Informatics

Analyzing the Learning Curve: The Influence of Multimodality And Gamification Feedback on Motor Skill Acquisition

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Ana Vesić, B.S.E.
Matrikelnummer 12229250

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioğlu
Mitwirkung: Ambika Shahu, M.Sc.

Wien, 15. Februar 2025

Ana Vesić   Hilda Tellioğlu

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU WIEN Informatics

Analyzing the Learning Curve: The Influence of Multimodality And Gamification Feedback on Motor Skill Acquisition

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Media and Human-Centered Computing

by

Ana Vesić, B.S.E.
Registration Number 12229250

to the Faculty of Informatics

at the TU Wien

Advisor: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioğlu
Assistance: Ambika Shahu, M.Sc.

Vienna, February 15, 2025

Ana Vesić Hilda Tellioğlu

Erklärung zur Verfassung der Arbeit

Ana Vesić, B.S.E.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel" habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 15. Februar 2025

Ana Vesić

Danksagung

Mein Masterstudium an der TU Wien war eine Herausforderung, etwas Neues und Aufregendes, eine Gelegenheit zur persönlichen Weiterentwicklung, aber auch eine Lebensaufgabe – der Umzug in ein anderes Land und das Erstreben von Erfolg auf eigene Faust. Dafür brauchte ich viel Unterstützung und hatte glücklicherweise genau diese, nicht nur während des Studiums selbst, sondern auch während der Arbeit an meinem Masterprojekt und meiner Masterarbeit. Meine Familie, darunter meine Mutter Jasna, mein Vater Ljutvija und meine Schwester Lana, sind meine größten Unterstützer und meine Stärke, und ich bin ihnen unendlich dankbar, dass sie mich in jeder Hinsicht ermutigt und unterstützt haben. Dafür, dass sie an mich geglaubt, mich angefeuert und immer für mich da waren.

Auch von den Menschen der TU Wien erhielt ich großartige und entscheidende Unterstützung. Ich bin dem Helmut Veith Stipendienkomitee zutiefst dankbar, dass es mir die Möglichkeit gegeben hat, nach Wien zu ziehen und an der TU Wien zu studieren. Es war eine große Ehre und ein Privileg, ein wahr gewordener Traum. Ich bin nicht nur dankbar, sondern wünsche mir auch, den Erwartungen und der Ehre gerecht zu werden, die mir zuteil wurden, und diese Botschaft und Gelegenheit an andere Studentinnen weiterzugeben. In diesem Prozess möchte ich mich auch bei meiner Mentorin Hilda Tellioğlu bedanken, die mich geführt, unterstützt und mich an einen Standard gebunden hat, den ich mir erhofft hatte, als ich an die TU Wien kam – nicht nur als Professorin, sondern auch als Mentorin und Mensch. Ich werde die Unterstützung und die Chancen, die mir die TU Wien als Ganzes gegeben hat, immer in Erinnerung behalten und schätzen.

Ich möchte auch meinen Freunden in Wien und Belgrad danken, die an mich geglaubt und mich angefeuert haben, sowie meinem Partner, der mich auf jedem Schritt des Weges unterstützt hat.

Vielen Dank!

Acknowledgements

My masters studies at TU Wien were something challenging, new and exciting, an opportunity for growth and improvement, but also a life challenge of moving to another country and thriving on my own. For this I needed and thankfully had a lot of support, not just during the studies themselves but also during working on my masters project and thesis as well.
My family including my mother Jasna, father Ljutvija, and sister Lana, are my greatest supporters and strength and I am immensely thankful to them for encouraging and supporting me in every way I needed. For believing in me, rooting for me and being there for me.
I also had a great, and crucial support from the people of TU Wien as well. I am immensely thankful to Helmut Veith Stipend committee for giving me the opportunity and believing in me to move and study in Vienna, at TU Wien. It was a great honor and privilege, a real dream come true. I am not only thankful, but also wish to fulfill the expectations and the honor I was given, and send the message and opportunity further to other female students. In this process, I wish to thank my mentor Hilda Tellioğlu as well, for guiding me, helping me and holding me to a standard I hoped for while coming to TU Wien, not only as professor, but also a mentor and a fellow human. I will always remember and cherish the support and opportunities given to me by the TU Wien collective as a whole.
I also want to thank my friends, in Vienna and Belgrade for believing and rooting for me, as well as my partner, for supporting me every step of the way.

Thank you!

Kurzfassung

Das motorische Lernen umfasst das Erwerben von Fähigkeiten, die eine freiwillige Kontrolle von Gelenk- und Körpersegmentbewegungen erfordern, um spezifische Ziele zu erreichen [Mag07]. Dieser Prozess ist von Natur aus komplex und zeitaufwendig und erfordert konsequentes Üben. Motorische Fähigkeiten umfassen eine Vielzahl von Aufgaben, von Musikinstrumenten bis hin zu sportlichen Aktivitäten, wodurch ihre Aneignung ein wesentlicher Bestandteil des täglichen Lebens ist. Die Entwicklung von Systemen, die diesen Prozess beschleunigen können, wäre für Lernende von großem Nutzen.

Diese Masterarbeit untersucht den Einfluss verschiedener Feedbackarten auf das motorische Lernen, mit einem Schwerpunkt auf Handgesten-Choreographien (Hand Mudra), die sowohl als Yoga-Praxis als auch als Tanzchoreographie angesehen werden. Die Studie vergleicht visuelles Feedback (Videos und/oder Bilder), auditives Feedback (Musik und Audioanweisungen), multimodales Feedback sowie gamifiziertes Feedback, um deren Effektivität zu analysieren. Die Forschung widmet sich vier zentralen Fragen: 1. Welche Art von Feedback (auditiv, visuell oder gamifiziert) hat den positivsten Einfluss auf die Aneignung motorischer Fähigkeiten? 2. Wie unterscheiden sich Lernkurven bei motorischen Gedächtnisaufgaben mit und ohne Gamifizierung? 3. In welchen Lernphasen sollte Gamifizierung ein- oder ausgeschlossen werden, um die Fähigkeitsaneignung zu optimieren? 4. Wie unterscheiden sich wahrgenommene Freude und Anstrengung zwischen den verschiedenen Lernmodi, und wie hängen diese mit den Lernkurven zusammen?

Für diese Studie wurde eine maßgeschneiderte Anwendung entwickelt, die drei Lernmodi bietet: visuell, auditiv und gamifiziert. Quantitative Daten wurden mithilfe standardisierter Fragebögen (NASA-TLX und IMI) und Tests, die während des Lernprozesses durchgeführt wurden, erhoben. Qualitative Daten wurden durch Interviews mit den Teilnehmern:innen gesammelt.

Die Ergebnisse zeigen, dass gamifizierte Lernmodi die schlechtesten Leistungen und den höchsten Anstrengungsgrad verursachten, während multimodales Feedback, das visuelle und auditive Elemente kombiniert, die besten Ergebnisse sowohl hinsichtlich Leistung als auch Freude erzielte. Der Hauptbeitrag dieser Forschung ist ein Softwaresystem, das für effektives motorisches Lernen, insbesondere in den frühen Phasen der Fähigkeitsaneignung, entwickelt wurde, sowie Richtlinien für die Gestaltung solcher Systeme. Systeme sollten Flexibilität bieten, sodass Lernende verschiedene Modi kombinieren, Videos bearbeiten, Notizen machen und Aufgaben segmentieren können. Gamifizierung sollte schrittweise eingeführt werden, mit dem Fokus auf positivem Feedback, um Selbstvertrauen und Motivation zu stärken. Gamifizierte Lernmodi sind am besten in späteren Phasen geeignet, um die Beherrschung der Fähigkeiten und die Leistungssicherheit zu unterstützen.

Schlüsselwörter: *motorische Fertigkeit, Lernen, motorisches Lernen, Fertigkeitserwerb, frühes Lernen, visuelles Feedback, auditives Feedback, multimodales Feedback, Gamification, gamifiziertes Lernen*

Abstract

Motor learning involves acquiring skills that require voluntary control of joint and body segment movements to achieve specific goals [Mag07]. This process is inherently complex and time-consuming, requiring consistent practice. Motor skills encompass diverse tasks, from playing an instrument to engaging in sports, making their acquisition an integral part of everyday life. Developing systems to expedite this process could significantly benefit learners.

This master's thesis investigates the impact of different feedback types on motor skill learning, with a focus on hand gesture choreography (Hand Mudra), which is regarded as both a yoga practice and a dance choreography. The study compares visual feedback (video and/or images), auditory feedback (music and audio instructions), multimodal feedback, and gamified feedback to analyze their effectiveness. The research addresses four key questions: 1. Which type of feedback (audio, visual, or gamified) most positively impacts motor skill acquisition? 2. How do learning curves vary in motor memory tasks with and without gamification? 3. At which stages of learning should gamification be included or excluded to optimize skill acquisition? 4. How do perceived enjoyment and effort differ across learning modes, and how do they relate to learning curves?

For this study, a custom application was developed, offering three learning modes: visual, auditory, and gamified. Quantitative data were collected through standardized questionnaires (NASA-TLX and IMI) and tests administered during the learning process. Qualitative data were gathered through participant interviews.

Results indicate that gamified learning modes resulted in the lowest performance and highest levels of effort, while multimodal feedback combining visual and auditory elements produced the best outcomes, both in terms of performance and enjoyment levels. The main contribution of this research is a software system designed for effective motor skill learning, particularly in the early stages of skill acquisition, as well as guidelines for designing such systems. Systems should offer flexibility, enabling learners to combine different modes as they learn, manipulate videos, take notes, and segment tasks. Gamification should be introduced slowly, focusing on providing positive feedback to boost confidence and motivation. Gamified modes of learning are best employed in later stages to support skill mastery and performance confidence.

Keywords: *motor skill, learning, motor learning, skill acquisition, early stage learning, visual feedback, audio feedback, multimodal feedback, gamification, gamified learning*

Contents

CHAPTER 1

Introduction

1.1 Motivation

Motor skill acquisition is an umbrella term for various different skills starting from playing an instrument, typing, performing a dance choreography to complex sports activities such as snowboarding, to posture and rehabilitation cases when re-learning or improving walking [SJHB09, HBDH10]. The widespread presence of motor skills and their necessity in the everyday lives of individuals and hence society highlights the importance and potential of their improvement and optimization. Motor skill, as mentioned, does not necessarily have to be a skill gained from a hobby; it includes proper posture or the performance of any physical activity. When understanding the essence of motor skills that extend to the working environment, we see not only the wellness but also the monetary importance of motor skills and their optimization. Injuries at work are one of the leading causes of paid leave and losses of monetary gains for companies where physical labor and proper motor skills play a key role in the optimal work process [OO23]. Musculoskeletal disorders (MSDs) contribute to 15% of productivity losses in the EU, with national studies highlighting significant economic impacts, such as Germany's EUR 17.2 billion in production loss and EUR 30.4 billion in loss of gross value added in 2016, representing 0.5% and 1.0% of the GDP, respectively [fSaWdK+19].

In order to prevent or overcome difficulties that can negatively affect one's health it is necessary to be able to work on motor skills, whether that includes learning a new motor skill or improving the ones already learned. Motor (physical) learning is the ability to acquire motor skills that require voluntary control over movements of the joints and body segments to achieve a goal [Mag07]. The process of acquiring motor skills is fundamental to numerous real-world activities, ranging from sports to performing arts, rehabilitation, and daily life tasks. Understanding and optimizing this process is crucial for enhancing performance and efficiency in these domains. Motor skill acquisition is inherently complex and time-consuming, requiring repeated practice and often guided feedback to achieve proficiency. This feedback can take various forms, such as visual, auditory, or multimodal combinations, and its effectiveness can significantly impact the learning curve and time required by an individual to adopt the skill.

For the purpose of learning and improving motor skills it is important to keep the learners interested, engaged and motivated. For this purpose many different modes of learning are used, and one of the new emerging ones is gamification. Gamification is a concept of adding game elements [DDKN11], also known as “serious game” design or “persuasive game” design [VVAvdK13]. Enjoyment in the learning process is something that is found to be very important, it acts as a natural reward, and studies in motor learning indicate that such rewards can boost motor skills performance. [CHQK13, RLGS15, SHA16]. One reason why gamification has the power to enhance enjoyment is that it can fulfill the basic psychological needs of autonomy, competence and relatedness described by Self Determination Theory [RD00a]. For this reason gamification and enjoyment are a part of researching the learning curve in the motor skill learning process, researching to what extend they have an impact. In work by Sailer et al. [SH20] it has been found that the learning environments enhanced with game design elements can potentially affect learning outcomes. Additionally, from the perspective of self-determination theory, different types of feedback play crucial roles in the learning process and are often triggered by game design elements. Continuous feedback to learners is a fundamental feature of serious games [Pre01, WH12, WNOS13, SH20].

Furthermore, topics of enjoyment and effort are very closely connected and have significant influence on learning process, knowledge or skill adoption and performance at work. Understanding what spikes enjoyment and decreases the effort and how the two relate is of crucial importance in order to properly approach and address the learning process for best results in both the process itself and the outcome. Understanding what modes of learning provide optimal combination of enjoyment and effort and how they compare among each other can also lead to important conclusions when making a decision how to design a learning process of a new motor skill.

In summary, understanding how different feedback mechanisms affect motor skill acquisition, with a particular emphasis on the role of gamification, as well as on the perceived enjoyment and effort can enable us to draw useful conclusions and improve such an important process of motor skill acquisition, which influences greatly individuals and society on every day life, and in many different professions and spheres of life.

1.2 Problem Statement

A wide array of experimental and research activities has been conducted to explore the factors influencing the process of motor skill acquisition. Nevertheless, the main focus of these papers has mostly been on the resultant outcomes, with a limited emphasis on the learning process itself.
Traditionally, the study of motor skill acquisition has focused on isolated feedback mechanisms, such as visual cues or verbal instructions, or combination of the stated.

Lately, there has been ongoing research including VR technologies, having a synergy of visual, auditory and haptic feedback. However, innovative learning methodologies have introduced new dimensions to feedback, including gamified learning environments that integrate elements of play to enhance engagement and motivation. Despite the growing interest in these novel approaches, there remains a gap in understanding the comparative analysis of different types of feedback throughout the motor learning process. More specifically, comparative analysis of more traditional modes of feedback such as visual and auditory and a different approach to learning such as gamified learning.

This thesis aims to fill this gap by systematically analyzing the impact of various feedback types — visual, visual and auditory or multimodal, and gamified — on the learning curves of individuals engaged in motor skill acquisition. Specifically, it investigates which type of feedback yields the most positive outcomes in terms of performance improvement, perceived effort, and enjoyment. By focusing on the learning of hand choreography, specifically the Hand Mudra used in yoga and dance, this research provides insights that are both practical and applicable to broader contexts.

This master thesis presents the following research questions:

- RQ1: What type of feedback (audio, visual, or gamified) has the most positive impact on motor skill acquisition?
- RQ2: How do learning curves vary in motor memory tasks involving hand gesture choreography with and without gamification?
  - RQ2.1. What considerations should be made regarding the inclusion or exclusion of gamification techniques at various stages of the learning process for hand gesture choreography?
- RQ3: How do perceived enjoyment and effort vary between different learning modes and how does it relate to the learning curve?

By identifying the most effective feedback types and their optimal introduction moments in the process, this research aims to inform the design of more efficient and engaging motor skill learning systems, ultimately improving performance and learner experience.

1.3 Contributions

The contributions of this thesis are numerous, spanning theoretical insights and practical applications in the field of motor skill learning. The main contributions are summarized as follows:

1. Software designed for the purpose of this study
2. Comparative analysis of feedback types

3. Learning curve insights
4. Gamification in motor learning
5. Design recommendations for learning systems

Software designed for the purpose of this study offers administering different modes of learning from video, video with auditory instructions to a game that assesses players' performance in real-time and provides feedback on it, using machine learning. In the development of the software, the development of the game is included as well as the real-time test mode in order to evaluate participants' performance. Developing test mode also provides the possibility of precise control and measurement of feedback types and learning outcomes. This system of machine learning based scoring and data collection enables a comprehensive quantitative data for further analysis.
Furthermore, the extensive comparative analysis of different learning modes, namely providing visual feedback, auditory with visual also regarded as multimodal, and gamified one, provides insight on most effective learning modes and guidelines for improving learning outcomes of motor skills.
Additionally and on the note of improved learning, this analysis provides detailed insights into how different feedback types influence the learning curve of motor skills.
By incorporating gamification into the learning process, this thesis explores its impact on participant engagement, motivation, and performance. It provides recommendations on the optimal timing and conditions for implementing gamified learning with the goal of maximizing its benefits.
Based on the findings, this thesis offers insights in most effective types of feedback for motor skill acquisition as well as on the most engaging, enjoyable and the ones requiring the most effort. These findings can further be extrapolated for designing more effective and engaging motor skill learning systems. These recommendations aim to improve the efficiency and enjoyment of the learning process, leading to better performance and skill retention.
By addressing these contributions, this thesis improves the understanding of motor skill acquisition and provides valuable insights for the development of more effective learning tools and methodologies.

1.4 Structure of the Thesis

The structure of the thesis is as follows. In Chapter 2 we will cover the necessary understanding of learning theories, with addition of state-of-the-art comparative analysis of different learning modes as well as role of gamification in learning of motor skills.
In Chapter 3 we will cover the software made for gamified learning, administering of other modes of learning, as well as tests, it's architecture and the utilized machine learning algorithm.
In Chapter 4 we will describe methodologies used in this research in detail, as well as the

user study design and execution.
In Chapter 5 we will talk about results, of both quantitative and qualitative methodologies, covering and comparing all different modes of learning.
Chapters 6 and 7 cover discussion and conclusion including future work, respectively.

CHAPTER 2

Background

In order to perform comprehensive literature review and pave a solid foundation for this work, a systematic search was conducted across multiple academic databases, including PubMed, IEEE Xplore, SpringerLink, ResearchGate, Google Scholar, Scopus and other, with the following keywords: "motor learning", "multimodal feedback motor skill acquisiton", "comparative feedback methods", "motor learning with video and audio", "video feedback motor learning", "gamified learning in motor skills", "serious games for motor skills", "enjoyment and effort skill acquisition", "learning curves in motor skills", "learning theories", "motor skill learning". Main focus was put on articles published between 2010 and 2023, however, for the theories of learning and other fundamental theories that time frame is expanded. Studies focusing on motor skill acquisition, feedback mechanisms, or gamification have been included, as well as peer-reviewed articles, empirical studies providing quantitative or qualitative data. The exclusion criteria were based on articles, editorials, studies not available in English and papers unrelated to motor skill learning or feedback mechanisms.

2.1 Learning theories

Here we will present several learning theories which are making the foundation for this study and for motor skill acquisition. We will start with the well known Behaviorism Learning Theory, move further to Social Learning Theory and proceed with the research done with respect to learning with different modes of learning, including gamification, with the focus on motor skill learning.

2.1.1 Behaviorism Learning Theory

Behaviorism, as one of the most prevalent schools of thought in psychology emerged at the beginning of the twentieth century. Often cited, important work by John B. Watson in 1913 [Wat13], expresses an opposition attitude towards introspective methods, focusing on the observable behavior and the environmental factors which influence it. During this time, new waves and ways of understanding social sciences, such as psychology, emerged that advocate that psychology should be grounded in empirical evidence and research, having objective measurements [AS14].

2.1.2 Watson and Little Albert Experiment

Watson is often regarded as a founding father of behaviorism. His pivotal work in this topic "Psychology as the Behaviorist Views It" [Wat13], Watson argued that psychology should abandon the study of consciousness and instead focus on behavior that can be observed and measured. He believed that behavior is a result of conditioning, a process through which organisms learn to respond to stimuli in their environment. This will later be important when we talk about motor skill acquisition. Watson demonstrated the importance of conditioning through the famous experiment with Little Albert. This experiment showed how emotional responses can be conditioned in humans. It was conducted by Watson and his student Rosalie Rayner, at John Hopkins University. For this experiment Watson conditioned a young child to fear a white rat. He did this by curating environmental conditions, so whenever the white rat appears loud, frightening noise would follow. Before the start of the experiment Little Albert, the name that was given to the child only for the context of the experiment, was given series of baseline emotional assessments by exposing the child for the first time to a white rat, a rabbit, a dog, a monkey, masks, cotton, wool, and other. During these exposure the infant did not show fear to any of the stimuli. The infant was placed on a table with a mattress on top of it in the middle of a room. The white rat was placed next to the child and the child was allowed to engage and play with the rat. As the child started engaging with the rat Watson and Rayner made a loud sound, and repeated doing so every time the child engaged with the rat. After the noise was made Albert showed signs of fear and distress by crying. When several such instances were repeated the child was presented with the rat only, after which the child started crying and crawling away, showing sever signs of distress. What happened here is that the neutral stimulus which was the rat, as proven before the experiment, after conditioned turned into the conditioned stimulus and was evoking an emotional, conditional, response which is unconditioned response made by noise which is unconditioned stimulus [Sch86] as shown in the Figure 2.1 The experiment illustrated the principles of classical conditioning, showing that emotional reactions could be learned through association.

