe*. https://cs.wellesley.edu/~cs215/Lectures/L17-IntroGamesJigsawPuzzle/ScottKim-What_is_a_Puzzle.pdf

Kinetic Games. (2020). *Phasmophobia* [Video game].

Kirke, A. (2018). When the Soundtrack Is the Game: From Audio-Games to Gaming the Music. In D. Williams & N. Lee (Eds), *Emotion in Video Game Soundtracking* (pp. 65–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-72272-6_7

KranX Productions. (2008). *Musaic Box* [Video game].

Krekhov, A., Emmerich, K., Rotthaler, R., & Krueger, J. (2021). Puzzles Unpuzzled: Towards a Unified Taxonomy for Analog and Digital Escape Room Games. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 269:1-269:24. https://doi.org/10.1145/3474696

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

LB ART. (2021). *Don't Forget the Lyrics* [Video game].

Legacy Gaming. (2023). *Remnant 2—Secrets of the Water Harp | All Solutions, Rewards, Loot & Secrets* [Video recording]. https://www.youtube.com/watch?v=LO9Cj0jHRc0

Lesser, A. J. (2020). An Investigation of Digital Game-Based Learning Software in the Elementary General Music Classroom. *Journal of Sound and Music in Games*, *1*(2), 1–24. https://doi.org/10.1525/jsmg.2020.1.2.1

Li, D., Cowan, N., & Saults, J. S. (2013). Estimating Working Memory Capacity for Lists of Nonverbal Sounds. *Attention, Perception & Psychophysics*, *75*(1), 145–160. https://doi.org/10.3758/s13414-012-0383-z

Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S., Xiang, J., Fu, Q.-J., & Zhang, F. (2016). Musicians Are Better than Non-musicians in Frequency Change Detection: Behavioral and Electrophysiological Evidence. *Frontiers in Neuroscience*, *10*, 464. https://doi.org/10.3389/fnins.2016.00464

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189

Lim, K. Y. T., Truong, K. M., & Wu, Y. (2023). Melody Mystery: Learning Music Theory through Escape Room Puzzles. *Education Sciences*, *13*(5), 491. https://doi.org/10.3390/educsci13050491

Linehan, C., Bellord, G., Kirman, B., Morford, Z. H., & Roche, B. (2014). Learning curves: Analysing pace and challenge in four successful puzzle games. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '14*, 181–190. https://doi.org/10.1145/2658537.2658695

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479–1498.

Lonely Few. (2013). *Blendoku* [Video game].

Lotz. (2021). *Partitas* [Video game].

Margolin, U. (2005). Russian Formalism. In M. Groden, M. Kreiswirth, & I. Szeman (Eds), *The Johns Hopkins guide to literary theory & criticism* (pp. 814–820). Johns Hopkins University Press.

McAuley, J. D., Henry, M. J., & Tuft, S. (2011). Musician advantages in music perception: An issue of motivation, not just ability. *Music Perception*, *28*(5), 505–518. https://doi.org/10.1525/mp.2011.28.5.505

McCrindle, R., & Symons, D. (2000). Audio space invaders. *Proceedings of the 3rd International Conference on Disability, Virtual Reality and Associated Technologies*, *3*. https://www.icdvrat.com/2000/P/2000_09.pdf

Melhart, D., Togelius, J., Mikkelsen, B., Holmgård, C., & Yannakakis, G. N. (2024). The Ethics of AI in Games. *IEEE Transactions on Affective Computing*, *15*(01), 79–92. https://doi.org/10.1109/TAFFC.2023.3276425

Mi, Q., & Gao, T. (2025). Engagement-Oriented Dynamic Difficulty Adjustment. *Applied Sciences*, *15*(10). https://doi.org/10.3390/app15105610

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Mizener, B., Vandenberghe-Descamps, M., Abdi, H., & Chollet, S. (2022). Musical Listening Qualia: A Multivariate Approach. *Auditory Perception & Cognition*, *5*(1–2), 46–75. https://doi.org/10.1080/25742442.2022.2036074

Mobius Digital. (2020). *Outer Wilds* [Video game]. Annapurna Interactive.