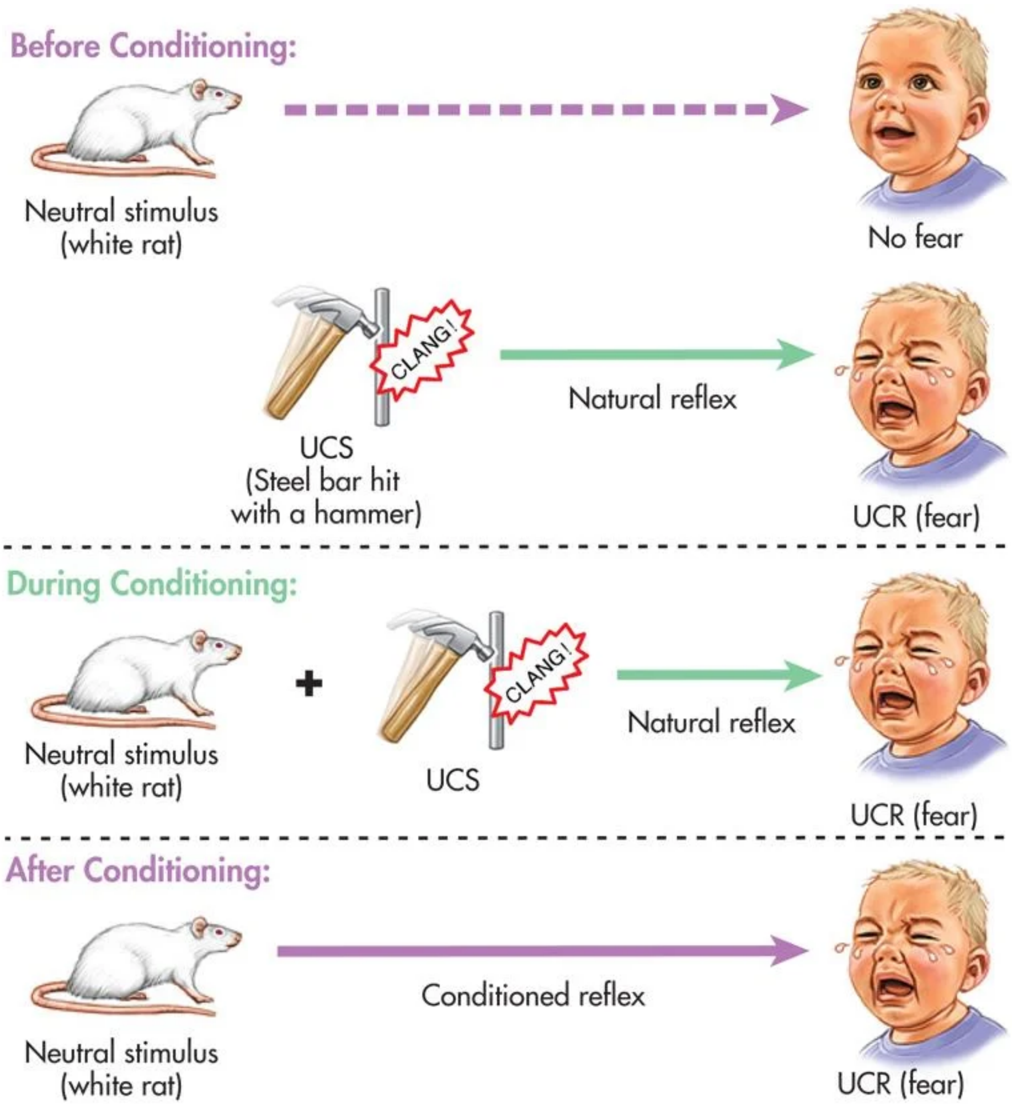


Figure 2.1: Experiment with Little Albert showing neutral stimulus, natural reflex and how their pairing yields conditioned reflex [Mcl23]

2.1.3 Pavlov's Reflex and Forward Conditioning

Furthermore, adding to the list of people who contributed to the theory is Ivan Pavlov with his famous experiment on conditioning, Pavlov's reflex. Pavlov's work presents the key and ground work for the learning theories with behaviorists. He set the foundations for classical conditioning, in his work "Conditioned Reflex" [Pav27]. In this work he proved that neutral stimulus when paired with an unconditional stimulus, neutral stimulus can be evoked. The experiment he conducted involved dogs, presenting them with food that naturally starts salivation in them. He introduced a bell, a sound of the ringing bell every time the food is introduced. After a couple of times of pairing the bell with the food, the sound of the bell alone started eliciting salivation in dogs. In this experiment the food presents unconditioned stimulus, the salivation the unconditioned response, and the bell the conditioned stimulus. After several pairings we have the conditioned response which was salivation which started every time the bell rang. Pavlov noticed this phenomena happening when the people who feed the dogs enter the room dogs started salivating, expecting food. With this experiment he proved that the behavior can be learned through association, which later became the key concept of the behaviorist theory. This is especially interesting and useful in regards to motor skill learning, and especially dance choreography which we will later cover in this chapter.

To further extend this to the study done in this thesis, it is important to mention forward conditioning. In the forward conditioning we also have conditioned stimulus and unconditioned stimulus. Key concept to understand in forward conditioning is that the conditioned stimulus comes before unconditioned stimulus, signaling that the unconditioned stimulus will follow [Cha88, CSM04]. Here we will talk only about delay conditioning, however forward conditioning has both delay and trace conditioning. With delay conditioning the conditioned stimulus overlaps with the presentation of the unconditioned stimulus. This overlapping is repeated which makes the person act the same even with the presentation of unconditioned stimulus. One of the examples can be a sound that is played simultaneously with water sprinkled to someone's face. After some time, even if only the sound is played, the person would turn away in order to escape the water. This is tightly connected with music being tied and implicating certain dance choreography moves that the person should perform.

2.1.4 Skinner's Box and Concept of Reinforcement

Something that is very important for this work and gamification is the work of B.F. Skinner. Skinner expanded Watson's and Pavlov's work and introduced the theory of operant conditioning, focusing on consequences and how they shape one's behavior. He introduced the concept of reinforcement, presenting it in his book "The Behavior of Organisms" [Ski38], where reinforcement makes any event that increases the likelihood of certain behavior. The theory of operant conditioning uses stimulus that reinforces and the ones that punish certain behavior in order to modify it, and elicit the wanted one. Positive reinforcement presents a rewarding stimulus after a desired behavior is shown, increasing the likelihood of repeating such behavior. Negative reinforcement, or punishment, involves some negative stimulus or removing the positive one in order to decrease the likelihood that such bad behavior happens [Ski53].

Skinner is the founder of the experimental apparatus, the Skinner box, or operant conditioning chamber [Ski38]. Besides operant, this chamber helps studying classical conditioning as well [CMB09, Kre83]. This experimental apparatus studies animal behavior, usually a rat, placed in a controller environment in which it displays specific behaviors [Ski38]. Rat usually presses a lever in order to get a reward, often times coming as food [FS57]. A rat or pigeon is placed in a box with a lever or lights, with the goal to press it if it wants to get a reward in the form of food or avoid something that irritates it, such as loud noise. The experiments show the reinforcement schedules, them being variable, fixed, ratio and interval, and how they impact the rate of learned behaviors [FS57]. This work and Skinner's overall contribution to operant conditioning has a significant impact on educational practices [Ski68].

2.1.5 Applications and Implications of Behaviorist Principles on Motor Skill Acquisition

Behaviorist principles focus on environmental influences and on the behavior that can be observed. Such approach has brought many different applications and solutions in different fields, including education where positive reinforcement is practiced in order to encourage desired behavior, such as teachers using reward system such as stamps in order to motivate good academic performance.

When speaking of motor skill acquisition, behaviorist approach can be very beneficial for designing training and learning processes. Behaviorists offer the basis of learning a motor skill, repetition, reinforcement. This includes feedback which is the crucial aspect of learning a new motor skill, acting as a reinforcement and a guide in the learning process, guiding the individual towards the correct performance and error correction.

In the following text we will focus on the use case of this thesis and how each of the mentioned behaviorist theories are related and could be applied to it. We will use the use case in order to explain how reinforcement and feedback are implemented in the context of behaviorist theories. In the process of learning a hand choreography using gamificiation, the scenario is as follows: showing the choreography steps in the bottom of the screen while the person is looking at themselves on the computer screen trying to replicate the images, receiving feedback in the form of animations. Animations for positive reinforcement are fireworks and for the negative one is a bouncing red letter "X". In this example the images indicating dance moves as cues are conditioned stimulus, and the immediate feedback serving as reinforcement is unconditioned stimulus. This depicts operant conditioning, where correct performance of the dance move, in this context behavior, is reinforced by positive feedback in the form of animated fireworks, representing the reward, increasing the likelihood of the desired behavior being repeated.

An interesting point to make is on the topic of scheduling of the reinforcement. As in Skinner box, different reinforcement timings can influence the behavior, similarly varying, different, reinforcement timings influence learning outcome of dance choreography. If beginners are presented with continuous feedback they could benefit from it, while the more experience dancers can benefit more from sporadic feedback.

When turning to Thorndike and his experiment with cats, we can also apply his discoveries to learning the dance choreography. Learning a new choreography is a process of trial-and-error. Each attempt that leads to correct performance of one move is then reinforced by a positive feedback, increasing the likelihood of it being repeated.

When providing immediate feedback on the quality and precision of performance of each move, instructors can shape dancers' behavior over time. This relies on the theory of Law of Effect, where positive outcomes reinforce the desired behavior [KC09].

Going further, we will talk about Pavlovian conditioning in dance choreography. In classical conditioning, the music or beat in the dance routine can serve as conditioned stimulus. Dancer's movements which represent unconditioned response, can be prompted by verbal instructions, or visually by demonstrations, which both represent unconditioned stimulus. Over time, the dancers learn to perform the movements, conditioned response, in response to the conditioned stimulus which is in this case music, alone [BPS10].

Continuing on the topic of auditory and visual cues, dancers can learn to associate specific auditory cues, parts of the song the choreography is performed to, certain movements through repetition and practice. This is based on the same principle as Pavlov's dogs learned to associate the sound of a bell with food. This associative learning helps in memorizing and executing dance routines [Ada71].

We will now discuss how emotional conditioning plays a roll in motor skill learning. If dancers are presented with a positive reinforcement in the form of positive emotional response, such as applaud or a praise, when performing successfully, they are conditioned to experience positive emotions. With such emotional conditioning their desire to perform well is being reinforced [SL11]. This approach can also help when wanting to reduce performance anxiety. Classical conditioning can help dancers manage performance anxiety, by practicing in a supportive environment. This happens when their negative experience riddled with anxiety, are overwritten with context of support and positive emotional reinforcement, where performance can be associated with calmness and confidence [WP01].

Behaviorism has significantly influenced the field of psychology by providing a framework for understanding how behavior is learned and modified. Through the foundational work of scholars like Watson, Pavlov, Skinner, and Thorndike, behaviorism has established key concepts such as classical conditioning, operant conditioning, and the Law of Effect. These principles have practical applications in education, therapy, and skill acquisition, highlighting the importance of environmental influences and reinforcement in shaping behavior. Behaviorism contributions remain integral to our understanding of learning and behavior.

2.1.6 Social Learning Theory

Building upon, but also opposing in some aspects to behaviorism theory, social leaning theory gives additional insight into how people learn, specifically in the context of motor skill acquisition. Social learning theory states that humans learn through observation of other humans and imitation. In more detail, learning, as a cognitive process, is taking place in a social context, and learning can happen through observation or direct instruction, alone, even without motor reproduction or reinforcement [BW63]. Additionally, learning happens not only when people observe others' behavior, but also rewards and punishments. If an action is perceived to be rewarded higher the chance for it to persist (building upon the behaviorist theory that the behavior is based and lead by reinforcement [Ban71, RCM12]. Likewise, if the behavior was punished, it most likely will not persist [RCM12].

The origins of this theory trace back to 1940 with Skinner, who proposed using stimulus-response theories to explain how language is learned and used [Ski57]. He argued that all verbal behavior is shaped through operant conditioning. Although a behaviorist, Skinner laid the groundwork for social learning theory, suggesting that parts of speech come from words and sounds people have previously heard. He further noted that parents encourage these "echoic responses", linking them to comprehensible speech [Ski57].

We will focus now on Hull's Drive Theory and Social Learning by Miller and Dollard. Hull introduces a natural, inherent, drive for imitation in humans, which suits as a learning mechanism. Clark Leonard Hull was opposing behaviorist stimulus-response theories [Hul30]. Miller and Dollard built upon Hull's theory of drive, which states that a drive is a need that prompts a behavioral response. In particular, a drive for imitation was created and propagated as a result of positive reinforcement from social interactions [Hul30]. This was the first time term 'social learning' was used. Building on this theory of imitation, Albert Bandura brings new contributions.
Albert Bandura, writing a book in 1977, under the name "Social Learning Theory" [Ban77]. He integrated social learning theory with cognitive learning theories, outlining in 1963 [BW63], that the theory is behavioral in its core, emphasizing that learning occurs through the interaction of behavior, environment, and cognition. What was novel about this is that it introduced the concept and highlighted the role of imitation, with it's famous Bobo doll experiments, however the book and the revision of his theory in 1977 changed its course and focus to the more cognitive nature. Key pillars of Bandura's social learning theory are that:

1. Learning is a cognitive process which is placed in a social context,
2. Vicarious reinforcement of observing consequences as well as observing behavior can help learning,
3. Modeling, also known as observational learning.

What his experiments with Bobo doll showed was that humans, especially children learn behavior through observation and imitation of adults', especially if the behavior is reinforced.

Important and related concept to Bandura's imitation theory is the theory of mirror neurons. This theory, first introduced in 1990s, by Giacomo Rizzolatti [RFGF96], a neuroscientist, provides the neurological foundation of observational learning and imitation. The mirror neurons are specialized brain neurons which are activated when an individual performs or is observing someone else performing the same action [RFGF96]. This further reaffirms Bandura's theory of learning by observing and imitating the behavior of other people [Ban77].

Once the mirror neuron is activated the observed behavior is internalized, making it easier to replicate it later [GG98]. The example for this can be the hand choreography used in this study. While learning the hand choreography, the participants are observing the instructor's movements in the video, which activates the same neural pathways that would be activated if the participant really performed those movements, preparing and making the brain ready for imitation [RC04]. This, besides Bandura's theory, supports and explains why imitation and reinforcement are effective in learning processes [Ban77].

Additionally, mirror neurons are regarded as main actors in facilitating empathy and understanding the intentions of others, reinforcing the social context of learning [RC04]. This ties into Bandura's framework by emphasizing that learning is not purely a behavioral process but one deeply rooted in both cognitive and social interactions [BW63, Ban77].

This can be directly tied to learning a new motor skill, especially a hand choreography. These theories show that imitation of observed behavior, such as hand movements, are driven by an intrinsic need, reinforced through social interaction. Reinforcement of correct dance moves or any other motor skill action, in a broader context, could be any positive social feedback such as praise in a dance class or a praise by an instructor, and in a more narrow context of this work it is a positive visual feedback. When it comes to imitation, learning a motor skill, more specifically a hand choreography, is nothing else but imitating another person performing the skill. While imitating, the central point, main goal of the skill acquisition is getting the right, positive, visual feedback, or a praise. Furthermore, positive or negative visual feedback is a social construct and the level of success as well. This becomes crucial in the process of learning where the person is improving the skill, as the feedback not only helps identify errors but also reinforces correct behavior, motivating further effort and refinement. The perceived quality of feedback shapes the learner's self-evaluation and progress, influencing both the rate and the outcome of skill acquisition. In the study performed for the sake of this thesis, the participants were looking at a video of another person, and afterwards at the screen where they saw themselves in the camera, both positive and negative reinforcement feedback was shown.

2.2 Theoretical Perspectives on Enjoyment, Effort, and Feedback in Motor Skill Learning

While designing an effective learning system which learners would easily and willingly opt to use often, it is crucial to understand the relationship between enjoyment and effort and how it translates to feedback in learning. The following review of theories offers an understanding of key theories that are relevant to different feedback modes taking into account enjoyment and effort. The theories that the fundamental overview will cover in order to give a frame for the research being done are the Flow Theory, Cognitive Load Theory, Learning Curve Theory, and Fitts and Posner's stages of motor learning.

2.2.1 Flow Theory: Sustaining Engagement Through Optimal Challenge

In order for someone who is learning a skill to fill engaged, immersed in a task but also to enjoy it, the challenge that is presented to them has to be carefully designed, so that it is not underwhelming or overwhelming. The Flow theory [Csi90] explains how to achieve the optimal level of the learning challenge. Flow occurs when there is a balance between the perceived challenge of the task and the individual's skill level. This balance prevents the learner from becoming either bored (when the challenge is too low) or anxious (when the challenge is too high). Csikszentmihalyi [Csi90] identified several key dimensions of flow, including:

1. Challenge-skill balance: Tasks must be designed to match the learner's current skill level.
2. Clear goals: Participants need a clear understanding of what they are trying to achieve.
3. Immediate feedback: Feedback must help learners adjust their actions to maintain focus and alignment with their goals.
4. Concentration on the task: The activity should demand full attention, reducing the likelihood of distractions.
5. Sense of control: Learners should feel that they are in control of their progress, even if the task is challenging.

Through empirical studies, in many different settings the Flow Theory has been proven to work. Examples supporting this include, studies in music education and sports psychology which show that students are more interested and enjoy learning when the tools are made to fit their skill levels and they get feedback on time [JC99].
Feedback modalities, such as the visual one, in the case of this work, the video-only mode of learning, as any other mode used in this thesis, has to balance the presentation of the task and what type and amount of information represents to the learner, or participant

in the study in order to start and maintain the learning flow. In this case, video-only mode of learning needs to have the right tempo of the choreography presentation, not too slow or fast, the image needs to be clear, the transitions should be clear so that there is no ambiguity, the music needs to be on adequate volume. When adding all these components the task and the learning mode need to maintain the focus. The mentioned video-only mode of learning have a risk of not being engaging enough or not providing additional cues for easier skill retention, the task and skill itself can become tedious and boring. On the other hand, the multimodal systems, e.g. the combination of video with audio instructions, or even a game can be overwhelming with information and require too much attention and effort where the retention drops, due to overbearing the person with processing of video and audio instructions simultaneously, disrupting the learning flow. In the case of learning through a game, the elements such as rewards or punishment even though they can foster engagement, they can also, if not designed carefully, represent a distraction while the learners need to actively process task requirements and remember them.

This theory is important and relevant for this study since this work explores how different modes of learning affect the learning curve, or the learning flow. What this study does is that it exactly explores what type of learning offers the greatest engagement and the highest levels of perceived enjoyment.

Cognitive load theory and Flow theory work together because staying in flow needs good use of the person's mental resources. When learners are overloaded with unnecessary information or jobs that are too hard for them, their cognitive load goes over the limit for staying in a flow state.

2.2.2 Cognitive Load Theory: Feedback Design for Learning Efficiency

An important theory for this study is the Sweller's Cognitive Load Theory [Swe88], which emphasizes how important it is to optimize the demands for best learning outcomes, meaning that the task demand should not be too high or too low. It focuses on the person's capacity to process and their work memory, as well as how it influences their learning efficiency. The working memory is limited in duration and capacity. This fact is crucial when designing a task and instructions in order not to cognitively overload the learner. Cognitive load theory recognizes three different types of cognitive load:

1. Intrinsic load: This refers to the inherent complexity of the material being learned. For example, complex choreography with rapid transitions inherently imposes a high intrinsic load, especially for novices.

2. Extraneous load: This load comes from poorly designed instructional materials that add unnecessary cognitive demands. Examples include disorganized visuals, unclear instructions, or irrelevant distractions.

3. Germane load: This refers to the cognitive effort devoted to schema formation and automation, which is critical for skill acquisition.

When focusing on the work done in this thesis there are a few things that need to be addressed while designing different learning modes. It is important to minimize the extraneous load and optimize the germane load in order to maximize the learning. For this example the motor skill learning through video only must make sure that the video is of high quality, that the visual cues are clear, transitions between movements are seamless, and pacing aligns with the learners' capabilities, e.g. in the beginning the pace must be slow. If there are abrupt transitions or mismatch in the music and the image there can be an increase in the extraneous load.

The mutlimodal learning modes in this study add another layer or complexity and require additional attention. According to Mayer's Cognitive Theory of Multimedia Learning [May09], integrating multiple learning modalities such as video and audio, has the potential to increase learning under the condition that they complement each other. However, if there is any redundancy or conflicting inputs, the attention can be split and the extraneous load increased, reducing learning efficiency. In the context of this study, if the audio instructions would not follow the image represented in the video, e.g. the audio would have the output "ring finger down" and the image in the video would be 2 seconds late or show pointing finger moving downwards, this would evoke split attention, confusion and extraneous load in processing the information, remembering and deciding which finger to move. In the context of the gamified learning, if the game offers unnecessary rewards or penalties which are taking up a lot of space on the screen or sound could also lead to distracting the learners, redirecting cognitive resources from the point of the task. Additionally, if the feedback is presented in a timely manner through rewards and dynamic feedback, and by presenting the information in smaller, digestible chunks such as small pictograms of current moves to be performed, the cognitive fatigue can be alleviated. However, it is important to understand when to introduce gamification so that it does not overcomplicate the task.

There are empirical studies that support this theory principles. One of them are studies on multimedia learning environments that show improved retention and comprehension in learners by reducing cognitive overload if they are carefully and properly designed [Swe88]. These findings emphasize the importance of aligning the task complexity with the instructional design with the person's cognitive capacities, and are directly tied to the feedback modalities researched in this thesis. Additionally, by measuring perceived effort, it evaluates whether multimodal or gamified feedback increases cognitive efficiency compared to video-only feedback, particularly during early-stage learning.

2.2.3 Fitts and Posner's Stages of Motor Learning: Feedback for Different Learning Stages

For understanding the process of learning a motor skill, it is crucial to understand the Fitts and Posner's proposed a thee-stage model of motor learning [FP67]. This model provides a developmental framework for understanding how individuals acquire and refine motor skills over time. Each stage represents a phase in the learning process and has its own characteristics and needs for instruction. The stages are following:

1. Cognitive stage
2. Associative stage
3. Autonomous stage

Cognitive stage is the initial stage of learning, where the individuals focus firstly on understanding the task and secondly on developing a basic strategy for execution. Their performance is inconsistent since they are still experimenting with different approaches and are prone to making errors quite often. The provided feedback is critical in this stage of learning to help the learners recognize and correct mistakes. Again, moving the focus for examples on this thesis, in the learning mode where the video is provided, clear and explicit visual instructions are essential to guide the individuals through the exploratory base. Images only could not give them enough detail to understand how to perform transitions, and if the pace is too fast they would not be able to understand and see them so quickly and easily as if the video was clear and slow. However, having auditory instructions of gamified learning can help in detecting errors in performance and improve timing.

Associative stage is reached through practice, the movements are more consistent and refined. In this stage the errors decrease and the learners start developing a sense of rhythm and timing. In this stage multimodal feedback can be very effective. Auditory cues, instructions, can reinforce visual instructions and help individuals to synchronize the movements with music. At the associative stage the greatest benefit is found in subtle refinements such as more precise timing.

In the final, autonomous stage, the skill becomes automatic for the individual, requiring minimal conscious effort, it is then in the muscle memory. The task is performed seamlessly, making the space for cognitive resources to adapt to new scenarios. In this stage even though the feedback and its type are less important, the motivational element such as rewards or praise, offered in the gamified mode of learning, can support the individuals in continuing practicing and refining the skills.

Fitts and Posner's model is backed up by a study in neuroscience which shows how neural plasticity helps people in learning a new motor skill. Prefrontal cortex is more active during the first, cognitive stage of motor learning which is a sign that the learning is a conscious process. As an individual translates from associative to autonomous stage, their brain activity changes as well, to the areas that are connected to movement. This is when their movements become automatic [KHX+19]. This shows that the Fitts and Ponser's model is founded in biology and can be a valid base for understanding the motor skill learning.

The Fitts and Posner model stresses that learning is always changing, and that the type and amount of feedback must change as the student does. This is closely related to the research done in this thesis for understanding when a video, auditory instructions with video or game come best in place in the learning process. This helps create learning systems that are tailored to each stage.

The findings of this thesis can have valuable and practical insights for designing learning systems that are optimized for early stages of motor skill acquisition. By integrating video-only, multimodal, and gamified feedback, this study provides actionable insights into how these modalities can be used to improve learning outcomes in contexts ranging from education to sports and rehabilitation.

2.3 Learning Curve Theory

Learning curve was first introduced by Hermann Ebbinghaus in 1885 [Ebb13], referring to the rate of learning represented in a diagram. This curve can oscillate and even decrease which was proven in his memory tests published in the same year [HTD+03][Ebb13]. Better understanding of the learning process can be further expedited and made more motivating and interesting to the learners. In previous work, the focus has predominantly been on the reward systems [VLDD22, VDD+21]. However, the interplay between reward mechanisms and various feedback modalities remains unexplored. The sustained interest and ongoing research in this area highlight its significance and relevance.

Ebbinghaus described learning curves as a representation of skill acquisition over time. Video-only feedback may produce slower initial learning as individuals must decode movements independently, relying only on visual cues. Multimodal feedback, by integrating auditory instructions, accelerates early-stage learning by enhancing learners' understanding of spatial and temporal aspects. Gamified feedback, with its engaging reinforcement mechanisms, may sustain learners' motivation, helping them overcome difficult challenges.

By comparing learning curves of these different feedback modalities, the study contributes to empirical evidence to how different modes of learning influence the rate of motor skill acquisition.

2.4 Motor Skill Acquisition

It is important to restate the significance of motor skill acquisition. It is a critical field of study in fields of such great importance to the society and individuals including sports science, education, rehabilitation, physical health and many other. It involves learning voluntary movements through practice, requiring a combination of sensory feedback, cognitive engagement, and physical execution [Mag07]. Motor learning is a lifelong process, including new skills as well as improving the ones already developed. It includes baby's first steps to an athlete's championship-winning performance [FP67, KHX$^+$19]. This process requires a great intrinsic motivation and dedication, with continued practice and effort put in [RD00b, ALDK07].

The purpose of this literature overview is to explore comparative impact that different learning modes have on motor learning including visual, multimodal and gamified mode of learning. Furthermore, it identifies the gaps and provides understanding of the relationships and impact of different learning methods in the case of hand choreography motor skill.

2.4.1 Comparing Feedback Methods: Audio, Visual, and Haptic

Multimodal feedback can be any combination of feedback including e.g. visual, auditory, haptic and more. Multimodal feedback should use and enhance the strengths of all modalities included in order to enhance learning. Type of feedback mechanism or the learning mode has an important role in motor learning, especially on cognitive and associative stages, according to Fitts and Posner. Visual feedback in the form of a video has long been recognized for its ability to provide a clear demonstration of the instructed movements, helping individuals understand spatial and kinematic specifics [PVH$^+$18]. In the cognitive stage, where learners are focused on "what" to perform, video feedback helps learning through observing by breaking down complex movements and patterns into manageable steps.

Auditory feedback complements the visual one very well since it provides the temporal cues, such as rhythm and timing for then to demonstrate a certain move or how. Work of Holland et al. [HBDH10] showed that the best learning outcomes while learning how to play drums gave the combination of audio and haptic feedback in comparison with audio-only and visual-only feedback, highlighting the enhanced temporal and tactile coordination facilitated by haptic feedback. This work showed that multimodal feedback yields best results when combining multiple channels of feedback during learning. Furthermore, Moinuddin et al. [MGS21] in their work on augmented motor rehabilitation and skill training proved that in both healthy and clinical individuals the combination of audio and visual feedback leads to better performance outcomes. Their study showed that dual-modal feedback significantly improved motor accuracy and reduced task completion time, emphasizing that spatial and temporal cues complement each other and how significant this mode is for motor skill improvement. Additionally, in a review article by Sigrist et al. [SRRW13] the role of augmented feedback modalities—visual, auditory, haptic, and multimodal are investigated in the motor skill learning process, and concluded that multimodal feedback including visual and auditory feedback can yield the best results if the system is designed carefully and purposefully. In a paper by Carmen et al. [PMMK24] it was found that the best correction feedback in sports is the multimodal one, including both visual and auditory feedback, as well as haptic. In this case visual and auditory information in the cognitive stage help individuals understand "what" to perform, while in the associative stage, it supports "how" the movements should be refined. Moinuddin et al. [MGS21] found that multimodal feedback is particularly effective for reducing error rates and improving coordination, critical in tasks requiring both spatial and temporal accuracy. The study as the use case had surgeries, where precision and timing are crucial. By combining visual and auditory cues, the participants where able to synchronize movements and correct the errors in real-time.

Although haptic feedback has shown promise in specific contexts, such as learning to play musical instruments, as in work by Starner et al. [HSD+10] where they explored the use of haptic gloves in learning fine motor skills, observing significant improvements in tactile sensitivity and movement coordination, it is not applied as much due to practical constraints of wearing complex devices that can be overwhelming or distract from the task and the process. This study, therefore, focuses on audio and visual feedback, as these modalities are sufficient for effective learning during the early stages of motor skill acquisition. In the case of this thesis we will consider multimodal feedback the one combining visual and auditory ones.

As the learning moves to the associative stage, the multimodal feedback becomes less demanding, making less of a load and more of a benefit, resulting in movement refinement [NL12].

However, poorly designed multimodal feedback can lead to cognitive overload, particularly for novices in the cognitive stage [SRRW13], hindering the learning process. For this reason it is important to stress the role of quality design of learning systems. The only purpose of the system is to help in the learning process, otherwise, poorly designed learning systems can make the learning process less enjoyable and less productive, loosing its purpose.

2.4.2 Gamified Learning

Gamification introduces game elements, such as goals, challenges, and rewards, into learning process and systems. Gamified learning helped keep people interested and motivated [BGRS23], but it's not as clear how well it works in the cognitive stage of movement learning. Van der Kooij [vdKvDvV+19] found that gamified balance and gait exercises significantly increased enjoyment, which is a critical motivator to start and keep learning. Nevertheless, the same research noted that gamification's benefits are often motivational rather than performance-based, aligning with findings that gamification may not directly enhance skill acquisition in the early learning process.

Some research suggests that incorporating games into learning may cause distractions or mental stress if done too early on. Learners in the cognitive stage are still building basic skills, so they might find it hard to interact with game-like features in a meaningful way. Instead, gamification might work better when it's presented during the associative stage, when students are ready to improve their skills and get more out of long-term engagement [TA16]. In the same way, Almeida et al. [AKUF23] found that those who are learning the skill for the first time and had early gamification inclusion in the learning process, made mistakes more often because they didn't have enough basic knowledge.

Combining gamification with multimodal feedback offers a possibility to address both cognitive and motivational aspects of motor learning. However, as highlighted by Wiemeyer and Schneider [WS12], the effectiveness of gamification depends heavily on the how familiar is the person with the skill and the design of the gamified elements. Beginners, especially those in the cognitive stage, may struggle to adapt to gamified systems without sufficient foundational knowledge.

Van der Kooij et al. [vdKvDvV+19] emphasized that gamification's motivational benefits could complement traditional feedback methods, particularly in later stages of learning. By phasing gamification after an initial period of structured instruction, learners can build foundational skills before engaging with more complex, game-based interactions.

2.4.3 Research Gaps and Contribution

While existing literature highlights the effectiveness of multimodal and gamified feedback, several gaps remain. First, there is limited research of gamified learning during the cognitive stage of motor learning. Second, the comparative analysis of different modes of feedback and their timing of inclusion in the learning process remains underexplored, as well as the relationship of enjoyment and effort in the learning in the same analysis. Connecting how the enjoyment and effort relate to the learning curve and comparing it against different learning modalities is the gap that this thesis is addressing through systematic comparison. By focusing on the timing and context of gamification, but also modes as well, this study aims to understand the design of effective learning systems for early-stage motor skill acquisition.

CHAPTER 3

Application for Motor Skill Learning

3.1 Application Description

For the purpose of making a consistent learning experience for different users a tool was needed to administer different modes of learning, from administering video for visual learning, to audio instructions to gamified learning through playing a game. Additionally, in this application testing mode was implemented as well.

The application is a web based application, available on a localhost, which takes in the user's ID, name and the mode the user is using to learn with. There are four modes of application, as follows:

- VIDEO
- AUDIO
- GAMIFICATION
- TEST

Each mode has normal speed, and the slow speed which is twice as slow as the normal version(0.5x speed). We will now cover each mode, and explain why each was implemented.

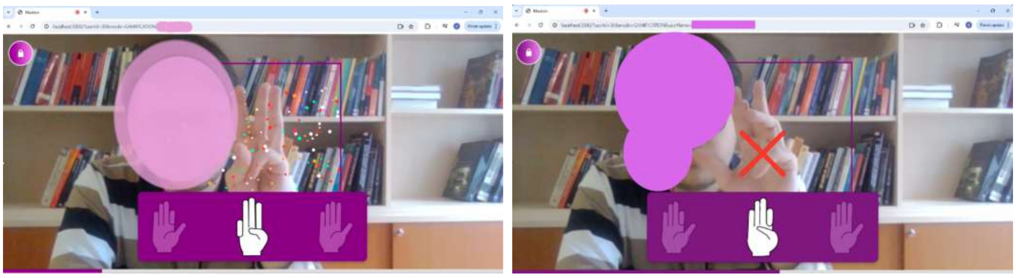
VIDEO mode offers the user ability to see the video of the motor skill, more specifically the hand choreography, however, without the ability to pause the video. This was chosen so that all users have the same learning experience, and that the process is consistent. On the screen only the video is shown when the page is loaded. There is a play button on the image of the video, once clicked, the video starts. The goal was for the user to follow the choreography along, or choose to watch, whatever the their choice is. Once the video stops, the user can replay it. The video has the music that follows the along, and is 30 seconds long. There is an additional slow mode, where the video lasts 60 seconds, however everything else remains the same.

AUDIO mode does not differ too much from the VIDEO mode. The only difference is that in the same video with music audio instructions are recorded and placed over music, following the steps from the video. This makes AUDIO a multimodal mode, including visual and auditory mode of learning. The slow mode is exactly the same, with the slowed music and video, but the audio instructions are following the video with bigger gaps, pauses between the steps, however in the same, normal speed of audio, voice sounding normal. The instructions placed over the video are written, scripted by the author, following the video, and the sound. The voice in the video is generated by a video editing software, Clip Champ [Des24], for the purpose of voice and accent familiarity. AI generated voice is a synthetic voice generated by artificial intelligence algorithms that are able to imitate human speech, trained on big textual and audio datasets.

TEST mode includes live testing, where the music from the video is played and the user, participant, can see themselves in the screen in camera. Test mode doesn't include feedback, points or mistakes, user only see themselves in the screen. There are two modes, slow and normal speed, with accompanying music, depending on the speed of learning on that day. Frames from the live recording are extracted and processed using Google's algorithm MediaPipe, where keypoints of the hand are detected and extracted, angles between them calculated and sent to backend where the values are being compared to predetermined values, and points or mistakes assigned and saved in the database.

GAMIFICATION mode as the main inspiration for this application includes the live evaluation of the motor skill performance, as in the TEST mode, however the points are not being counted, but feedback is shown. If the user, participant, performs the motor skill correctly at the right time they get an animation of firework on the screen over the live camera, and if not, a big red letter "X". In the bottom of the screen there is a line, track, with three icons that are changing along with the music. All three icons are presenting one step, one move, of the hand choreography, one position of the hand and fingers. Farthest left icon represents the previous step, the one in the middle is the current step, position that the user should perform, and the farthest right is the one that is coming after the current one so that the user can prepare. The evaluation of steps and feedback is processed live, in the same manner as in the TEST mode, using MediaPipe.

While using this application, as previously mentioned, the interface is very simple. There is only one button to click on, in order to start the learning session. While switching between learning modes, starting a test, changing the speed or userID, the address of the localhost is changed, there is no user interface or buttons, in order for the participants to be as focused as possible on the task, as well as preserving the possibility of changing the named properties only by the researcher when needed. For this reason the flow chart will not be designed specifically since the user interface for the participants themselves is consisted of only one button to start the session.



(a) Correct Performance (b) Wrong Performance

Figure 3.1: Game showing correct and wrong performance with anonymized participant

3.2 Goal of the Application

The primary goal of this application is to expedite the motor skill learning and elevate the quality of motor skill acquisition. In this case, more specifically, the Hand Mudra choreography as the motor skill. More specifically, this application provides the opportunity to compare different modes of learning and their impact on motor skill acquisition. Furthermore, and most importantly, it provides the possibility to test the performance of the motor skill real time, and quantify the progress of the learning process for analysis and empirical study.

As previously mentioned, there are three different modes of learning. We will now cover the goal of the first two, VIDEO and AUDIO, following with GAMIFICATION where we will go in depth of the mode's goal. The application has the goal of unifying different ways of learning, representing a learning tool offered during the process. The reason why the application offers all three, even though VIDEO and AUDIO are quite simple, is for the consistency purposes and ease of use. This application, therefore, can be expanded if some other ways of learning would be explored. Most importantly, it is a tool for learning, offering consistency and familiarity along different modes of learning. Even though VIDEO and AUDIO modes are fairly simple in the term of application options and implementation, there are there to offer unimodal feedback, that being VIDEO, and multimodal feedback, being AUDIO. These ways of learning have been studies before, and are found to be a good benchmark practice for comparing the new or other modes of learning. Hence, AUDIO and VIDEO modes are there to provide ease of use and consistency throughout the experiment, a base for further tool development and experiments and for ease of comparing different modes, especially GAMIFICATION in this case, providing a standardized protocol in this study.

The gamified learning part of application, representing the main motive and inspiration for developing the tool itself, has the goal of enhancing the speed and quality of motor skill learning, more specifically Hand Mudra choreography. Since there were no previous tools or applications made for this purpose, the game for learning this motor skill was developed, inspired by the existing game “Just Dance Now” [Ubi24], which is a popular consumer console game, for replicating different choreographic. Since the motor skill in this case is also a choreography it is useful to measure performance and progress of the participants through checking if the moves were correct, and quantifying it. This was done by incorporating interactive and visually engaging elements within a web-based platform. As opposed to the first two mentioned modes of learning, VIDEO and AUDIO, also regarded as more traditional modes, since the instructions rely on passive and linear content, this approach provides active learning experience. This is done by showing icons that indicate previous, current, and next positions, in the bottom of the live camera feed, enabling the users to see themselves in real-time as they learn, practice. This setup is designed not only to improve motor skill acquisition but also to transform the learning process into an engaging, enjoyable experience with reduced perceived effort.

Besides engaging new approach to learning, gamified learning has another critical objective. That is, to determine if it can make the process more enjoyable and reduce the perceived effort, as opposed to the traditional approaches. This is due to traditional learning methods requiring prolonged focus and concentration, without elements that boost enjoyment, which can feel tiring and strenuous, especially while learning precise and sequential movements. This application therefore introduces gamification elements, including real-time feedback and progression cues, with the foal to create a sense of achievement and engagement. By doing so, the application aims to:

- Increase intrinsic motivation by making the learning process feel like a rewarding activity rather than a task. The visual progress indicators and dynamic feedback encourage users to remain committed and view the experience as enjoyable and attainable.
- Reduce cognitive load and perceived effort by presenting clear, sequential guidance, allowing users to focus only on the immediate actions they need to perform rather than processing complex, whole-task instructions. The step-by-step visual guidance has a goal to lower mental strain and make the experience easier to comprehend and more manageable, especially for beginners.

Furthermore, besides testing which approach yields the best results from gamified, multimodal or unimodal, it also tests whether a gamified, interactive one can accelerate the learning curve for motor skill acquisition, by applying cognitive principles such as anticipatory learning and self-correction. With visual cues, icons, or small images in the bottom of the screen, indicating the sequence of the hand positions, users can develop natural rhythm, which encourages smoother transitions between movements and helps solidify each step in memory. By showing the next step in advance, users can mentally rehearse the transition, promoting faster and more intuitive skill and ability. Additionally, because gamified learning involves active participation, this approach is hypothesized to improve long-term retention compared to passive video-only methods.

Beyond these practical goals, as previously mentioned, the application is also designed as a research tool to examine the broader impacts of gamification in motor skill learning. This is done by providing real-time feedback and engaging sequence indicators in the form of the line of icons indicating the steps of the choreography. This can serve as a viable alternative to traditional approaches, exploring new and innovative skill acquisition techniques. Such new engaging techniques could lead to improvements or offer different methods in educational and therapeutic applications.

3.3 Design Principles and User Experience Considerations

AI generating voice software offer standardized options given to the users, with familiar voices and consistent tones and accents. This consistency and uniformity helps users develop familiarity and trust with the application, supporting better understanding. This is key for reducing cognitive lead connected to needing to adjust to different and varying voices and accents. Furthermore, it minimizes misunderstandings that could occur with having varying accents and languages, especially accents coming from people of different language backgrounds. This fosters clarity of the instructions given in the application to users, and in this particular case bridges the issues of author and the participants having different mother tongues, providing consistent auditory instructions, easing communication of information and commands, fostering comprehension.

Consistency and familiarity are guiding principles and strongly supported in the design of this application. Both of these items are key for technology acceptance and adoption, since it is more likely for users to use and interact with systems, technologies, that are consistent in all interactions. "Consistency is classified into behavioral consistency (e.g., operational consistency) and object-based consistency, which includes consistency in information, system and service in terms of the information systems success model." [SSW14]. Behavioral consistency refers to how functions are consistently executed, and object-based consistency refers to uniformity, consistency in the system's information, interface and services. In this application, behavioral consistency is supported by structuring each learning session in the same way. Furthermore, each video starts with the same image, standardized starting image, and the entire sequence of each learning round and among different modes, maintains the same format, reducing potential distractions or any type of confusion for participants, including the previously talked about audio, voice feedback. Object-oriented consistency refers to having each video occupying the same size, position on the scree, as well as all UI elements remaining static across sessions and modes. Furthermore, this approach ensures that the focus of the participants remains on adopting the new motor skill, supporting the learning experiment to be legitimate and comparable between all participants and modes. Continuing on the note of AU generated voice, consistency and familiarity are strengthened by using the same voice in every instruction. This design decision supports consistency and uniform auditory experience, which is crucial for user comfort and reduction of redundant cognitive and attention load, enabling participants to focus on learning. This is especially important in the multi session studies, particularly those over multiple days, which in this case, is a two-day study. This approach ensures familiarity and consistency from day one to day two, supporting user comfort and engagement. When users encounter the same voice and content across sessions, they are more likely to feel more confident and comfortable with the learning tool and the experience itself, improving both the quality of their participation and the validity of the study, throughout the entire learning experience.

The accent on consistency and familiarity contributes to the technology adoption. Technology adoption is more than simple use of the system, it reflects users' readiness and willingness to engage with the new technology. When providing consistent and familiar learning environment, the application encourages a positive user experience, making participants more comfortable and willing to engage in each session. This readiness is crucial for effective motor skill acquisition, since a steady and familiar environment enhances focus, reduces excessive cognitive load, and supports continuous engagement across multiple learning sessions.

Overall, the application's consistent use of standardized AI-generated voices and uniform structure promotes user familiarity and builds trust in the tool. This stability is crucial in a learning context where user engagement and willingness to participate can directly impact outcomes. By focusing on these elements, the application not only facilitates smoother technology adoption but also creates an optimal environment for learning and skill acquisition, providing a stable environment for experiment testing, supporting its legitimacy.

3.4 System Architecture

3.4.1 Technology Stack

The following is the overview of the technology stack, what technologies are used, how and why they are chosen. TypeScript [Cor24] is selected for the development of the application. TypeScript is a superset of JavaScript, with extended programming paradigms that allow for an object-oriented style of programming. The TypeScript code is transpiled to JavaScript and served through a Node.js [Ope24b] application that binds to a server port to deliver static files using Express.js [Ope24a] serving the main logic of the game that is executed by the browser. Additionally a web-socket[Hic24] server is mounted to allow real-time data communication during the game, such as to collect points. A more detailed description of the software architecture as well as an game activity flow is described below. TypeScript enables rapid application development while following best practices for web client development and object-oriented programming paradigm. These two factors play a critical role in the selection of this technology.

3.4.2 Client and Server

The application consists of a server-side and a client-side. The server-side has three main components: the database, a Node.js server using the Express.js framework, and a WebSocket server. The database serves as the storage solution, responsible for tracking and managing game states. The Node.js Express server and WebSocket server handle in-game changes and events, such as collecting points. Complementing the server-side, the client-side consists of the game user interface, a client WebSocket that connects to the server WebSocket, Google's MediaPipe library, and the StateManager, which serves as the core of the client-side logic.

The user interface is built using HTML Canvas. It is responsible for drawing the game layout and dynamically responding to changes based on input tracked by MediaPipe.