Moore, S., & Smallwood, S. (2022). The Way In: Videogame as Composition and Composition as Videogame. *Interference Journal*. https://www.academia.edu/96189026/The_Way_In_Videogame_as_Composition_and_Composition_as_Videogame

Mora-Cantallops, M. (2018). Transhistorical perspective of the puzzle video game genre. *Proceedings of the 13th International Conference on the Foundations of Digital Games, FDG '18*, 1–9. https://doi.org/10.1145/3235765.3235768

Moss, R. (2011, January 26). A truly graphic adventure: The 25-year rise and fall of a beloved genre. *Ars Technica*. https://arstechnica.com/gaming/2011/01/history-of-graphic-adventures/

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Naughty Dog. (2003). *Jak II* [Video game]. Sony Computer Entertainment.

Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience & Biobehavioral Reviews*, *140*, 104777. https://doi.org/10.1016/j.neubiorev.2022.104777

New Dawn Game Studio. (2023). *Square Logic* [Video game].

NEXT Studios. (2019). *Unheard—Voices of Crime* [Video game].

Nintendo. (1998). *The Legend of Zelda: Ocarina of Time* [Video game].

Nintendo. (2000). *The Legend of Zelda: Majora's Mask* [Video game].

Nintendo. (2002). *The Legend of Zelda: The Wind Waker* [Video game].

Nintendo. (2004). *Paper Mario: The Thousand-Year Door* [Video game].

Oliva-Zamora, M. Á., & Mangiron, C. (2025). How to create video games with cognitive accessibility features: A literature review of recommendations. *Universal Access in the Information Society*, *24*(3), 2805–2818. https://doi.org/10.1007/s10209-025-01231-5

Östblad, P. A., Engström, H., Brusk, J., Backlund, P., & Wilhelmsson, U. (2014). Inclusive game design: Audio interface in a graphical adventure game. *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound, AM '14*, 1–8. https://doi.org/10.1145/2636879.2636909

Pajitnov, A. (1985). *Tetris* [Video game].

Paney, A. S. (2015). Singing video games may help improve pitch-matching accuracy. *Music Education Research*, *17*(1), 48–56. https://doi.org/10.1080/14613808.2014.969218

Partitas on Steam. (n.d.). Retrieved 14 February 2026, from https://store.steampowered.com/app/1590270/Partitas/

Pasinski, A. C., Hannon, E. E., & Snyder, J. S. (2016). How musical are music video game players? *Psychonomic Bulletin & Review*, *23*(5), 1553–1558. https://doi.org/10.3758/s13423-015-0998-x

Peppler, K., Downton, M., & Lindsay, E. (2011). The Nirvana Effect: Tapping Video Games to Mediate Music Learning and Interest. *International Journal of Learning and Media*, *3*, 41–59. https://doi.org/10.1162/ijlm_a_00062

Phonopath. (2014, October 23). *Phonopath—IGF Crash Course Introduction* [Video recording]. https://www.youtube.com/watch?v=q8g3v5ApGuI

Pietsch, S., & Jansen, P. (2012). Different mental rotation performance in students of music, sport and education. *Learning and Individual Differences*, *22*(1), 159–163. https://doi.org/10.1016/j.lindif.2011.11.012

Pilchmair, M., & Kayali, F. (2007). Levels of Sound: On the Principles of Interactivity in Music Video Games. *Situated Play, Proceedings of DiGRA 2007 Conference*, 424–430.

Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.

Pomini, E., Pecunioso, A., & Agrillo, C. (2025). Musicians are faster to process hierarchical Navon letters. *Psychology of Music*, *53*(4), 643–652. https://doi.org/10.1177/03057356251320975

Pusey, M., Wong, K. W., & Rappa, N. A. (2021). The Puzzle Challenge Analysis Tool. A Tool for Analysing the Cognitive Challenge Level of Puzzles in Video Games. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 276:1-276:27. https://doi.org/10.1145/3474703

Pusey, M., Wong, K. W., & Rappa, N. A. (2022). Using Case Studies to Explore Need Satisfaction and Frustration in Puzzle Video Games. *Games and Culture*, *17*(5), 752–772. https://doi.org/10.1177/15554120211056126

Rare. (1998). *Banjo-Kazooie* [Video game]. Nintendo.

Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., & Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research*, *19*(1), 2–8. https://doi.org/10.1080/01616412.1997.11740765

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of Nonverbal Assessment* (pp. 223–237). Springer US. https://doi.org/10.1007/978-1-4615-0153-4_11

Red Candle Games. (2024). *Nine Sols* [Video game].

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J.-C., McTear, M., Raman, T., Stanney, K. M., Su, H., & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Commun. ACM*, *47*(1), 57–59. https://doi.org/10.1145/962081.962106

Regamey, K. (2011). *Phonopath* [Video game].

Röber, N., & Masuch, M. (2004, July 6). Interacting With Sound: An Interaction Paradigm for Virtual Auditory Worlds. *Proceedings of the 10th Meeting of the International Conference on Auditory Display*.

Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities—Findings from a longitudinal study. *Applied Cognitive Psychology*, *28*(4), 545–557. https://doi.org/10.1002/acp.3034

Rollings, A., & Adams, E. (2003). Gameplay. In *Andrew Rollings and Ernest Adams on Game Design* (p. 648). New Riders Press.

Román-Caballero, R., Vadillo, M. A., Trainor, L. J., & Lupiáñez, J. (2022). Please don't stop the music: A meta-analysis of the cognitive and academic benefits of instrumental musical training in childhood and adolescence. *Educational Research Review*, *35*, 100436. https://doi.org/10.1016/j.edurev.2022.100436

Romero-Mendez, E. A., Santana-Mancilla, P. C., Garcia-Ruiz, M., Montesinos-López, O. A., & Anido-Rifón, L. E. (2023). The Use of Deep Learning to Improve Player Engagement in a Video Game through a Dynamic Difficulty Adjustment Based on Skills Classification. *Applied Sciences*, *13*(14). https://doi.org/10.3390/app13148249

Roshanian Ramin, M., Aliabadi, K., Delavar, A., & Rastegarpoor, H. (2022). The effect of puzzle games on the creativity of students. *New Media Studies*, *8*(32), 180–155. https://doi.org/10.22054/nms.2022.44342.774

Ross, J. M. (2002). Guiding students through programming puzzles: Value and examples of Java game assignments. *ACM SIGCSE Bulletin*, *34*(4), 94–98. https://doi.org/10.1145/820127.820175

Rovithis, E. (2012). A classification of audio-based games in terms of sonic gameplay and the introduction of the audio-role-playing-game: Kronos. *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound, AM '12*, 160–164. https://doi.org/10.1145/2371456.2371483

Rovithis, E., Floros, A., Mniestris, A., & Grigoriou, N. (2014). Audio games as educational tools: Design principles and examples. *2014 IEEE Games Media Entertainment*, 1–8. https://doi.org/10.1109/GEM.2014.7048083

Rovithis, E., Mniestris, A., & Floros, A. (2014). Educational audio game design: Sonification of the curriculum through a role-playing scenario in the audio game 'Kronos'. *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound, AM '14*, 1–6. https://doi.org/10.1145/2636879.2636902

Rovithis, E., Papadopoulou, A., Komianos, V., Garneli, V., & Floros, A. (2024). Speech Puzzles (Spuzzles): Engaging the Reduced, Causal, and Semantic Listening Modes for Puzzle Design in Audio Games. *Applied Sciences*, *14*(9), 3858.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R., Rigby, C., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*, 344–360. https://doi.org/10.1007/s11031-006-9051-8

Sala, G., & Gobet, F. (2017a). Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training. *Current Directions in Psychological Science*, *26*(6), 515–520. https://doi.org/10.1177/0963721417712760

Sala, G., & Gobet, F. (2017b). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, *20*, 55–67. https://doi.org/10.1016/j.edurev.2016.11.005

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2

Salter, A. (2014). *What Is Your Quest?: From Adventure Games to Interactive Books*. University of Iowa Press. https://muse.jhu.edu/pub/165/monograph/book/34990

Samat, J., Baharum, A., & Andin, C. (2022). Identifying Elements of Gamification for Reading Music Notation in Music Education. *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 563–567. https://doi.org/10.1109/ICTC55196.2022.9952727

Sarathy, V., Rabb, N., Kasenberg, D. M., & Scheutz, M. (2024). Using Puzzle Video Games to Study Cognitive Processes in Human Insight and Creative Problem-Solving.