MediaPipe tracks and detects fingertips, wrists, and hand joints, enabling the identification of gestures to process the game sequence and award points (e.g., correct or invalid gestures) to the server via the WebSocket connection for real-time processing.

This setup ensures that all game events are stored in the database, with the server game logic processing instructions received from the client's WebSocket connection.

The StateManager also keeps the game state in memory while executing and processing the game loop. This includes performing the gesture sequence to start, play, and finish the game. User-specific and session-specific metadata are tracked through attributes such as gameUUID, gameMode, userId, and userName, ensuring each game instance is uniquely identifiable.

Game setup and initiation configure the UI elements based on the game mode, initializing canvas animations for gamified experiences, and setting up the game. The StateManager sends a signal to start a new game session, begins the game loop, updates gesture indices, monitors progress, and uses the detectors to handle gesture detection.

It updates the game state, plays instructional videos if video mode is set, checks if the game is actively running, and ends the game session while clearing intervals. It also manages the time indicator for gesture stages, updates the UI with the remaining time for the current stage, and displays the next hand gesture to the user.

The StateManager works closely with the GameSocket for server communication, UI-Elements for managing the interface, and detectors for gesture recognition. The gesture detection is handled using detectors, with the Hand-Detector class wrapping the MediaPipe Hand Landmarker API to provide a high-level interface for hand tracking. The detectors module provides utility functions for gesture detection based on hand landmark data. Further details on hand detection will be covered in the next section.

3.4.3 Detectors Used For Gesture Recognition

There are nine detectors responsible for identifying and categorizing gestures based on the response results from MediaPipe's video input processing.

Each detector outputs a true expression if the gesture is validated and meets its defining criteria; otherwise false is evaluated, indicating an invalid gesture.

These detectors use parameters provided by Google's MediaPipe to make their assessments. To achieve this, landmarks are passed to each detector as the primary input for interpreting gestures.

Each landmark describes the position of the hand within the input frame by identifying 21 key points corresponding to the fingers, wrist, tips, and joints.

The parameters are defined for each landmark, as follows:

- x, y, and z points of the landmarks represent the normalized axes:
  - x: Represents the horizontal axis, normalized against the width of the frame.
  - y: Represents the vertical axis, normalized against the height of the frame.
  - z: Represents the relative depth

More can be viewed in Figure 3.2

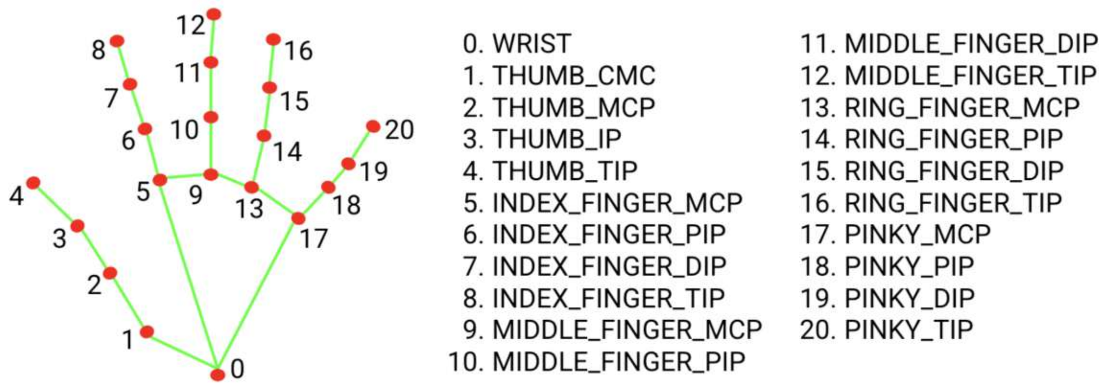


Figure 3.2: Hand Landmarks [Goo24b]

The data is further passed to detectors to categorize gestures.

For practical understanding of detectors, the definition of ONLY_INDEX_FINGER_UP detector is described below, while the others are omitted as they define gestures in a similar manner.

The ONLY_INDEX_FINGER_UP detector identifies when only the index finger is extended, while all other fingers—thumb, middle, ring, and pinky — are pointing down toward the palm, with the palm facing the video input device. This is determined by calculating the distances between the index finger and the palm. The positions of the key points are extracted, specifically the key points representing the tips of each finger. The distances between the index finger and each of the other fingers are then calculated. This process is applied to each gesture. If the position of the fingers matches one of the predefined gestures during the calculations, that gesture is recognized and subsequently processed to trigger a game interaction.

3.4.4 Database

A PostgreSQL database is used to manage game points capturing and user information. A typical game record consists of a generated primary key id, a user id and name, a game mode, and a creation time that marks the start of the game.

For each game session, players collect points for correct or wrong gesture recognition. Each gesture recognition event is defined within a time sequence, using the start and end attributes to indicate the beginning and end of a sequence within the game. Within the Point domain, the type of gesture is recorded, along with points marked as correct or incorrect.

The relationships are defined using a well-situated object-relational mapper (TypeORM), allowing efficient data schema modeling to abstract away data manipulation against the database.

The table below provides a descriptive overview of the database domain with its attributes and their entity relationships.

Table 3.1: Game Entity Attributes

Name	**Type**	**Relationship**	**Description**
Id	uuid	Primary Key, Non-null	Identifies one specific game played by a user.
userId	varchar(255)	None	A non-unique user ID as a string to identify a player within a game.
userName	varchar(255)	None	An optional user name for descriptive naming of a user for adding context to a game.
gameMode	varchar(255)	None	Defines the operational game mode, specifying the mechanisms of the game.
createdAt	Timestamp without time zone	None	Marks the start of a game. Each game has a fixed period of length.
points	Reference Key to Point Entity	Foreign Key	Represents a one-to-many relationship between a game and points

3.5 Live Hand Tracking - MediaPipe

In this application for the purpose of hand recognition, live tracking and pose detection MediaPipe was used. MediaPipe is a framework developed by Google, which is a set of libraries and tools for easy application of artificial intelligence and machine learning techniques in any application [Goo24a]. It can be used across different development platforms, and it is an open source project. Examples of real time hand tracking can be seen in Figure 3.3.

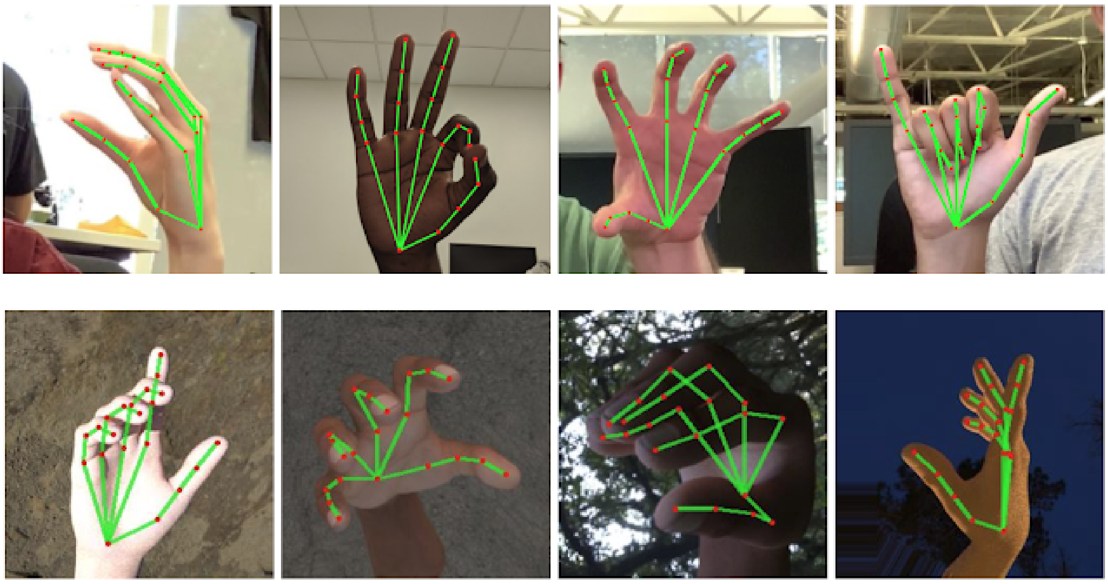


Figure 3.3: Examples Of Real Time Hand Tracking Using MediaPipe

MediaPipe is divided into different functionalities, such as MediaPipe Tasks, which includes cross-platform APIs and libraries, MediaPipe Models which includes pre-trained models. Additionally, there are MediaPipe Model Marker which enables the user to customize the model for custom solutions where user's own data would be used, and last but not least, MediaPipe Studio which enables visualization, evaluation and benchmark solutions in browser [Goo24c]. MediaPipe has plethora of solutions, each including one or more models, but also leaving the room for the user to customize the models. Some of the solutions include object detection, image classification, image segmentation, hand landmark detection, gesture recognition, face detection, pose landmark detection etc. In this application object detection and hand landmark detection are crucial.

Diving deeper into the hand landmark detection, its main task is to detect the landmarks of the hand detected in an image. There are 21 landmarks, compromising of all joints in the hand, the palm and root of the arm, and tips of fingers [ZBV+20]. This task is useful for locating the key points and rendering visual effects based on it. Landmark detection uses image data, and by applying machine learning algorithm model as "static data or a continuous stream" [Goo24b] outputting hand landmarks in the image coordinate, world coordinates and handedness of multiple detected hands [Goo24b]. The hand landmark detector processes input images by performing rotations, resizing, normalization and color space conversion in order to prepare them for the efficient and accurate landmark detection. It calculates the score, filtering results based on a score threshold, ensuring that only results that are meeting certain confidence level get considered [ZBV+20]. The input for the hand landmark detector can be in the form of still images, decoded video frames or live video feeds. The output contains, as previously mentioned, handedness of detected hands, meaning detecting and assigning left or right hand, image and world coordinates of hand landmarks [ZBV+20]. There are several configuration options for optimizing detection based on user requirements including running mode, number of hands, minimum hand detection confidence, minimum hand presence confidence and minimum tracking confidence. Description of each of these configuration options follows [Goo24b].

- Running Mode allows the application, tool to operate in different modes, depending on the input type
  - IMAGE: processing single images
  - VIDEO: processing frames from pre-recorded, loaded video
  - LIVE_STREAM: processing live camera feed
- num_hands: defines the maximum number of hands the detector can, should track, default value is 1
- min_hand_detection_confidence: determines minimum confidence score required for the palm detection model to consider detection successful, default is 0.5
- min_hand_presence_confidence: sets minimum confidence score for hand presence in the landmark model. f the score is below this threshold, the palm detection model is triggered to re-establish hand location. Otherwise, the tool relies on a lightweight hand tracking algorithm for subsequent frames, default is 0.5
- min_tracking_confidence: specifies the confidence threshold for tracking continuity between frames. This ensures smooth transition across frames in video and live stream modes, default value is 0.5

Continuing further to model architecture for both models, palm detection model and hand landmark detection model. Both models are based on convolutional neural networks (CNNs). CNNs are often used for real-time performance for low-power and low-memory devices, such as phones [HZC+17]. In the following text there will be explanation of CNNs, followed by specific CNNs that the palm detection model and hand landmark detection model are built by, respectfully.

A convolutional neural network is a type of deep learning model designed to process data such as images. Hence, they are used and good tools for image recognition and classification, due to their ability to automatically and adaptively learn spatial hierarchies of features from input data [LBBH98]. CNNs are constructed of convolutional layers, activation function, pooling layers and fully connected layers. The data is processed through multiple layers, each layer learning to detect different features. Early layers usually focus on patterns such as edges, while deeper layers identify more complex structures such as shapes or objects [IGC16].

Convolutional Neural Networks, or CNNs, are a subgroup of Artificial Neural Networks, also referred to as ANNs, however they differ from other forms of ANNs. They are analogous to traditional ANNs, However, they differ in a way that "instead of focusing on the entirety of the problem domain, knowledge amount the speciifc type of input is exploited" [IGC16]. The advantage of CNNs is that they can have a far simpler architecture than other ANNs.

Convolutional layers are layers of neural network that apply a set of learnable filters, called kernels, to the input data. These layers perform convolution operations that extract local features such as edges, textures and patterns [IGC16]. Each filter slides over the input data, producing a feature map that highlights a presence of specific features at various spatial locations [LBBH98]. Activation functions are applied after convolution, with the goal to introduce non-linearity into the model, enabling to learn complex patterns. One of the usual activation function used is Rectified Linear Unit, or ReLU activation function [NH10]. Pooling layers reduce the spatial dimensions of feature maps, which decreases computational load and helps in making the detection of features invariant to minor translations and distortions. Usual pooling layer used is max pooling [BPL10]. Fully connected layers are implemented in the final stages of the model architecture implementation, since they take high-level features extracted by previous layers and perform classification tasks. Each neuron in these layers is connected to all neurons in the preceding layer, integrating the learned feature to make predictions [IGC16, LBBH98].

Palm detection model is a CNN-based single-shot detectors, which means that uses a CNN architecture, as previously described, similar to a single-shot detector(SSD) optimized for detecting small, rigid objects, such as palm [ZBV+20]. Single-shot detectors are object detection models that enable the identification and localization of multiple objects in images or video frames in a single pass through the network [Goo24d]. The architecture is inspired by RetinaNet [LGG+17], which incorporates an encoder-decoder structure to capture broader context and better localize small, low-contrast objects, such as hands against different backgrounds. An encoder-decoder structure is a neural network architecture designed for transforming input data into output data through a two-step process, where the encoder compresses the input into a fixed-size representation and the decoder reconstructs the output from that representation [BKC17]. The convolutional neural network model in this case is fine-tuned [HR18] to handle complexities of palm detection, especially issues such as self-occlusion and diverse rotations. Fine-tuning is the process of taking a pre-trained model and adjusting its parameters on a new, related task to improve performance [JYL14].

Moving onto the hand landmark detector model, it is a regression-based CNN. Regression-based convolutional neural networks are used to predict continuous numerical values, such as the x, y, and z coordinates of hand landmarks, rather than discrete class labels. This approach is essential for tasks requiring precise localization, like hand landmark detection [TS14]. Hand landmark detector model works in such way that after the hand region is detected, the landmark model performs direct regression to output the x, y and z coordinates of each of the 21 landmarks on the hand. This CNN model is optimized for efficient inference and uses dense convolutional layers to output precise spatial coordinates to each landmark withing the cropped hand region [ZBV+20]. In short, the palm detection model provides the initial bounding box. while the landmark model predicts precise key points on the hand.

CHAPTER 4

User Study

4.1 Different Modes of Feedback

For this study a new way, mode, of learning was introduced with the aim to test if this innovative approach would have positive and significant impact on the learning process and the speed of learning a new motor skill. This new mode is a gamified approach to learning, offering the users to learn through a game. For the purpose of testing and proving the successfullness of the new approach, traditional approaches are included in this study as well for the purposes of comparative analysis and proof support. These traditional modes include a unimodal way of learning, only through visual feedback, as well as the already scientifically proven more impactful approach, a multimodal way of learning, which besides visual includes auditory mode of learning as well, through spoken instructions on motor skill performance serving as a guidance. A more detailed description of each mode follows.

4.1.1 Visual Learning

Visual learning is a powerful and essential modality in motor skill acquisition, since it is the most natural and inherent way of learning of human and living beings, through seeing and repeating. It is the basis of human ability to process visual information effectively and fast. When learning motor tasks, visual learning through video demonstrations and static images plays a crucial role in enhancing skill comprehension, imitation, and execution.

In order to specify in more detail what type of visual learning this study refers to, here will be talked about only about video and images of the motor skill performance. Visual learning through video demonstrations provide the participants to observe the precise movements and timing required to perform the task, skill. This process supports Social Learning Theory presented in the beginning of this thesis, in the scientific background discussion. One of the main pillars of this theory is the observation and imitation. This process activates and uses the mirror neuron system, which is activated when an individual observes an action performed by someone else, in this case a person in the video or images, which fascilitates imitation and learning of that action [RC04]. Research has demonstrated that observing motor tasks can support creation of neural connections related to those movements, preparing the brain for improved physical execution [Hey11]. More specifically, in the context of this work, motor skill hand choreography, which involves complex sequences of finger and hand positions, benefit from visual cues as they allow learners to accurately see and process transitions and precise positioning of each movement.

Furthermore, having a video as a tool to learn a motor skill the participant has a double benefit, first of spatial dimension in the sense of where to move, and the temporal one indicating when to move. This information the participant connects so they know when to do what and how. Watching a task performed in real-time or slow motion helps the participants understand the rhythm and pacing, coordination required to replicate the motor skill [RBJ+24]. When talking about images as a visual way of learning, they provide a static reference that highlights gestures, making it easier to remember them, helping in the focus.

Additionally, visual learning aids in the formation of motor memory. Motor memory is stored representation of movement in the brain. Studies suggest that combining visual learning with active practice helps remember and embed the motor skills deeply, improving long-term memory retention and performance [EK80].

In summation, the benefits of visual learning are that the learning is accelerated by observing detailed videos, since it is clear for the participant how to perform the motor skill in question [JKG09]. Furthermore, visual content helps break down the complex movements into smaller parts that are easier to process to a learner, and later replicating it with a reduced cognitive load [TSO19]. Last but not least, having continuously the opportunity to see the task enables the learner to refine their movements and improve precision [MWW21].

The visual mode of learning entails a video in the center of the screen in a shape of a square with only the hand in the frame. The video lasts for 30 seconds and contains only the choreography that follows the music. There is a slow version that is used in the first day of learning, and the twice as fast version, the normal paced version on the second day. The participant, learner, can see only the mentioned video on the screen, with no other buttons, graphics or text, so that the participant's focus in purely dedicated to the video. The video contains a real-life video of a person performing the motor skill. Since in this mode there is only visual way of learning, it is regarded as a unimodal way of learning.

4.1.2 Audio Learning

In this study auditory learning is comprised of the same content as video learning case, meaning the video that shows the choreography, lasting 30 seconds, placed in the middle of the screen. However, in audio learning case, when the mentioned video is played, a sound of audio instructions indicating to the learner what to do next is played over the given video. The audio of instructions is embedded in the mentioned video. Same as mentioned before, there are two paces of this case, the slow one and the normal paced one. In the slow mode, the audio instructions are spoken, recorded in a normal speed, however with longer breaks in between them, so that the user has no confusion or discomfort while listening, hence enabling them to focus better. As opposed to the visual learning which is a unimodal way of learning, offering only one mode of learning, the visual one, auditory is multimodal, counting in visual besides the audio instructions.

Auditory feedback provides an additional sensory channel and tool for better understanding, remembering and adopting the skill. When combined with the visual learning the auditory instructions improve the learning process by reinforcing actions, guiding participants through movements, improving engagement and retention. This multisensory approach to learning is shown that it improves information processing and memory retention [SS08]. While having both visual and audio instructions learners benefit from a more detail input, helping and improving their ability to replicate motor tasks accurately. This is further helped since auditory instructions provide critical timing cues which helps participants pace their movements correctly at the right time, and connect them properly. This is critical in a task such as a hand choreography where timing and sequence play a crucial role for executing the movements accurately. Furthermore this approach helps in reducing the cognitive load, since it distributes the information processing among different sensory channels [Art08]. This happens since when a person is receiving two types of cues, through two different channels, they can process information more effectively, without overloading the visual memory. This further helps them to understand and remember the movement patterns better, which is crucial for complex sequences and skills, such as Hand Mudra.

Audio instructions can help ensure that the participants are not only mimicking the movements, but that they are consciously performing them, directly contributing to the focus, hence accuracy and fluidity of each gesture. Additionally, as already mentioned previously, this approach can help the participants manage the pacing, to be synchronized to the rhythm, performed at the right time, with the right speed, which directly contributes to their retention and performance overtime. The main advantages of this approach can be summarized and listed as follows:

- Improved timing and coordination
- Enhanced precision
- Deeper engagement
- Smaller cognitive load
- Greater retention

4.1.3 Gamified Learning

Gamified learning could also be regarded as a part of visual learning, since it relies on the instruction track and images in the bottom, as well as the camera feed posing as a visual element of a mirror for the participant. However since it has its own factors, engagement tactics and requirements of the participants it has its own section, explaining all of the mentioned in detail.
Gamified learning, or gamification of learning, is putting elements of games into the learning process [DDKN11], which in this case includes immediate, real-time feedback of participants' performance in the form of fireworks if the performance was correct, or in a form of big, red letter "X", across the camera feed, as well as indicating the previous, current and next step in the form of images in the bottom of the screen. This approach can enhance the level of engagement, to optimize their skills and learning itself [SRM+20]. Furthermore, in the case of this study it represents an innovative strategy in motor skill acquisition, since it combines real-time feedback and motivational design in the form of the mentioned animated feedback. This specific design is inspired by the popular game Just Dance Now [Ubi24]. The nature of gamification is such that it makes use of the principle of intrinsic and extrinsic motivation [RD00a], which then creates a rewarding environment that supports focus and practice.
So far gamification has shown increase and improvement in the learning process in educational contexts, including increased engagement, retention and intrinsic motivation [DDKN11]. For motor, and also any other learning this approach increases engagement, and through the constant real-time feedback it helps with not just engagement but also correction and hence motor skill acquisition that is faster and more precise.

In this study, the gamified learning is used for teaching a hand choreography, a motor skill that requires precise hand and finger movements, and in the right moment. The gamified learning mode has the following components:

- Camera feed: Participants see themselves in a live feed, enabling immediate self-assessment and needed correction
- Instruction track, positioned at the bottom of the screen with the following three elements:
  - Previous step: The farthest left image showing the past step, or move, that the participant should have made for the reflection and context
  - Current step: Highlighted image that shows the action that the participants should perform real-time
  - Next step: The farthest right image which should prepare the participant for the upcoming movement

This interface offers participants with two dimensions, the temporal one and the spatial learning dimension. This is important for helping the participants in synchronization and sequencing, which are critical in motor skill learning [MA13]. In this mode of learning there are two variant, as in the visual and audio learning versions, which are the slow-paced and the normal-paced one. In the slow version, which is at half the speed from the normal one movements are performed at a reduced speed, with a longer break between them, so that the person can look at the upcoming image and focus on performing it, feeling the new movement. The reduced speed is helping in initial comprehension of the learning material by reducing the cognitive load [Swe88]. The normal version has the normal speed music, with a shorter break between the movements, being introduced on the second day, simulating a real-world pacing, enhancing rhythmic synchronization [PJVT24]. Having choreography sequences shown in the instruction track in the bottom of the screen, synchronized with music, enhances timing and rhythm understanding with participants, which is a key factor in motor learning [EFS+16]. Participants learning with this mode see only the game interface on the laptop screen and no other interaction elements such as buttons or text, ensuring the undivided attention on the task, increasing focus and concentration.

There are a couple of advantages of a gamified learning approach, especially in the context of cognitive and behavioral mechanisms. These advantages are shown in increased intrinsic motivation, where gamified approach and reward systems naturally increase engagement, fostering a state of flow [Csi90], furthermore, it awakes a focused attention in participants since the interactive interface reduces distractions. Last but not least, and very crucial, this approach offers participants an opportunity for iterative refinement, since the real-time feedback enables participants to continuously adjust the movements, supporting incremental improvement [EKTR93]. The research shows that gamified systems are very effective in supporting engagement during repetitive and challenging tasks, which makes them especially sutiable for motor skill learning where repetition is expected [HKS14].

The gamified approach aligns very well with the motor learning theories which are putting a great focus, highlighting active practice, feedback and repetition [Ada71]. This systems engages the mirror neuron network which gets activated during observation and imitation of the movements. This supports neural adaptation and skill acquisition and improvement [RC04]. Additionally, combining anticipatory cues with instant feedback improves motor memory retention and retrieval, both of which are critical for learning complex movement patterns [Kra06]. Additionally, via repeated practice, the dynamic interaction helps participants internalize sequences, promoting procedural learning [SL19] makes use of the principle.

4.2 Visual and Gamified Learning Combined

In the previous three titles basic theory and reasoning behind the choice of learning modes was explained. In the combined mode of learning some of the previous modes are combined in the learning process. For this specific mode, visual and gamified learning combined, more precisely, video and gamified learning combined, is designed as follows:

- Day 1: Visual learning only
- Day 2: Gamified learning only

On the first day, participants are learning in the slow mode, as all other participants, however, with the exact same mode as the Visual Learning. The video showing the choreography in slow mode, with only the video shown in the screen of the application, with no other buttons or possibility to go back or rewind is shown and the participants are learning as if the participants in the group which learns only with visual mode, using only the video.
On the second day, participants are, figuratively speaking, moved to the other group, the one which learns only with gamification. Hence, the second day includes the normal pace learning speed, using the game with the camera feed where the participants see themselves, following the instructions in the instruction track in the bottom of the screen showing the images indicating the previous, current and next move.

The goal of this mode of learning is to explore if later inclusion of gamified learning would make a difference in the learning process and if it would accelerate the learning, as opposed to the gamified mode of learning. The assumption is that the participants would get familiar with the skill on the first day with a traditional mode of learning, such as the visual one, learning only with a video tutorial, and that on the second day the game would capture their attention and focus and increase their engagement, hence increasing their performance and accelerate the learning.

4.3 All Modes Combined - Free Learning

The free learning mode includes all modes of learning mentioned and more. The nature of the sensory channels and feedback is the same, however additional elements are added. The group assigned to this mode has the freedom to choose what they will learn with and when. The main difference is in the choice what to learn with within visual learning group. In this case, besides the video the participants had the opportunity to choose printed images of movements to replicate, in the correct order, they could color code the images with color pencils, plastic flowers of different colors, different color stickers and similar. Additionally they could cut the images and write notes on the images or on a separate peace of paper. If the video is chosen, on the first day the participants would be given the slow version of the video, however as the learning would progress they could choose any speed of the video at any given moment.
Additionally, they could choose the already described audio learning mode, with the same video from the visual mode, with the audio instructions layered over the music and the video, in both slow and normal pace mode. This mode could be chosen whenever during the learning process, for how many time a participant wants within the given timeframe for learning, during one session.

As with the previous modes, gamification mode is exactly the same as already described. It has both slow and normal pace, and can be combined with any other mode at any given time, for how many times the participant is willing to learn with it within the session.
Participants have complete freedom to choose what they would learn with, at what point during the session. They can change their strategy at any given moment, and they do not have to predefine the learning plan before starting, they can learn freely with the given options. However, they can not manipulate the videos, rewind or pause, as other participants couldn't do it either. They can only choose from the modes other participants learned with and choose additional visual tools to learn with such as images, using color coding, notes and similar.

The main goal of this learning mode is to compare if there are any differences in the performance and learning process if a person has a free will in choosing the mode of learning. The point of comparative analysis among different modes is to explore if the participants who could combine and choose freely the mode of learning if they would have accelerated learning and overall better performance as opposed to other modes of learning. Additionally, if there is any preference within the group of Free Learning towards a certain mode of learning.

4.4 Task Design

Motor skill acquisition is fundamental in disciplines such as dance and yoga, where learning complex movement sequences is essential. In dance, motor learning begins with perception, as students observe and organize their experiences when a teacher demonstrates a combination or skill. This process involves attention and perception, which are critical components in the initial stages of motor learning [WKC09].

The cognitive processes involved in dance, such as learning and memory, visually and spatially orienting, mental imagery, and multimodal sensory-motor integration, shows the complex nature of motor learning in this art form [HBT24]. Understanding these processes can improve teaching methods and enhance skill acquisition in both dance and yoga.

Incorporating insights from these disciplines into motor learning research can provide a more comprehensive understanding of how complex motor skills are acquired and refined. This knowledge is valuable for developing effective training programs and instructional strategies in various motor skills and practices.

Motor skill is an umbrella term for wide spectrum of skills, from sports activities such as snowboarding [RD00a], to playing instruments such as piano [vdLSBJ11], drums [HBDH10] to choreography [KHX+19], and many other. Hand mudra choreography is a fine motor skill since it involves precise and coordinated movements of small groups of muscles in the hands and fingers. Fine motor skills are used and are of greatest importance in tasks involving precision and control [Mag07]. Studies so far stress that when practicing fine motor skills neural plasticity and coordination are greatly improved, which further makes them suitable for a study such as this one, for investigating the motor learning processes [MA13]. In addition to that, the sequential nature of hand mudra movements fall perfectly under theories of motor skill learning since motor skill acquisition involves encoding, retention and retrieval of movement sequences [Kra06]. These factors that include not just physical movement but the sequential nature of the Hand Mudra choreography, make it ideal for exploring and understanding motor learning in controlled experiment. Motor skills are more than muscle contraction and physical work. They involve cognitive work, with certain sequences in their performance, retention and retrieval of those movements [SLW+19]. For these reasons Hand Mudra choreography is chosen as the adequate skill for participants to learn.

Moreover, since it involves only one hand and its fingers, the choreography is very suitable for analysis, scoring, and quantitative data gathering and analysis for more careful and precise learning process and progress recording and analysis. With only laptop camera the hand and fingers are recorded and evaluated on the success of performance in real time. It is far more simple than including the entire body, it is suitable for university setting, it requires less effort, hence participants are more keen to practice it in the span of two days, and it doesn't require any additional equipment, hence it is easier to analyze with only one machine learning library and simple calculations, which makes it possible to be done precisely in real-time. Furthermore, the advantage of consistency in the motor skill given to learn helps the participants replicate the same gestures over all trials, making it easier for evaluation and process analysis as well.

The main advantages include:

- Simple implementation: easy for participants to learn and practice and for researcher to conduct the study, evaluate and analyze
- Consistency: In replicating and learning the same gestures for participants and for careful analysis and progress tracking for researchers
- Feasible task to learn in 2 days

The task itself is a hand choreography that lasts 37 seconds, comprised of 8 different Hand Mudra hand gestures, which make a twice repeating sequence of 18 gestures in a row, with 36 steps in total, part of choreography can be seen in Figure 4.1.

4.5 Methodologies

4.5.1 Overview of Methodological Approach

In this study, mixed-methods approach was chosen with the goal to use the strengths of both qualitative and quantitative methodologies. The dual approach enhances the reliability and depth of findings. Qualitative and quantitative methods support each other and provide deeper understanding, by capturing user experiences and providing measurable data on performance and engagement. User experiences provide better understanding of the quantitative results, describing them.

The quantitative methods include NASA-TLX and IMI questionnaires as well as the data gathered from the machine learning hand gesture recognition algorithm, the mentioned MediaPipe Hand recognition framework, which offers objective measure of participants' performance and workload. These methods and tools allow for consistent scoring and the identification of patterns across different learning modes. The machine learning scoring system ensured accuracy in performance evaluation, and mitigates observer bias, especially in the case if the ML algorithm was not used but rather the observer counted the number of correct steps, leading to more valid and reliable results, supporting validity of the study itself.

Qualitative data, in this case structured interviews, provide insight into the participant's subjective experiences and perceptions during the learning process. Qualitative methods in this case have the role of supporting and deepening quantitative data, enabling more meaningful conclusions. The insights gathered from the interviews give deeper understanding of emotional and cognitive aspects of the learning process, such as strategies, perceived competence, enjoyment and similar.

By combining these methods, the study gives researchers access to a wide range of data that is both qualitatively rich and quantitatively precise. This method makes it easier to look into how participants feel and how those feelings relate to the results that can be measured. This, in turn, gives us a full picture of how different learning modes impact motor skill acquisition, perceived effort and enjoyment, and how those relate and impact one another.

4.5.2 Questionnaires - NASA-TLX

In this study two questionnaires are used namely NASA-TLX and IMI, for investigating perceived load (e.g. mental and physical), as well as perceived effort and enjoyment while performing a given task. NASA-TLX is short for NASA Task Load Index, developed by NASA, more specifically the Human Performance Group at NASA. It is a widely recognized and used tool, as previously mentioned, for measuring perceived workload while performing a given task, across six dimensions [NAS24]:

1. Mental Demand
2. Physical Demand
3. Temporal Demand
4. Performance
5. Effort

6. Frustration

These six dimensions provide a comprehensive view of the cognitive and physical challenges experienced by participants during a task [HS88].

NASA-TLX in its standard form has a 20-point scale, however in certain studies, as in this one, the 20-point scale has been reduce to the 7-point Likert scale, where 1 represents "Very Low" and 7 represents "Very High". This approach was chosen since it simplifies the rating process for participants, especially when administered multiple times across different sessions, interchangeably with other questionnaires as well, which is case in this study. Previous studies, such as the one by Hamari in 2014, have demonstrated that such modifications can yield valid and reliable results [HKS14].

Likert scale is named by the scientist Rensis Likert, and his dissertation at Columbia University in early 1930s [Lik32]. His goal was to create a more practical and statistically robust method with the goal to measure attitudes. It is a psychometric tool used in questionnaires to asses attitudes, opinions, perceptions. It is a scale that represents levels of agreement and disagreement, usually starting from "Strongly Disagree" to "Strongly Agree" [JKCP15]. This helps in assessing people's responses quantitatively. Likert scale over time gained great attention and adoption, due to its ability to capture direction and intensity of opinions. Overtime, the scale was introduced with certain adaptations with different point ranges e.g. 7-point scale, in order to improve granularity and precision in research. The scale is symmetric, having equal number of positive and negative options, having the midpoint often as the neutral option such as "Neither Agree nor Disagree".

For the purpose of keeping high level of granularity, but also simplifying and making it easier for the participants to take the questionnaire in this study the NASA-TLX is a 7-point scale, as in other, mentioned and approved research examples. The decision to use a 7-point Likert scale instead of the original 20-point scale was made to simplify the survey process and reduce cognitive load on participants. The 7-point scale still captures meaningful variability in responses while making the assessment more intuitive and user-friendly. Research has shown that such adaptations can retain the sensitivity of the tool while enhancing participant compliance and ease of use [MA00].

The NASA-TLX provided a structured, validated framework for quantifying subjective workload across diverse dimensions. Its integration into the study enabled:

- Comparative analysis: Measuring changes in workload perception between the two days offered insights into the impact of task familiarity and repetition.
- Complementary insights: When combined with performance data from the machine learning algorithm and qualitative feedback from interviews, the NASA-TLX added depth to the evaluation of participant experiences.

The NASA-TLX was administered at the end of each day of the study to assess participants' perceived workload after performing the hand choreography tasks. Each participant completed the questionnaire independently, rating their experience across the six dimensions. The survey captured participants' reflections on how demanding they found the task, their perceived performance, and their levels of effort and frustration.

- Day 1: Participants engaged with the task for the first time, completing the NASA-TLX afterward to record their initial workload perception.
- Day 2: The same task was repeated, with the NASA-TLX administered again to identify changes in workload perception, potentially influenced by increased familiarity and task efficiency.

This consistent approach across the two days allowed for direct comparisons of perceived workload, highlighting the impact of task repetition and learning on participants' experiences.

4.5.3 Questionnaires - IMI: Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory, IMI, is used for measuring intrinsic motivation and engagement during given activities. It represents a validated tool used for this purpose, and has been used widely in the research community [LBU24].

The tool was originally developed in early 1980s through foundational research by Ryan in 1982 [Rya82], as well as Ryan, Mims and Koestner in 1983 [RMK83]. Major changes and developments were then brought in 1990 by Ryan, Connell, and Plant [RCP90], and later by Deci et al. in 1994 [DEPL94]. These studies helped establish the instrument's reliability and validity for assessing different dimensions of motivation across diverse activities and populations.

The IMI comprises seven subscales, each capturing a distinct aspect of motivation and related experiences:

1. Interest/Enjoyment: Measures intrinsic motivation; often considered the primary indicator of a participant's engagement and enjoyment in the activity.
2. Perceived Competence: Assesses feelings of capability or effectiveness during the activity.
3. Effort/Importance: Evaluates how much effort participants exert and the significance they place on the task.
4. Value/Usefulness: Captures participants' perceptions of the activity's relevance or utility, especially in internalization studies.
5. Pressure/Tension: Reflects feelings of stress or coercion experienced during the task; negatively associated with intrinsic motivation.

6. Perceived Choice: Measures the extent to which participants feel autonomous and volitional in their engagement with the activity.

7. Relatedness: A newer subscale that examines feelings of connection to others during the activity; commonly used in studies on interpersonal interactions.

The biggest strength of IMI is it's flexibility, since it can be tailored to fit a specific task or a context, without compromising reliability. For example, the generic option in the questionnaire, "I tried very hard to do well at this activity" can be adapted to "I tried very hard to learn this material". The biggest advantage is that the custom items can be adjusted to be more accessible and understandable for the participants, hence easier to administer.

However, IMI also has it's weaknesses such as redundancy or items as well as subjective self-reporting of participants. Even though the overlapping items can improve reliability, it can also cause participant fatigue, hence it is important to be cautious when putting the questionnaire together. Furthermore, correlations between self-reports and behavioral measures are often not high, due to the factors such as ego involvement, self-presentation bias and similar.

There are several variations of the IMI, each providing a design of the study according to the research context:

1. 22-item standard version: Includes four subscales—interest/enjoyment, perceived competence, perceived choice, and pressure/tension.

2. 9-item short version: Focuses on reading tasks, with subscales for interest/enjoyment, perceived competence, and pressure/tension.

3. 25-item version for internalization studies: Includes value/usefulness, interest/enjoyment, and perceived choice.

4. 29-item version for relatedness: Includes five subscales—relatedness, interest/enjoyment, perceived choice, pressure/tension, and effort.

The inventory, as already mentioned, has several subscales, however for this study the following subscales were used:

- Interest/Enjoyment
- Perceived Competence
- Effort/Importance
- Pressure/Tension

These subscales were chosen to provide insight into participants' motivation, perceived performance as well as their emotional responses during the learning process [RD00b]. The subscale of Interest/Enjoyment is a direct measure of intrinsic motivation, while other subscales provide additional context about participants' experience and engagement with the task. These dimensions contribute to understanding how different learning modes, such as visual, auditory and gamified learning, affect participant motivation and overall satisfaction.

The IMI was administered at key points of the study in order to capture participants' perceptions across various stages of the learning process, more specifically at the end of each session, or more precisley at the end of each day.

Each IMI subscale included a series of statements rated on a 7-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree). Example items included:

- For Interest/Enjoyment: "I enjoyed doing this activity very much."
- For Perceived Competence: "I think I am pretty good at this activity."
- For Effort/Importance: "I put a lot of effort into this activity."
- For Pressure/Tension: "I felt very tense while doing this activity."

In conclusion, the IMI provided a good insight in understanding participants' emotional and cognitive engagement with the learning process. Its inclusion in the study offered understanding in how different modes impact intrinsic motivation and satisfaction, as well as how perceptions of competence and effort influence participants' experiences and performance. Together with the scores gathered from machine learning algorithm analysis, the comparison between subjective experiences and objective performance along different modes of learning can be done and deeper understanding obtained.

4.5.4 The Scoring System - ML Algorithm

In order to further support and have a deeper insight into subjective perceptions of participants about their invested effort, success and performance, but also having a hard a data on showing the progress of learning, tests where participants are showing their level of choreography knowledge were introduced. The tests were administered multiple times, in the middle and at the end of learning sessions, in order to track the progress, and compare the data among different modes of learning. This system allows precise tracking of participants' performance without the need for an observer, reducing subjectivity and potential observer bias or mistake.

Tests were designed to measure participants' knowledge and skill of the motor skill, choreography, at multiple points: mid-session and post-session on both days. These checkpoints offer insight and tracking of the learning process along different learning modes. During the tests, the participants could see only the live video of themselves performing the choreography on the laptop's camera, and hear the music, enabling them to know when to start and finish their choreography. This approach made sure that the participants rely solely on their knowledge and skill, without relying on real-time feedback.

The scoring system uses MediaPipe framework, the mentioned machine learning library that extracts hand and finger position in real time. The data is processed dynamically and saved in a database capturing:

- The timestamp of each movement
- Score for correct or wrong movement performance, "correct" or "mistake"

Even if the participants performed an incorrect movement before or after the current one, any correct position a participant displays gets recorded as an increased score, a point, it doesn't depend on a context, hence even a random correct performance will be recorded as a point. This approach makes sure that all successful movements are captured, even if occasional errors occur. The database stores performance data for each session, number of correct and false steps. This allows us to perform the following:

- Accuracy scoring, counting the number of correct movements as a measure of knowledge and precision, hence allowing us to perform:
- Progress tracking, which allows us to compare performance data across test sessions to understand progress over time, and later to compare it across different learning methods.

The main advantages of this approach include the following:

- Objective scoring
- Scalability
- Consistency
- Real-time data processing

In order to achieve objective scoring the observer who manually counts correct moves needs to be replaced with a system that is consistent, without biases or most common human errors that occur due to lack of focus or subjectivity. With such system in place, that computes errors real-time with a consistent rules for counting objective scoring is possible. Furthermore, the system automates the scoring process, making it possible to test larger groups of participants without additional observers to rotate, which further strengths scalability and consistency, since the scoring remains uniform across all participants and test sessions. Real-time data processing enables immediate storage and analysis of metrics, which further ensures consistency and precision, since the calculated and recorded scores are immediately updated.

The machine learning-based scoring system plays a vital role in supporting qualitative methods. The dual approach of both qualitative and quantitative methods enables better understanding in how participants learn a new motor skill, and what their experience is like, by connecting subjective feedback from interviews and questionnaires, with the objective, real-time performance data. This further provides validity and reliability of the findings, and deeper understanding of the learning process, how effort and enjoyment relate to success of learning.

4.5.5 Interviews

The interviews are chosen in order to gain a deeper understanding of participants' experiences and thoughts which initially couldn't be captured by NASA-TLX and IMI. More specifically, interviews give a greater insight into emotions, opinions and attitudes, which further explain participants' engagement with the learning process. By integrating interviews in this study, there is a opportunity to complete and improve findings of quantitative data together with the qualitative findings and better understand poor or good performance paired with certain attitudes extracted from the mentioned questionnaires.

For this purpose a semi-structured interview approach was chosen in order to balance between the guided questions and flexibility to explore perhaps hidden or deeper reasons for participant's experiences. This format of interviews gave the opportunity to keep a conversational flow if a person was willing to share more but also stay focused on the goals of the study, in order not to deviate too much from the intended topics and their understanding. This approach ensures that all topics and questions are covered, that could further explain the answers and personal performance, but also have the space for the person to share more if willing.
The interviews were conducted after each learning session, or more precisely, at the end of each day, after all tests and questionnaires, as the last step and an opportunity for the participants to reflect on the entire process, including the questionnaires and tests themselves. The interviews were conducted as follows:

- Setting: in person with each participant, in the same room where the experiment was conducted. The interviews were recorded with a mobile phone, with a previous consent approval, at the beginning of the experiment. Keynotes were taken during the sessions.
- Number of participants: 56, each of them giving two interviews, one at the end of each day
- Duration: There was no strict duration of the interviews but an average duration is estimated around 5 minutes each.

Interviews were chosen as a method in order to further support NASA-TLX and IMI questionnaires, as a confirmation, to highlight certain themes such as high workload, physical strain or perception of difficulty and personal success rate. On the other hand interviews were chosen to highlight certain contrasts such as certain frustrations, unique strategies used, or anything that would add further detail to the study, and not just confirm what was already collected by other research methods. In conclusion, interviews were chosen in order to confirm and strengthen findings from other methodologies, giving stronger validity for the research, but also showing some contrasts and giving deeper detail and insight, explaining further or inspiring for further research.

For better understanding some of the questions are listed below:

- What was hard in this challenge? Why was it hard?
- Did you derive any strategy to learn the skill?
- How did you improve performance? What helped you in it?
- How did you improve performance? What helped you in it?
- Did something in the learning process influence your enjoyment?
- Did something in the learning process increase or decreased effort? When do you feel it happened?

4.5.6 Ethical Considerations

Steps for ensuring ethical research have been implemented from the beginning of the research process, more precisely, from planning, recruiting, conducting research to storing the data, analyzing and reporting.
The recruiting process of participants was done online, with a call distributed to the students of Technische Universität Wien, with a detail description of what the research is about and what is expected of the participants. They were informed that the study takes place on two days in the University premises, and that for the learning process they would play a game or learn a motor skill in a different way.