Proceedings of the 46th Annual Conference of the Cognitive Science Society, *46*, 3004–3010.

Schaeffer, P. (1966). *Traité des objets musicaux*. Paris : Éditions du Seuil.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*(2), 457–468. https://doi.org/10.1037/0022-0663.98.2.457

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology (London, England: 1953)*, *102*(3), 283–302. https://doi.org/10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*(Volume 75, 2024), 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music*, *38*(2), 209–221. https://doi.org/10.1177/0305735609339473

Sepulveda, G. K., Besoain, F., & Barriga, N. A. (2019). Exploring Dynamic Difficulty Adjustment in Videogames. *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 1–6. https://doi.org/10.1109/CHILECON47746.2019.8988068

Shepard, R. N. (1964). Circularity in Judgments of Relative Pitch. *The Journal of the Acoustical Society of America*, *36*(12), 2346–2353. https://doi.org/10.1121/1.1919362

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, N.Y.)*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Shklovsky, V. (1965). Art as Technique (L. T. Lemon & M. J. Reis, Trans.). In *Russian Formalist Criticism: Four Essays* (pp. 3–24). University of Nebraska Press. (Original work published 1917)

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Shute, V. J., & Wang, L. (2015). Measuring Problem Solving Skills in Portal 2. In P. Isaías, J. M. Spector, D. Ifenthaler, & D. G. Sampson (Eds), *E-Learning Systems, Environments and Approaches: Theory and Implementation* (pp. 11–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-05825-2_2

Sierra On-Line. (1994). *King's Quest VII: The Princeless Bride* [Video game].

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's Area Supports Enhanced Visuospatial Cognition in Orchestral Musicians. *Journal of Neuroscience*, *27*(14), 3799–3806. https://doi.org/10.1523/JNEUROSCI.0147-07.2007

Small, C. (1998). *Musicking: The meanings of performing and listening*. University Press of New England.

Solo Philosopher. (2022). *Hear Your Path* [Video game].

Somethin' Else. (2010). *Papa Sangre* [Video game].

Somethin' Else. (2013). *Papa Sangre II* [Video game].

Sonority on Steam. (n.d.). Retrieved 14 February 2026, from https://store.steampowered.com/app/1432390/Sonority/

Square. (1994). *Final Fantasy VI* [Video game].

Square. (1996). *Super Mario RPG* [Video game]. Nintendo.

Steam Community: Circuits. (n.d.). Retrieved 14 February 2026, from https://steamcommunity.com/app/282760/reviews/?browsefilter=toprated&snr=1_5_100010_

Steam Community: Partitas. (n.d.). Retrieved 14 February 2026, from https://steamcommunity.com/app/1590270/reviews/?browsefilter=toprated&snr=1_5_100010_

Steam Community: Sonority. (n.d.). Retrieved 14 February 2026, from https://steamcommunity.com/app/1432390/reviews/?browsefilter=toprated&snr=1_5_100010_

Stockfelt, O. (1997). Adequate Modes of Listening (A. Kassabian & L. G. Svendsen, Trans.). In D. Schwarz, A. Kassabian, & L. Siegel (Eds), *Keeping Score: Music, Disciplinarity, Culture* (pp. 129–146). University Press of Virginia.

Strauss, H., Reiche, S., Dick, M., & Zentner, M. (2024). Online assessment of musical ability in 10 minutes: Development and validation of the Micro-PROMS. *Behavior Research Methods*, *56*(3), 1968–1983. https://doi.org/10.3758/s13428-023-02130-4

Summers, T. (2016). *Understanding Video Game Music*. Cambridge University Press.

Swaminathan, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude. *Intelligence*. https://doi.org/10.1016/J.INTELL.2017.03.005

Targett, S., & Fernström, M. (2003). Audio games: Fun for all? All for fun. *Proceedings of the 2003 International Conference on Auditory Display*, 216–219.

Team Cherry. (2025). *Hollow Knight: Silksong* [Video game].

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. https://doi.org/10.1007/s00221-004-2044-5

Thekla, Inc. (2016). *The Witness* [Video game].