Once the participants arrived for the experiment they were informed again, and in more detail what is the study about, what is the goal of the study and they were verbally taken and informed on the process of the study that they would experience. Afterwards they were given the consent form where they could read what the study is about, what is the goal and what is expected of them. Furthermore, in the consent form they were informed and asked to give their consent for the data gathering, describing what data will be recorded and later used. Additionally, they were informed that the data used will be anonymized, since each participant was given an ID number to be used for the data analysis and later reporting. Once signed they gave their consent for the study to take place and for the data to be recorded and later used under the conditioned that it is

anonymized.
Data reported here will be anonymized and only referenced by the given ID on the participant's consent form.
During the experiment itself, the participants were not forced to do anything, and it was not expected of them to do anything that would cause them harm in physical, emotional or any other way.

Each participant was thanked after the experiment and hosted with professional and warm conduct of the researcher. Each of them had their time which was uninterrupted by other participants, the interview or their performance was not monitored by anyone other than the researcher themselves, and all data and experience was left between the researcher and participant in the room where the experiment was conducted on both days.

4.6 User Study Design

The user study was organized with 56 participants in total. All participants have the same educational background in technical sciences. Most of the participants were students at bachelor level studies, however there were some studying master studies and one graduate as well. Participants were not asked about their age, however the estimate is that the most of them are between 22 and 25 year old, with a few of them being older. More detail on number of participants is represented in Results section.
They were all tasked to learn the same hand choreography without prior knowledge of the gestures. All of them were unfamiliar with the task and gestures before coming in, with an equal distribution of participant playing an instrument, having experience in dance or similar games such as the one offered. All participants were learning in the same environment over the course of 2 days, 30 minutes each day, in the same room, at TU Wien premises. They were tested for their improvement in total 5 times. Graphical representation of the user study can be seen in Figure 4.2. Additionally, to see the room at university premises and where the study was conducted, see Figure 4.3.

As described previously, the task at hand was to learn the Hand Mudra choreography, a dance choreography involving one hand, composed of combination of different hand gestures. In general, Hand Mudra has no predefined sequence, but a defined gestures that one can combine in their performance. For the sake of mensurability and comparability in the study all participants needed to learn the same sequence, the same choreography. Repeating once more, this task is chosen since it is a motor skill which is a part of yoga and dance choreography, changing hand gestures in a specific order. Furthermore, the task is learnable over the course of 2 days, since it is consisted of a repeating pattern, choreography being 37 seconds long. The choreography is accompanied by a suitable melody. This task represents a balance between simple and learnable one over the course of two days, however still imposing a challenge, since it is comprised of 8 different, repeating gestures, with a once repeating sequence of 18 gestures which is

made of the previously mentioned 8 different ones, combining them in a sequence of 18. Additionally, Krakauer et al. [KHX+19] suggest that while simple learning tasks might not fully capture the complexity of acquiring real-world skills, they do provide crucial insights into the basic elements of the learning process. These elements are required, if not completely enough and adequate, for describing and understanding acquiring more complex and sophisticated skills [KHX+19].
Furthermore it is important to say that feedback plays a vital part in the motor skills acquisition providing individuals information which the individuals generated themselves through different intrinsic sensory receptors, which come from their own movements [GPC13, OGS01, SL19]. All motor tasks generally include this feedback system [OGS01].

The study contained five different groups of participants, where in each group participants are learning with a different mode of learning. Participants are learning with the same mode on both days, which they were assigned at the beginning of the study. Building on the previous statement, this study is a between-group study, meaning that each participant will belong and participant only in one group, the one they were initially assigned to. The five different modes of learning are: visual, multimodal including visual and auditory, gamified, a combination of visual on the first day and gamified learning on the second, and last but not least, a mode combining all of the previous ones.

As each mode of learning was described in detail in the pervious text, here is only a short description, for easier following.

- VIDEO of choreography accompanied with music, making it a visual mode,
- AUDIO containing the same video with music as in previous mode overlayed with audio instructions,
- GAMIFICATION mode using a game in which the participant sees themselves real-time in camera, following steps shown in images at the bottom of the screen, with immediate feedback on their performance (right or wrong shown in animation)
- VIDEO condition used on the first day, and GAMIFICATION used on the second day of learning, to investigate if later inclusion of gamified learning has benefits or any differences from using the gamified learning from the beginning,
- FREE LEARNING, accounts all of the mentioned possibilities, modes of learning with the additional option of learning from images printed on a paper and keeping notes. Participants develop their own strategy for learning and within the same amount of time as the other groups have, choose how they will learn.

The study was organized and conducted as follows: The participants were greeted and given the consent form. Signing of the consent form was followed by a short introduction on what is the purpose of the study and how the study will play out so that the participants have idea about what will go on and so that they could focus better.

Introduction is followed by introducing the task, and showing the participant the choreography that they should learn over the course of two days. For this purpose the 37 seconds video of the choreography is shown in normal speed so that the participants know what to expect and what is the end goal. If the participant has no question the study session starts.

On the first day the studying was organized in two batches of three learning rounds, meaning that the participant is met with watching a video in a slow mode, of 0.5x speed of the original choreography, three times in a row, for two times, with a test in between. They are allowed to follow along or just watch. In the case of gamified learning they perform the choreography by playing the game, trying to score the correct move.
After the first learning round the test is administered. During the test the participants see themselves in the camera, and hear the music of the choreography. Once the test is started there is a countdown of 3 seconds before the evaluation starts. The participants during the test do not see any feedback on the screen except themselves as in a mirror. After the test, NASA-TLX questionnaire is administered for the first time.
This was followed by the second batch of learning starts, organized in the same way as the first one, followed by another test for the end of the day.
After finishing the learning and studying for the day, participants were introduced with the IMI questionnaire, which was followed by the interview.

On the second day, participants learn with normal speed. The learning is organized in three batches, with six rounds of learning in the first batch, six rounds again in the second batch, and three rounds in the third batch.
After each batch of learning the test is administered, in the same pace as the one participants learned with.
The questionnaires, NASA-TLX and IMI are administered again, followed by the interview.

However, the Free Learning mode of learning differs. The learning is organized on the first day also in two batches, or rounds, with tests following them. Participants have the freedom to combine different modes of learning offered to other participants with addition to images of gestures placed in the order of choreography, and ability to take notes. The image map and pictograms that were used in game and also were offered as a pictogram map can be seen in Figure 4.1 and Figure 14, respectively. Within the same timeframe as all other participants had, the ones belonging to Free Learning mode had 10 minutes in each batch to learn before being examined with the test. On the first day, they learned for 10 minutes, followed by a test, which was repeated once more.
On the second day they learned also in three batches. First two batches were organized in 8 minutes, and the last one in 4. Each batch was followed by a test.
After the first test on the first day participants were introduced with NASA-TLX, and at the end of the day with IMI questionnaire, followed by the interview. On the second day questionnaires and the interview were administered at the end of the day.

The tests are used to assess how much of learning happened in that time frame. They happened after each batch of learning, 2 times on the first day and 3 times on the second day.
The first day was used for the participants to get to know the choreography, task and develop muscle memory, while on the second day they had more extensive training sessions. However, the time dedicated to learning is almost the same, with only the last batch of 3 rounds exceeding the equal time from the first day. Participants showed everything that they learned up to that point on a computer camera, looking at themselves, following the music to perform the choreography. Their performance was be rated by the hand gesture recognition algorithm. The initial test and the final test are followed by NASA-TLX[Hum86] survey(described in “Methodologies” section), while the learning and testing sessions at the end of each day are followed by IMI survey[Rya82] (described in “Methdologies” section). For Consent Form and Experiment Guide see Figure 15, Figure 16, and Figure 17, Figure 18 in Appendix B.

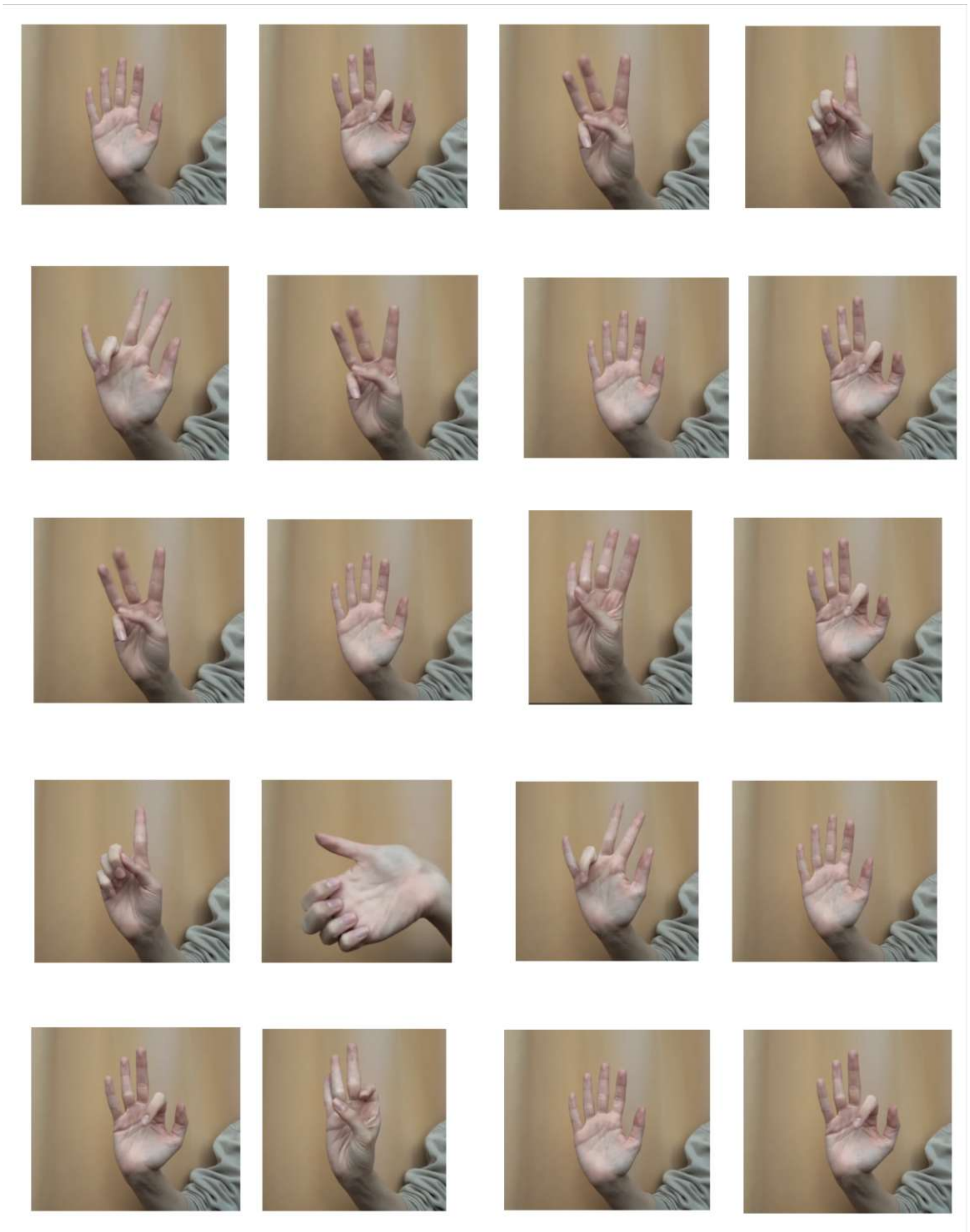


Figure 4.1: Part of choreography shown in image map

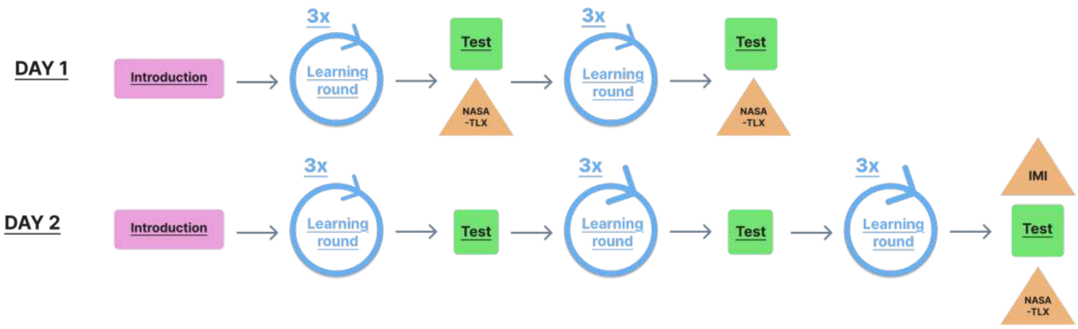


Figure 4.2: Visual representation of study design

(a) AUDIO mode of learning (b) FREE LEARNING mode of learning

Figure 4.3: Study setup, TU Wien

CHAPTER 5

Results

Quantitative analysis was all done in Python using following libraries:

Qualitative analysis on interviews included performing thematic analysis and looking for themes that would further explain or confirm quantitative data analysis results or show new patterns. Interviews were transcribed using DaVinci Resolve software.

For the sake of better readability of visualisations and tables, learning modes will have the following form:

- VISUAL - VIDEO
- AUDIO - AUDIO
- GAMIFICATION - GAMIFICATION
- VISUAL AND GAMIFICATION - VL
- FREE LEARNING - FL

Participants were students of TU Wien, mostly coming from a bachelor course on scientific research and writing, with a couple of more masters students and one working person who also studied and successfully completed their studies at TU Wien. Due to the nature of the study participants were not asked for their age, hence there is no exact insight, but it is estimated that the students ranged 22-26 years of age. Each group initially was intended to have equal number of participants, 11 per group, however, due to a mistake in counting the numbers while executing the experiment, number of participants in each group are as follows:

- VIDEO - n=9

- AUDIO - n=11
- GAMIFICATION - n=11
- VISUAL AND GAMIFICATION - n=14
- FREE LEARNING - n=11

In total there are 56 participants in 5 different groups. What is important to mention is that the following statistical tests will show that data is still valid and comparable. Additionally, the group with least participants is VIDEO, which is a group used as a baseline but also not as a group that is estimated to yield valuable results in study as the previous research proved that multimodal learning as AUDIO is will yield better results. Hence, VIDEO mode is included for comparison reasons, as a reference point. The group with most participants is the most important group which is being observed and researched as a new way of learning, hence increase in number of participants helps to validate the conclusions even more.

5.1 NASA-TLX

The NASA-TLX responses were analyzed with the goal to compare the workload experiences among the five learning modes, namely VIDEO, AUDIO, GAMIFICATION, VIDEO AND GAMIFICATION and FREE LEARNING. The responses from NASA-TLX were grouped by the mode of learning as well as by the session, having one on the first and the other on the second day of learning.
In this case the unweighted average of all six scales was applied while calculating the individual NASA-TLX scores, since the participants were not asked to compare subscales pairwise based on their perception of importance of each scale, they were all equally weighted, since all subscales were equally important. The six NASA-TLX subscales - Mental, Physical, Temporal, Performance, Effort and Frustration, were used to calculate both individual and later subscale scores, including overall group scores as well. The goal was to find the differences in the workload experiences across groups and iterations. The data was not filtered from outliers for the following reasons. Firstly, NASA-TLX does not instruct for the outliers to be filtered out, but rather to take average values of reported scores. This is crucial since the goal of NASA-TLX is to gather subjective, personal perceptions of workload, understanding how perceptions of effort(and later enjoyment as well) influence the learning process. Outliers for this purpose are valid and important pieces of data in order to understand the perceptions of participants and make valid conclusions. Furthermore, as it can be seen in the Figure 4, all conditions have outliers, which further reaffirms the importance of keeping the outliers in the analysis. Last but not least, in the statistical analysis tests that are robust and more resistant to outliers are used such as Kruskal-Wallis test. This approach ensures a detailed overview and proper understanding of participants' perceptions, keeping all answers and participants

present in the analysis.
The formula for calculating unweighted scores is the following:

$$\text{NASA-TLX Score} = \frac{\text{Mental} + \text{Physical} + \text{Temporal} + \text{Performance} + \text{Effort} + \text{Frustration}}{6}$$

The data was stored in a csv file which was filtered and changed so that it has the following columns:

- UserID
- Condition(Mode of Learning)
- Iteration
- Mental
- Physical
- Temporal
- Performance
- Effort
- Frustration

UserID is a column which was not use for the analysis but it initially helped with ensuring that all participants were included in the analysis in each iteration. Data was further grouped by iterations, more precisely sessions or days on which the NASA-TLX was administered. After grouping the data by iterations, it was further grouped by the condition, or learning mode. This way, data was organized in such a way that it was possible to compare the results among different learning modes on both days separately. The NASA-TLX scores were analyzed to compare workload experiences across groups and iterations. Individual participant-level NASA-TLX scores were calculated as the unweighted average of six subscale scores. These scores were afterwards aggregated to compute group-level statistics, including mean and standard deviation for both overall TLX scores and subscale averages. Additionally, subscale-specific group-level averages were calculated for Mental, Physical, Temporal, Performance, Effort, and Frustration dimensions. This approach ensures that each participant's data contributes to group-level trends while providing detailed insights into how specific workload dimensions vary across conditions and iterations, emphasizing both broader differences and subscale-level patterns.

The following workflow of data analysis was implemented:

- Normality testing: before moving to statistical tests Shapiro-Wilk test was performed

- Statistical tests:
  - Between-condition comparison: Kruskal-Wallis test and post-hoc pairwise Mann-Whitney U test were used to compare groups
  - Within-condition comparison: Wilcoxon signed-rank tests were used to compare iterations for the same condition
- Descriptive analysis: computing average scores for each subscale, condition and iteration
- Visualisations in the form of boxplots and barcharts were used to visualize the differences between groups and iterations

Detail analysis for each of these tests follows, starting with the normality tests. Normality test of the individual NASA-TLX scores, for each group and iteration was conducted. Normality tests are conducted in order to apply adequate statistical tests later, parametric or non-parametric. For testing normality Shapiro-Wilk test was used. Shapiro-Wilk test was first published in 1965 by Samuel Sanford Shapiro and Martin Wilk [SW65]. It is testing the null hypothesis that $x_1, \ldots, x_n$ is coming from a normally distributed population [SW65]. In null-hypothesis significance testing p-value is the likelihood that, if the null hypothesis is valid, test results will be at least as extreme as the actual ones [Asc15, WL16]. A small p-value, or a p-value below the chosen value of 0.05, means that the extreme observed outcome is very likely not going to happen under the null hypothesis, hence:

- **p-value** > 0.05: The data is likely normally distributed.
- **p-value** ≤ 0.05: The data is likely not normally distributed.

What can be seen in Table 1 found in Appendix A, is that VIDEO and GAMIFICATION modes do not have normal distribution. For this reason ANOVA(Analysis of Variance) will not be applied, bur rather Kruskal-Wallis H-test for the non-normal distributions, as a non-parametric test. ANOVA and Kruskal-Wallis are performed in the case where there are three or more groups being compared against each other, as in this study where there is five different groups, and since some of the data has non-normally distributed data Kruskal-Wallis test is chosen. In order to perform comparison against iterations within the same learning mode group since there are some conditions with the non-normal data Wilcoxon signed-rank test. Wilcoxon test is chosen in this case since only two groups are compared against each other. What is clear from this analysis is that there is no significant difference among the two conditions, indicating stable perceived workload over time within conditions, suggesting a gradual learning over the two days, as shown in Table 2.

For between-group comparison, meaning across conditions, Kruskal-Wallis test was used due to non-normality in data distribution in two mentioned groups. The results for the first iteration among groups are that the p-value is much larger than 0.05, with the value of 0.283, which indicates no significant differences among groups regarding workload. Concluding that in the first iteration all participants perceived and experienced the same amount of workload, which is to be expected as everyone is facing the skill for the first time, starting with no knowledge or previous skill. However, in the second iteration p-value drops below 0.05, with the value of 0.032 indicating significant differences between conditions, seen in Table 5.1.

Table 5.1: NASA-TLX: Kruskal-Wallis Test Results

Iteration	**Test**	**Statistic**	**P-Value**
First	Kruskal-Wallis	5.0421	0.2830
Second	Kruskal-Wallis	10.5733	0.0318

Even though there are no big changes in perceived workload in participants of the same learning mode over time, there are significant changes among different learning modes, meaning that some learning modes evoked feelings of higher working load than others.

Since there were significant differences in second iteration among different conditions Post-Hoc Pairwise Comparison was done. This comparison was applied for conditions in the second iteration using Mann-Whitney U test for pairwise comparisons. Significant differences were found in the following pairs:

- GAMIFICATION vs FL, with $p = 0.045$
- VG vs FL, with $p = 0.0047$

However, after applying Bonferroni correction in order to avoid false positives, only significant difference can be seen between VG and FL. For detail result overview see Table 3, Appendix A.

Participants in FL condition reported significantly lower workload scores than the ones in VG. This indicates that participants who had gamified learning offered to them experienced significantly higher workloads than other participants. This can be further seen in Table 4, Appendix A, and Figure 4, Appendix B.

When analyzing average scores this is especially noticeable on the Physical subscale when comparing GAMIFICATION and FL, with the differences of score 4.57 and score 1.73 respectively, but also for all other subscales, with Mental subscale following with differences in scores 5 vs 3.56, and perception in Performance, where the difference was 3.07 vs 5.27 in favor of FL. When comparing VG and FL the differences are most noticeable in Physical subscale with VG scoring 5 and FL scoring 1.73, however Mental subscale not falling too far begin with a slightly smaller difference than in the previous pair, VG scoring 4.91 and FL 3.36. The differences continue on the rest of the scales as well. Regarding overall scores, FL had by far the lowest score in the second iteration of

3.27, while GAMIFICATION with 4.06 and VG with the score of 4.45 where the ones with the highest overall load score.

To further explain and support these foundings, descriptive analysis follows.

The most noticeable difference can be seen in reduction of all workload scores from first to second iteration for all conditions. This was expected, since the participants are more familiar with the task on the second day and overall effort and workload should be lower since the learning progresses and increases. What is most noticeable is that the FL condition includes participants which had the most significant improvement from all other conditions, reporting lower mental and physical demand, as well as higher performance. Furthermore, Mental Demand score decreased for all conditions, most noticeably for AUDIO and FL, and Frustration decreased in AUDIO, FL and VG conditions. The Table 5.2 for detailed subscale-level analysis for each group follows:

Condition	Iteration	Mental	Physical	Temporal	Performance	Effort	Frustration
AUDIO	First	5.18	2.91	4.55	2.36	4.18	3.82
AUDIO	Second	4.09	3.64	4.09	4.36	4.55	2.55
FL	First	4.55	2.73	3.91	2.73	4.09	3.18
FL	Second	3.36	1.73	3.09	5.27	3.64	2.55
GAMIFICATION	First	4.86	4.29	4.36	1.79	4.00	3.86
GAMIFICATION	Second	5.00	4.57	4.14	3.07	4.71	2.86
VG	First	5.36	3.45	5.18	2.00	5.09	4.18
VG	Second	4.91	5.00	4.91	3.18	5.00	3.73
VIDEO	First	4.56	2.56	4.33	2.11	3.56	4.33
VIDEO	Second	3.78	3.56	4.56	3.44	4.67	3.44

Table 5.2: Descriptive Analysis of NASA-TLX Subscales by Condition and Iteration

Continuing with the detailed analysis, focusing on the first iteration, across all six subscales. In the mental load scale for the first iteration results show that the participants learning with VG mode experienced highest mental load, followed by AUDIO, while those learning in FL mode experienced the least mental effort. However, differences are relatively small, indicating that all participants had a similar experience when faced with a new, unfamiliar task.

Additionally, when it comes to physical load lowest reported perceived load is reported in FL and very closely following by VIDEO, while GAMIFICATION group reported high levels of physical load. This implies that those learning through a game had to focus to perform physically as better as possible in order for the system to accept their movement as the correct one, receiving positive feedback, while processing it mentally, and remembering was not their priority.

Moving to the Temporal subscale perceived load is somewhat high on all subscales, with VG showing the highest and FL the lowest.

When it comes to the participants' perceived performance it is overall very low for the first iteration which is to be expected as the participants are getting to know and become familiar with the skill. Highest reported score is recorded with the participants from FL group, closely following with VIDEO and AUDIO, while GAMIFICATION recorded the

poorest score.
While analyzing perceived effort the highest score is seen in VG, which correlates with highest mental load and its' respective reasoning, while VIDEO records the lowest score. Interestingly, participants in VG and VIDEO in the first iteration both learned the same way, and the participants in VG were not aware that the learning mode would change on the following day.
Lastly, in Frustration subscale most frustrated were participants in VIDEO group, following with VG and GAMIFICATION, while least frustration reported participants of FL group.

In the second iteration highest mental load is proven to be highest in GAMIFICATION, scoring 5 on the scale. This scores is followed by VG, while the lowest score was reported in FL.
When it comes to the physical load, highest load was reported by the participants from VG group, scoring 5, followed closely by GAMIFICATION, with the lowest score in FL, with the score of only 1.73.
In Temporal subscale highest score was reported in VG, followed by VG, while lowest was reported in FL.
When it comes to performance most confident are participants learning in FL group, followed by a drop in AUDIO, while the least confident were the ones in GAMIFICATION and VG.
Highest effort was reported by participants learning with VG, following right after by GAMIFICATION, VIDEO and AUDIO, respectively, while lowest effort was reported in FL.
Highest frustration was reported in VG, while the lowest was reported in AUDIO and FL equally. This can be explained with the results from other methodologies used in this study, explaining that the personal expectation of participant in GAMIFICATION were that they would learn and score much better on the second day while their confidence was turned down and frustration rose after seeing little progress. Adding on to that, even though in AUDIO load was reported high on other scales, the performance was very high as well as in the FL group, where participants tackled the skill and accomplished the learning process successfully.

When comparing first and second iteration it is noticeable that mental load drops in all conditions except a slight increase in GAMIFICATION mode, but physical load increases in all conditions except in FL. On Temporal subscale, as in the Mental, a drop of load score is shown across all conditions, except VIDEO, which showed a slight increase. Perceived performance increased in all scales, usually doubling which is a good indicator of a successful learning process in all learning modes. When it comes to effort, results vary a lot, where an increase from first to second day is recorded in AUDIO, GAMIFICATION and VIDEO, while it slightly dropped in FL and VG. Frustration levels all dropped on the second day as the familiarity and knowledge increases and performance rises. These trends can be seen in the Figure 3, see Appendix B.

When analyzing overall raw NASA-TLX score, which is crucial for answering the third research question posed in this study, we can see the following trends, recorded in Table 4 found in Appendix A, and in the Figure 4 found in Appendix B. What is immediately noticeable is that the overall load score is lower on the second day, second iteration, for all groups except GAMIFICATION and VG where the overall score increases. The lowest load scores are noticeable in FL on both days, while the highest load score on both days is displayed in VG and GAMIFICATION. AUDIO group reported higher load score on day 1, but a lower on on day 2 when compared to VIDEO group. This data further reaffirms what has been proven with analysis of subscale loads as well as the statistical analysis in the second iteration, the pairwise comparison of results, which is furhter explained with the followed analysis.

Summarizing, while descriptive analysis highlighted general trends, statistical tests confirmed that the most significant workload reductions occurred in FL, with improved mental demand and performance. In contrast, GAMIFICATION and VG maintained higher workload levels, particularly for physical and temporal demands. The combined approach ensures that observed trends are both robust and meaningful.

5.2 IMI: Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory was administered in order to analyze the perceived enjoyment and positive attitudes during the learning process among the learning modes, in order to answer the third research question. It has been administered on both days with the goal to compare the results and understand if there are any differences between the two learning sessions. The IMI has several subscales, out of which four were used in this study, namely:

- Interest/Enjoyment
- Perceived Competence
- Effort/Importance
- Pressure/Tension

Each participant answered questions from this subgroups that were randomly ordered in the questionnaire. For each subscale the average of the corresponding answers were taken, except for the questions that had the reverse score. After subscores for each participant were calculated, the calculated scores were aggregated and averaged based on the groups, for each iteration separately. The formula for score calculations is the following:

1. Reverse score the items marked with (R) by subtracting the item response x from 8:
$$x' = 8 - x$$

2. Calculate the subscale score S by averaging across all item scores (using reverse-scored values x' where applicable):
$$S = \frac{\sum_{i=1}^{n} x'_i}{n}$$
where n is the total number of items in the subscale.

The goal was to compare the differences in the perceptions of interest and enjoyment, and how effort and pressure relate to that across different group learning modes and iterations. The data was not filtered out of outliers since the IMI instructions do not explicitly instruct to do so, and since there already isn't an equal number of participants in each group, hence for the purpose of keeping the data as balanced as possible across groups all data was kept. Furthermore, since the IMI is a subjective report of participants' perceptions all data is kept as it is relevant for understanding how perceptions relate to objective results. For this reason, and for the reason of absence of normal distribution in all subgroups, robust and more resistant tests to outliers are used including Kruskal-Wallis test in statistical analysis. The same approach is applied as in NASA-TLX analysis also for the sake of consistency of analysis and comparable results across methodologies.

The data was stored in a csv file with filtered out columns ending up with columns including UserID, Iteration, each question resulting in a separate column. The questions are then grouped by the subscale based on theoretical constructs, resulting in subscale columns where averages are calculated for each user. Data was later grouped by modes of learning calculating group averages from individual user scores for each subscale and iteration.

The worfklow of IMI data analysis is the same as in NASA-TLX:

- Normality testing: before moving to statistical tests Shapiro-Wilk test was performed
- Statistical tests:
  - Between-condition comparison: Kruskal-Wallis test and post-hoc pairwise Mann-Whitney U test were used to compare groups
  - Within-condition comparison: Wilcoxon signed-rank tests were used to compare iterations for the same condition
- Descriptive analysis: computing average scores for each subscale, condition and iteration

- Visualisations in the form of boxplots and barcharts were used to visualize the differences between groups and iterations

Normality test performed for groups per iteration showed that not all data is normally distributed, as it can be seen in the Table 5 found in Appendix A, namely, AUDIO and VG groups in second iteration for Perceived Competence, VIDEO in Effort/Importance in first iteration, and GAMIFICATION also in first iteration for Pressure/Tension.

Since not all data is normally distributed, and for already mentioned reasons, the following analysis continues by performing the Kruskal-Wallis test finding significant statistical differences among different conditions. In the groups Interest/Enjoyment in the second iteration, Perceived Competence in both iterations and Effort/Importance in first iteration p-value is found to be lower than 0.05, indicating significant differences, this can be seen in Table 6 found in Appendix A.

When it comes to the pairwise comparisons significant differences are found in Perceived competence in the second iteration for pairs GAMIFICATION and FL, as well as VG and FL, as well as in Effort/Importance in first iteration for pairs VG and FL, and GAMIFICATION and FL, which can be seen in the Table 7, Appendix A.

When looking for significant differences between iterations, significant difference was only found for FL condition for the scale Perceived Competence, meaning that the participants in FL group reported significant changes in their Perceived Competence form first to second iteration, as seen in the Table 8, Appendix A.

The statistical analysis included Kruskal-Wallis tests to compare subscale scores across conditions and Wilcoxon Signed-Rank tests to evaluate changes between iterations within each condition. The Kruskal-Wallis tests identified significant differences for some subscales, such as Perceived Competence and Effort/Importance, indicating that conditions influenced these outcomes. For changes across iterations, the Wilcoxon Signed-Rank test revealed a significant improvement in Perceived Competence for the FL condition ($p = 0.0125$), reflecting participants' increased confidence in their abilities during the second iteration. However, no significant differences were observed for other subscales or conditions.

These findings show the advantages that the FL condition has in developing perceived competence over time. While GAMIFICATION mode of learning did not yield statistically significant changes, it still showed reduction in tension. The AUDIO mode of learning showed consistent performance, while VG and VIDEO did not show significant changes. VG and VIDEO since there was no significant changes can be improved.

The FL condition consistently exhibited the greatest ratings in Interest/Enjoyment and Perceived Competence, with notable enhancements from the First to Second iteration (e.g., Interest/Enjoyment rose from 4.97 to 5.49, and Perceived Competence increased from 3.35 to 4.60). This indicates that the FL condition successfully engaged participants and improved their perceived competence with time. In contrast, GAMIFICATION demonstrated a little rise in Effort/Importance (from 3.67 to 4.36) and a notable decline in Pressure/Tension (from 3.71 to 2.66), indicating a reduction in stress while sustaining moderate engagement. However, Perceived Competence remained decreased in this mode of learning. This is shown in Figure 5.1, Figure 5 Appendix B, and Figure 6, found in Appendix B.

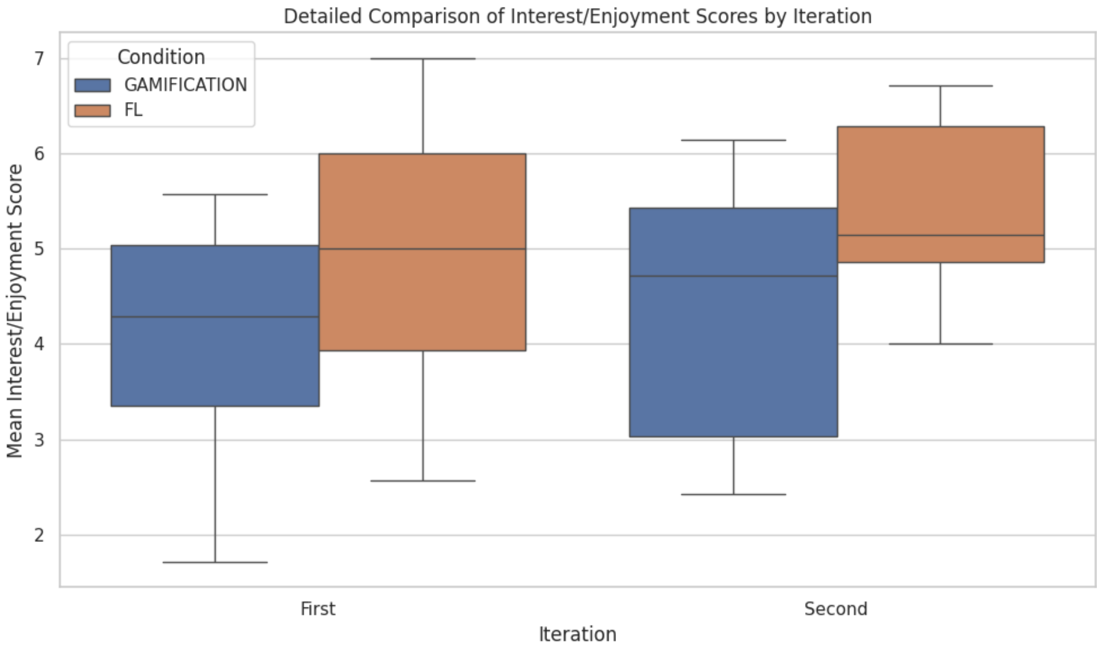


Figure 5.1: Pairwise Comparison Interest/Enjoyment per Iteration

The AUDIO condition exhibited moderate engagement and competence scores, with slight enhancements in Interest/Enjoyment (from 4.17 to 4.54) and Perceived Competence (from 3.13 to 3.95). Additionally, VG and VIDEO conditions continually exhibited lower results, especially in Perceived Competence, which remained below 2.0 for VG and below 3.0 for VIDEO across both iterations. This indicates that these modes of learning could require actions to improve participant engagement and perceptions of competence.

The analysis indicates that the FL condition is the most efficient for enhancing participant engagement and competence, while GAMIFICATION demonstrates potential for reducing tension. The AUDIO condition performs moderately well across subscales, but the VG and VIDEO conditions may require design enhancements to improve their effectiveness.

It is important to mention that Cronbach's Alpha test was performed showing results that all subscales show excellent reliability with values well over 0.7, as shown in Table 9.

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma_i^2}{\sigma_{\text{total}}^2}\right)$$

Table 5.3: IMI Subscale Scores by Condition and Iteration

Condition	Iteration	Interest/Enjoyment mean	Interest/Enjoyment std	Perceived Competence mean	Perceived Competence std	Effort/Importance mean	Effort/Importance std	Pressure/Tension mean	Pressure/Tension std
AUDIO	First	4.17	1.19	3.13	1.32	4.70	1.07	3.40	1.10
AUDIO	Second	4.54	1.03	3.95	1.77	4.69	0.90	3.00	1.23
FL	First	4.97	1.45	3.35	1.24	5.58	1.14	3.60	1.60
FL	Second	5.49	0.94	4.60	1.56	5.55	0.93	3.27	1.38
GAMIFICATION	First	4.09	1.17	2.50	1.12	3.67	1.40	3.71	1.50
GAMIFICATION	Second	4.34	1.31	2.71	1.34	4.36	1.62	2.66	1.14
VG	First	3.58	1.40	1.85	0.68	3.82	1.09	4.15	1.43
VG	Second	3.51	1.42	1.90	0.83	3.80	1.72	3.70	1.36
VIDEO	First	3.82	1.29	2.33	0.81	4.75	0.92	3.23	1.21
VIDEO	Second	3.87	1.23	2.89	1.18	4.63	1.36	3.42	1.15

Where:

- α: Cronbach's Alpha
- N: Number of items in the scale
- σ_i^2: Variance of the scores for the i-th item
- σ_{total}^2: Variance of the total scores (sum of all items)

5.3 Test Scoring

The analysis conducted provides a comprehensive evaluation of the test data across five learning modes: AUDIO, FL, GAMIFICATION, VG, and VIDEO. In total there was five tests, on each test descriptive statistics was performed, normality testing using the Shapiro-Wilk test, and Kruskal-Wallis tests to evaluate differences across learning modes. For tests with significant Kruskal-Wallis results, pairwise comparisons were conducted using Dunn's post-hoc test to identify specific group differences.

The normality test results reveal that most datasets deviate from a normal distribution (p-value < 0.05), particularly for the AUDIO and VIDEO conditions. Tests such as test1, test3, and test5 showed non-normal distributions across most groups, indicating that the use of non-parametric tests like Kruskal-Wallis is appropriate.

These findings suggest that the distribution of scores varies considerably between tests and learning modes, with FL and VG demonstrating more normal-like distributions compared to the highly skewed AUDIO and VIDEO learning modes.

The Kruskal-Wallis test revealed statistically significant differences among learning modes for four out of five tests:

- test1 ($p = 0.0126$): Demonstrates significant differences across learning modes, particularly between VG and VIDEO as identified in the post-hoc analysis.

- test3 (p = 0.0013), test5 (p = 0.0027), and test6 (p = 0.0020): Show strong evidence of differences among learning modes, especially involving the FL group.
- test2 (p = 0.0740): Did not exhibit significant differences among learning modes, indicating a more uniform performance across groups, with balanced improvements in the first day of learning, the early stages of learning.

For the tests with significant Kruskal-Wallis results, Dunn's post-hoc test provided deeper insights into pairwise differences:

- test1: A significant difference was observed between VG and VIDEO (p = 0.0074), suggesting that these two groups have distinct performance outcomes, even though the participants had the exact same starting point and mode of learning on the first day, learning using video.
- test3: Significant differences were identified for FL compared to both GAMIFICATION (p = 0.0093) and VG (p = 0.0034), indicating that FL consistently outperformed these groups. FL shows as an excellent mode of learning for mid-phase learning, while modes involving gamified learning start declining in progress.
- test4: Similar to test4, the FL learning mode was significantly different from GAMIFICATION (p = 0.0074) and VG (p = 0.0080), reinforcing the trend observed in test4.
- test5: The FL learning mode again demonstrated significant differences compared to GAMIFICATION (p = 0.0050) and VG (p = 0.0031).

These results highlight the consistent outperforming of participants from the FL learning mode, particularly in later stage of learning, where participants were all already familiar with the task and the initial introduction phase ended, and the real learning starts. This shows that modes with gamified learning were good for initial stages for engaging the users and letting them become familiar with the skill, but not to obtain and retain the skill. In contrast to what other forms show, even the unimodal VIDEO condition the learning started on the second day, with different speed of improvement, but still a reliable trend of learning.

The FL learning mode appears as the highest-performing group, with the highest mean scores for test3, test4, and test5. This indicates that participants in the FL learning mode consistently exhibited better performance, likely due to particular features of this mode that help with skill acquisition or improve performance. However, the variability (std) in FL scores is also high, indicating a wider range of participant outcomes.

In contrast, the VIDEO learning mode often displayed the lowest mean scores with lower variability. This suggests that the VIDEO learning mode may not have been as effective at improving performance compared to other learning modes, but participant outcomes were more consistent. The AUDIO learning mode, while showing moderate mean scores, had high variability across all tests, indicating that its effectiveness was inconsistent.

Nevertheless, VIDEO learning mode constant and steady rise, even though slow, while gamified learning showed decrease on the second day. When it comes to comparing the learning modes for each test, while using the pairwise comparison from Dunn's Post-Hoc Tests and descriptive statistics the analysis is as follows. In the early learning phase, which includes the first day, the first and second test, significant differences were noticed in the first test, however in the second test there were no significant differences in neither of the learning modes. In the first test the most significant difference is seen in VIDEO and VG, where VG significantly outperformed. Since both conditions on the first day, for the first and second test use exactly the same technique under exactly the same conditions there is no objective reason to have a significant difference among these two groups. The mean value for VIDEO was by far the lowest, having the value of 3, followed by GAMIFICATION of 4.92, almost 5, AUDIO with 5.45, and the most successful FL of 6.72, and VG of 7. Since the first and second test show mostly the cognitive stage of learning, getting to know the skill and recognizing different gestures, it should only be used as the reference point of how the participants started their learning and how much they improved, however the skill acquisition comes in later stages of learning. Second test, as previously stated, did not show any significant pairwise differences. The trend and ranking changed, where the most successful learning condition was shown to be FL, followed by AUDIO, showing the highest improvement, followed by VG with slight improvement, with GAMIFICATION and VIDEO with the lowest performance, respectively.

On the second day, in the fourth test FL learning mode showed the best performance, outperforming GAMIFICATION and VG significantly, with those two learning modes having the worst results, while there were no significant differences between VIDEO and other conditions. AUDIO took the second place, followed by VIDEO, GAMIFICATION and VG respectively, with a very slight difference between the gamified learning modes. In this mid-learning phase there are changes in trends in the sense that gamified way of learning started significantly underperforming and setting the trend for the future tests. It is shown that learning only through game in GAMIFICATION case, and including the game earlier in the process than it should be, as in case of VG, the skill acquisition significantly stagnates.

Starting the later phase of learning in test five the trends set in the mid-phase did not change. From fourth to fifth test the ranking stayed the same, however the GAMIFICATION and VG learning modes showed very little progress, and significantly less than all other. VIDEO showed slow but steady and noticeable progress. In the last test, pairwise comparison showed the same significant differences as in the previous tests, with FL significantly outperforming GAMIFICATION and VG. The ranking stayed the same for all modes of learning, however noticeable improvement can be seen in FL learning condition, while AUDIO showed little progress from the previous test, as well as the rest of the learning modes. In the last test, for comparison purposes, mean result in FL learning mode was 24.45, meaning that the participants learned all gestures, the entire choreography, however some were maybe not aware that the pattern is reoccurring, AUDIO with noticeable worse result but still good performance result of 16.45, followed by VIDEO having the mean of 12.11, dropping to 7.92 GAMIFICATION and 7.63 in VG. This all can be seen in Figure 7 showing results for test1, and Figure 8 showing results for test2, Figure 9 depicting results for test3, Figure 10 doing so for test4, and last but not least for test5 showing Figure 11, for all figrues see Appendix B. For these results see Table 5.4.

Summarizing test scores, FL learning mode demonstrated the highest score, being the most effective learning mode, significantly outperforming GAMIFICATION and VG from the fourth test onward, in the second day. VIDEO, even though showing poor initial performance, it has show stable and constant improvement, outperforming GAMIFICATION and VG in later stages. AUDIO on average showed good performance, being more efficient than VIDEO, however, depending on the participant's preferences showed high variability, hence FL confirming it's highest stability and effectiveness in the learning process, while still being the mode which includes and is consistent of multiple types of feedback.