Thinking Rabbit. (1982). *Sokoban* [Video game].

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Tong, T., Zingaro, D., & Engels, S. (2014). Design guidelines for audio-based game features. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '14*, 443–444. https://doi.org/10.1145/2658537.2661307

Total Mayhem Games. (2022). *We Were Here Forever* [Video game].

Toups, Z. O., Hammer, J., Hamilton, W. A., Jarrah, A., Graves, W., & Garretson, O. (2014). A framework for cooperative communication game mechanics from grounded theory. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '14*, 257–266. https://doi.org/10.1145/2658537.2658681

Trilobyte. (1993). *The 7th Guest* [Video game]. Virgin Interactive Entertainment.

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion (Washington, D.C.)*, *8*(6), 838–849. https://doi.org/10.1037/a0014080

Ubisoft Montreal. (2003). *Prince of Persia: The Sands of Time* [Video game]. Ubisoft.

Urbanek, M., & Güldenpfennig, F. (2019). Celebrating 20 Years of Computer-based Audio Gaming. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound, AM '19*, 90–97. https://doi.org/10.1145/3356590.3356605

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Vassena, E., Kochman, K., Latomme, J., & Verguts, T. (2016). Unimodal and cross-modal prediction is enhanced in musicians. *2045-2322*. https://doi.org/10.1038/srep25225

Vertigo Games. (2023). *The 7th Guest VR* [Video game].

Visani Scozzi, M., Iacovides, I., & Linehan, C. (2017). A Mixed Method Approach for Evaluating and Improving the Design of Learning in Puzzle Games. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '17*, 217–228. https://doi.org/10.1145/3116595.3116628

Víteková, L., Eichhorn, C., Pirker, J., & Plecher, D. A. (2026). Dynamic Difficulty Adjustment in Serious Games: A Literature Review. *Information*, *17*(1). https://doi.org/10.3390/info17010096

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised* [Data set]. https://doi.org/10.1037/t48859-000

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence—Second Edition* [Data set]. https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft15171-000

Weng, T.-S. (2022). Enhancing Problem-Solving Ability through a Puzzle-Type Logical Thinking Game. *Scientific Programming*, *2022*(1), 7481798. https://doi.org/10.1155/2022/7481798

Williams, S. M. (1994). Perceptual Principles in Sound Grouping. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 95–125). Addison-Wesley Publishing Company.

Winston, J. L., Jazwinski, B. M., Corey, D. M., & Colombo, P. J. (2022). Music Training, and the Ability of Musicians to Harmonize, Are Associated With Enhanced Planning and Problem-Solving. *Frontiers in Psychology*, *12*, 805186. https://doi.org/10.3389/fpsyg.2021.805186

Wishes Ultd., Lobanov, G., & A Shell in the Pit. (2018). *Wandersong* [Video game]. Greg Lobanov.

Wolf, M. J. P. (2001). The Medium of the Video Game. In M. J. P. Wolf (Ed.), *Genre and the Video Game*. University of Texas Press. https://doi.org/10.7560/791480

Wolf, M. J. P. (2008). *The Video Game Explosion: A History from PONG to PlayStation and Beyond*. Bloomsbury Academic.

Wuertz, J., Alharthi, S. A., Hamilton, W. A., Bateman, S., Gutwin, C., Tang, A., Toups, Z., & Hammer, J. (2018). A Design Framework for Awareness Cues in Distributed Multiplayer Games. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, 1–14. https://doi.org/10.1145/3173574.3173817

XD Games Studio. (2021). *Scale Enchanter* [Video game].

XD Games Studio. (2022). *Chords Enchanter* [Video game].

Xue, S., Wu, M., Kolen, J., Aghdaie, N., & Zaman, K. (2017). *Dynamic Difficulty Adjustment for Maximized Engagement in Digital Games* (p. 471). https://doi.org/10.1145/3041021.3054170

Yuan, B., Folmer, E., & Harris, F. C. (2011). Game accessibility: A survey. *Univers. Access Inf. Soc.*, *10*(1), 81–100. https://doi.org/10.1007/s10209-010-0189-5

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

Zombie LLC. (1996). *Zork Nemesis* [Video game]. Activision.