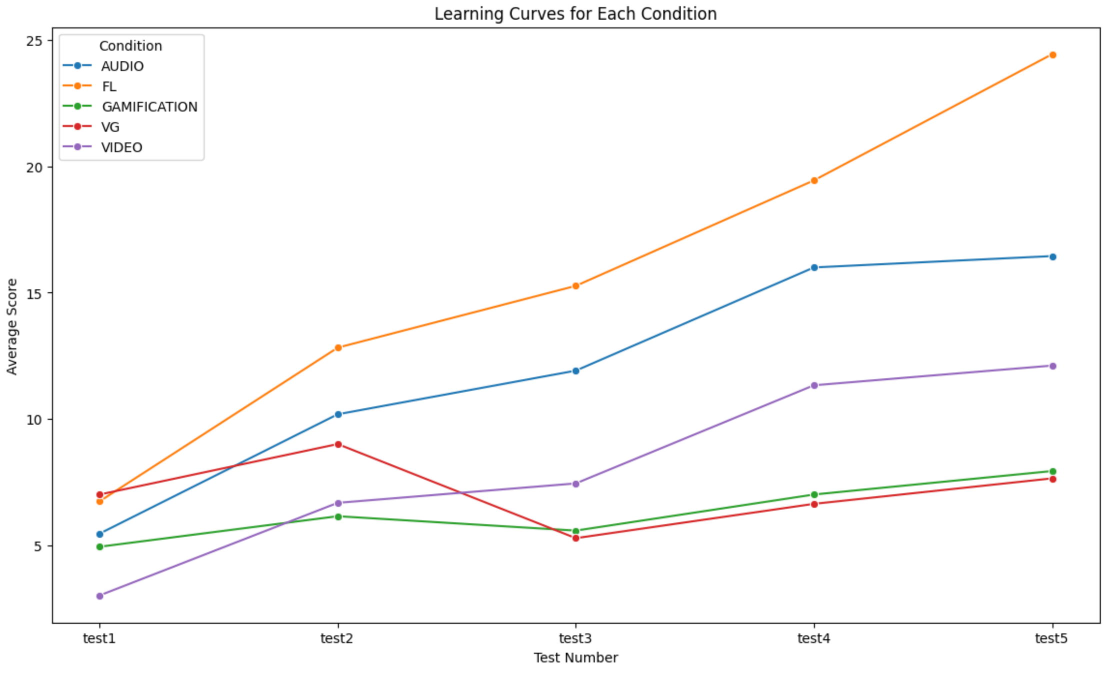
GAMIFICATION and VG proved to provide short-term benefits, initial boost of learning, however decline on the second day, as showing that being the only stand alone mode of learning is not enough for an effective skill acquisition.

In short, describing the learning curve in early stages including first and second test only small differences were shown between different learning modes, with VG, offering only video at that time, performing best. In mid-stage learning, fourth test FL takes the lead, while GAMIFICATION and VG start declining, while in later stages including test5 and test6 FL absolutely dominates all other modes of learning, while VIDEO definitely surpassed GAMIFICATION and VG. For learning curve progression see Figure 12.

Gamified way of learning showed to be useful for early engagement, however it fails to support long-term skill retention, including GAMIFICATION showing this trend, while VG, or more precisely, inclusion of gamified learning in the mid-phase of learning showed not to be effective, with learning declining. This could indicate that gamified learning could be useful for engaging the participants early on, to be interested in the topic, however in order for it to have positive and significant impact on the learning it should be included perhaps even later, in later phases of learning, for skill refinement. VIDEO showed steady improvement, suggesting long-term adaptability even though it had a weak start, proving to be steady and consistent but reliable mode of learning, while FL as the mixture of visual and auditory learning provides reliable improvement over time, being more stable than AUDIO, since participants had the opportunity to choose what type of visual feedback they need (static, dynamic) and at what stage, while AUDIO had predefined scenario. Nevertheless, AUDIO as the multimodal way of learning, with a predefined scenario showed the best results, as previous studies showed as well.

Table 5.4: Test Scores: Mean Descriptive Statistics Per Group and Test

Condition	**Test1 Mean**	**Test2 Mean**	**Test3 Mean**	**Test4 Mean**	**Test5 Mean**
AUDIO	5.45	10.18	11.91	16.00	16.45
FL	6.73	12.82	15.27	19.45	24.45
GAMIFICATION	4.93	6.14	5.57	7.00	7.93
VG	7.00	9.00	5.27	6.64	7.64
VIDEO	3.00	6.67	7.44	11.33	12.11



(a) Learning Curve: Results of Each Condition For Each Test

5.4 Interviews

All interviews have been transcribed using already mentioned DaVinci Resolve, and in some cases if the language of the interview was not English, using ChatGPT. Formatting of videos was done next, in a intelligent verbatim, where all stuttering or other noises were filtered out, and the conversation is divided into “Interviewer” and “Interviewee” for the sake of easy following and understanding.
In order to analyze the interviews, thematic analysis is applied, where the codes were

made based on focusing on answering the research questions. Interviews were coded, grouped by each learning condition and analyzed. Each learning condition was analyzed as a group to draw condition-based conclusions, afterwards relating the information among the conditions and with the quantitative data results.

Six themes emerged from the interviews, with nine codes in total. Themes are as follows:

- PREFERRED MODE OF LEARNING: answering if this mode and what it offers was helpful to the participant to learn, or would they prefer some other way of learning, e.g. audio cues, or game.
- CHALLENGES IN LEARNING: did the participant have difficulties in learning with the given mode of learning
- LEARNING CURVE IMPROVEMENT: answering if the learning gradual or sudden at a certain point of learning
- PERCEIVED EFFORT AND ENJOYMENT: did the participant enjoy the learning process, or associate positive feelings with the learning of the skill or not
- PERCEIVED EFFORT: did the participant invest too much effort, was it physically and/or mentally demanding, was it negatively influencing the experience, or was it a reasonable amount of effort that was motivating.
- PERCEIVED PERFORMANCE: was the participant satisfied with the performance or did they think they performed poorly.

  For the codes relating to the themes, see Table 5.5.

Table 5.5: Themes and Codes

Theme	**Codes**
Preferred Mode of Learning	**C1**: Keep the given mode
	C2: Change the mode
Challenges in Learning	**C3**: Difficulty learning, didn't overcome
Learning Curve Improvement	**C4**: Gradual
	C5: Sudden
Perceived Effort and Enjoyment	**C6**: Enjoyed
	C7: Neutral enjoyment
	C8: Did not enjoy
Perceived Effort	**C9**: Too much effort
	C10: Positive effort

Continued on next page

Table 5.5: Themes and Codes (Continued)

Theme	Codes
Perceived Performance	**C11**: Bad performance **C12**: Good performance

Each of the learning modes will be analyzed and later related and compared.

The VIDEO learning mode offered participants a room for structured and methodical learning, repetitive and monotonous, with no additional engagement, cues or guides. Hence, even though some were satsified, most expressed desire to learn with another mode of learning or additional tools such as “Having the gestures written on a sheet of paper in order would help me memorize them more easily.”, or “Using audio to get the speed right could help.”, or “Maybe visual feedback, like knowing when I did a gesture correctly or at the correct time.”, and even more precisley expressing their interest in gamified mode of learning “I think gamification, like a game similar to “Just Dance”, would help.”. Participants tried to find a pattern, with the main learning strategy being visual pattern recognition in order to memorize gestures and sequences. Many participants described their learning as gradual and steady. On the first day, the slower pace allowed the participants to focus properly, follow along and feel the motor skill on their own by following along, offering the opportunity to recognise and remember different gestures. However, most of the participants struggled with finding the pattern of the motor skill, but did manage to remember most of the different gestures. On the second day, most of the participants remembered a part of the sequence as well, usually a third of it, but there were those learning more, with one participant learning the entire choreography. Strategies that they used mostly involved how many fingers were up or down, trying to map the counting into some sort of a pattern, however that strategy did not yield results often “The fingers weren’t very helpful because there are several movements you can do with three fingers.”
However, the lack of interactive elements or immediate feedback limited engagement. Participants often expressed frustration when they couldn’t independently assess their performance or identify errors, especially during the test “During the test, you’re on your own, and if you’re blocked, you’re stuck.” . They expected to have some type of feedback if they were doing the skill correctly, since during the tests that was the only possibility for them to see themselves in the screen and try to perform the skill on their own. Many found memorizing the skill challenging, especially with long sequences, as participants had difficulty remembering movements without corrective feedback “I learned about the first half, but the rest was harder.”. The learning was gradual for most of the participants, and the difficulty was reasonable, however, even though not regarded as negative, the effort was present, both mentally and physically. On the second day, with faster speed and begin familiar with the task the participants reported increased

enjoyment “The faster pace made it more intuitive and pleasant.”. As they learned they also felt more competent and enjoyed the learning more. Despite these challenges, the VIDEO learning mode provided a strong foundation for gesture recognition, with many participants achieving incremental, steady progress. This method was a good way of learning for those who prefer more structure and repetition, but failed to help much those who seek engagement and dynamic elements.

The GAMIFICATION mode of learning offered an engaging and interactive approach that appealed and captivated some participants but introduced significant challenges for other. Real-time feedback mechanisms, rhythm synchronization, and gamified elements rating participant performance(correct or incorrect) motivated many. Participants described moments of good focus while they were adapting to the game's pace and feedback, which for some presented a lot of effort, while for some higher engagement and interest “I think a game is a good solution to learning in general. It grabs attention more.”.
However, since gamified learning often feels as if it is fast-paced, and requires quick adaptation, it also overwhelmed participants. Certain participants had difficulties adjusting to the system, synchronizing their gestures to the timing of the game, being too fast or too slow. Additionally, the system was not adjusted to have high tolerance, hence once a participant made or started making a wrong move but corrected themselves the system registered it as a mistake, bringing frustration to a rise in the participant. Concentrating on, and trying to get the speed right sometimes shifted focus away from practice, leading to errors and reduced confidence, which further led to feelings of frustration and incompetence “The tutorial kept rejecting my movements… it made it more complicated.”. Participants who successfully adapted their strategies, such as synchronizing gestures with beats or focusing on rhythm, reported noticeable improvements. However, those unable to overcome these challenges described the mode as mentally and physically demanding, indicating the need for adjustable pacing. Nevertheless, many participants experienced positive game feedback on the second day when the pace was in normal speed and more intuitive “Seeing faster results with the faster pace made it more enjoyable.”. Even though some of them adjusted better to the game it did not mean that they learned necessarily but that they rather focused on succeeding in the game. The learning was happening gradual for most of participants. Gamified mode of learning turned out to be a dividing mode, some participants found it to be engaging and challenging which kept them interested in the process, stating “It was fun. It was faster, so it was something new.”, however some stated that they would use different modes of learning such as “I'd rather just have some graphs and descriptions and then some examples.”. Many of them reported that they would use a game but later in the learning process for practicing and not for the initial learning. Even in the gamified system the participants tried finding the same strategies to learn as in the other modes, segmenting the choreography, counting the fingers that were put up or down, or mapping it to audio cues “Assigning numbers to the gestures would make it easier to remember sequences.”. Overall, most of participants found the gamified mode to be engaging, interesting and fun, which kept them motivated

and wanting to learn more, however there were also many that reported increased physical and mental effort and frustration. Participants focused more on succeeding in the game and not on learning the task, and reported that for initial learning they would use some other methods and would include gamification in later stages once they need only practice.

As VG incorporated both VIDEO and GAMIFICATION, on day one and day two respectively, the thoughts and feelings were corresponding to the participants in VIDEO and GAMIFICATION. On the first day, while they were learning with video only, they felt as if the task was sometimes monotonous but that it gave them ability to understand the task. They expressed the need for real-time feedback and not to rely only on themselves and their assumptions in such early stages of learning. Some expressed struggles and dissatisfaction with their performance since they didn't succeed in remembering much of the task. On the second day, they were more engaged, and even though they might not use the gamified way or learning for that phase, they still were interested and trying "The game made the process more relaxed for me compared to just watching the video.". However, those who did not initially have a good start on day one, felt even more frustrated on the second day. What is noticeable is that the satisfaction with the mode of learning is directly tied to the participant's perceived performance. Even though some participants did not have a good performance on the test, they felt in the game that they were doing good and reported that they felt relaxed, engaged and interested. More than half expressed that they would initially learned the skill differently but that they would learn and practice to perfect the skill later in the process, as in this phase they only focused to succeed in the game.

AUDIO mode of learning added integrated auditory instructions into the learning practice. Hearing the audio instruction on how to move the fingers, combined with visual ones offered great help to the participants to time and synch the gestures with rhythm, enabling deeper connection to the sequence and remembering "The audio instructions were helpful. They weren't too much because I was already used to the task." and "With audio instructions. I think it's easier to recall visually and with the voice at the same time.". Participants often mentioned how useful they believe audio instructions would be in other groups, mentioning that it would serve as guidance, however once offered, auditory instruction increased the mental load, especially in the beginning, stating "There was a lot of information at the same time. I had to control my hand, listen to the music, watch the video, listen to the audio, and memorize the order—all at once.", but even later once the pace was faster. Some also reported that the input was distracting, especially those who struggle to process multiple inputs at the same time. Majority of participants reported that this mode of learning was something that they liked and was useful "Yes, I enjoyed it. It was interesting and something I've never done before.". Strategies that the participants used were the same as in other modes, trying to section the task, find a pattern and similar. However in this mode of learning, more participants saw the pattern than in other modes. Participants learned gradually mostly, with some stating that the learning happened suddenly in one moment “Once I learned the sequence and realized it repeats, it was pretty self-explanatory.”. Those recognized patterns or aligned their gestures with the rhythm saw a significant progress. Additionally, some reported a decline in learning, or difficulty maintaining the focus especially on the second day, during the fast paced rounds. More than half participants enjoyed the task, due to its novelty and challenge, others were either of neutral reactions noting that even if it was interesting, it was not especially enjoyable, and those who experienced frustration noted that it was tied to their performance and inability to remember the sequence properly. Most of participants found that audio instructions were helpful, very few noted that it was distracting and overwhelming in the fast paced rounds of learning. More than half of participants felt competent that they mastered either the entire choreography or substantial part of it.

The Free Learning, FL, mode offered participants the most autonomy, allowing them to experiment with various strategies and select their preferred learning methods. Participants were overall satisfied with this mode of learning, being able to choose their own strategy and mode of learning “I liked the freedom of kind of being able to choose what to do.”. Many of participants decided for using audio instructions together with visuals “I switched to having audio in the background, which helped more than just the video.”. What many participants reported that would be helpful is to give then the hints and tips on the pattern of the choreography, however that is not a possibility for learning. What many other participants including some from this group struggled with was the physical effort they needed to put in in order to master the motor skill, they felt the movements were difficult to preform “The unnatural movements in the first gestures slowed me down.”. For most of the participants the learning was gradual, however those who saw the pattern once they found it they had a sudden increase in speed of learning and skill mastering “When I realized the sequence repeated, it became much easier.”. Most of the participants enjoyed the process, they found it interesting, or even calming “It was calming and almost meditative.”. Overall, the participants enjoyed the process and felt more competent than participants in other groups such as gamification. The choices were almost always multimodal, except for some, e.g. one participant chose to learn through writing the notes and figuring out the pattern. Afterwards they practiced with video, listening to the music for test practicing.

CHAPTER 6

Discussion

This research brings more understanding to the learning process of motor skills and how do effort and enjoyment influence it. More precisely, what mode of learning, whether it is unimodal, multimodal or gamified mode will have the fastest learning, and what mode offers least effort and most enjoyment, and if these two factors influence the speed of learning. The results of this study suggest that multimodal learning without gamification inclusion yields the best results, the learning happens the fastest and the enjoyment and effort are balanced. This section goes into detailed explanation of the findings.

For answering the first research question of what type of feedback, whether that is audio, visual, or gamified, has the most positive impact on motor skill acquisition, most important data is the quantitative analysis of the test scores. What can be seen is that multimodal modes yielded the best results, namely FL and AUDIO, respectively. This means that the most effective mode of learning was the one combining the video and auditory instruction, including that some of the participants from FL learning mode in the beginning also used image maps before moving to the video as well. VIDEO mode of learning came third with the moderate pace of learning, moderate effort and enjoyment. This traditional mode of learning proved itself to be effective and a reliable method to yield results, however the combined feedback proved to be more efficient, from the beginning. Gamified modes of learning showed that early inclusion did not offer good results. It kept participants engaged but also required increased both physical and mental effort, where participants focused on satisfying the game results instead of focusing on learning and remembering the task. This as a result, once when faced with the test, it showed that participants did not actually learn the skill, and that their subjective impression of knowledge turned out to be worse in reality, evoking feelings of incompetence and poor performance. This feeling made participants feel frustrated and less satisfied than in other modes. However, most of them reported that they were engaged and focused, and that they would use the game, however later in the learning process, once they practice to improve the skill, and not while initially learning it. The difference between GAMIFICATION and VG was that those learning with the game from day one, were not as surprised and later disappointed with their performance as those in VG. Participants in VG on the first day had far better performance learning in the traditional way with only the video instructions, and falling behind on the second day even behind GAMIFICATION, since the pace was faster and they were faced with a new mode of learning. The difference between the two modes on the second day was very small, however, those participants learning with the game on both days had slightly better performance than those in VG.

This brings us to answering the second research question of how do learning curves vary in motor memory tasks involving hand gesture choreography with and without gamification. Additionally, what considerations we should make when including the gamification techniques at various stages of the learning process. As previously stated, those learning with the game fell significantly behind than those learning without the game. Even the traditional modes of learning such as with video instructions only outperformed the gamified way of learning. The learning curve is much slower since the focus is on the game and not the learning itself. Participants are used to learning certain things in a traditional way in most of the cases, but do feel enthusiastic and willing to learn with the game in later stages. If the game is included too early in the process, whether that is from the very beginning or including it later, but still too early in the process the progress will decline and be much slower. Gamified way of learning, especially learning through a game should be included in motor skill learning once the skill is learned but needs more practice so that the learners are able to focus on steps already knowing them but trying to do them as better as possible. To summarize and point out, multimodal learning including FL and AUDIO modes lead to the fastest motor skill acquisition, VIDEO or unimodal feedback in this case provides slow but steady progress, while gamified modes GAMIFICATION and VG slow down early learning but are useful in later stages for practice and refinement. The learning curve is slower when gamification is introduced too early, as participants focus on the game rather than skill acquisition. Hence, the gamification should be introduced later in the learning stages, once the learners are ready to refine their skills. Moreover, positive and supportive feedback should be incorporated throughout the learning process. On top of that, learners should have control over how and when gamified elements are introduced. However, what is learned from the interviews, especially in modes which did not include gamification, participants do cherish certain amount of positive or corrective feedback along the traditional modes. This could come in the form of supportive messages or estimation of learning. Nevertheless, one must be very careful when designing such feedback. Participants usually expected some sort of feedback or help during the test itself, or some explicit tip, advice or guidance. Having this in mind, when considering the learning process on its own, perhaps only gamified elements of support to help feeling competent and successful in the process could be meaningful. As seen, the levels of frustration were directly tied to one's perception of performance quality. If participants felt that they did well, or that they did not have high performance pressure they overall felt better, less frustrated and enjoyed the process more. These facts lead us to answering the third research question.

Third research question enables us to understand how do perceived enjoyment and effort vary between different learning modes and how does it relate to the learning curve. As seen from quantitative data, but understood better through qualitative data analysis enjoyment is tied directly to the persons performance perception in most of the cases. Enjoyment was identified in the interviews as something that participants explicitly enjoyed, felt relaxed, found it interesting, fun, engaging. What could be noticed is that heightened effort did not decrease enjoyment, and vice versa. Effort was noticeable in almost all modes in the form of physical effort, and often in mental as well. The subjective reports of increased effort is directly tied to poorer test performance. Looking more closely, mental, physical and temporal perceived effort truly depicted this. The lowest effort was reported among those learning in FL and on the second place learning with AUDIO. The third place was taken by VIDEO, except in temporal category which was taken by GAMIFICATION by a slight difference. It is followed by the gamified modes of learning with VG showing the most increase of effort among all modes of learning. This is due to having a new mode of learning and faster speed of learning on the second day, leading to poorer performance and increased effort of achieving the results and adopting the new strategy. Perception of performance was a realistic reflection of test results. Best performance perception was shown in FL, followed by AUDIO, VIDEO, GAMIFICATION and lastly VG, as the test scores show the test scores as well. This is also reversely matched in frustration. Most frustrated were participants in VG and least in FL group. The same is shown in overall workload. Most workload was reported in VG, and least in FL. Important difference is that overall workload was reported higher in AUDIO group than in VIDEO on the first day, since AUDIO had more inputs and increased mental load due to higher number of inputs. Quantitative data coupled with interviews showed that enjoyment and interest were tied to participants perception of performance, and effort did not decrease the enjoyment. Effort was high in participants who did not find their way in the strategy that was offered, who felt that the motor skill given to learn was very difficult to perform physically, and those who felt overwhelmed or just challenged by the mode of learning, usually expressed in AUDIO and GAMIFICATION mode. However, in VIDEO mode, lack of engagement led to feelings of cumbersome task, pushing people to be less engaged and more difficult to push through the task. What is interesting is that even though participants in GAMIFICATION did not perform well as compared to other modes, participants still often reported that they were engaged and focused on the task, and that felt good since they felt immersed or challenged in a good way. It would be expected that gamification leads to the highest enjoyment and lowest effort, but enjoyment and effort do not directly influence each other. Instead, performance correlated with enjoyment, while effort was often seen as positive, keeping participants engaged and motivated.

These findings correspond to the previous research stated in this thesis. Gamified learning helps the individuals stay interested and motivated to learn [BGRS23], it keeps them engaged, however as Van der Kooij [vdKvDvV+19] found is that even though it can increase enjoyment it does not help much in the early-phase learning, and that it is beneficial for motivation but not for performance as well, it does not help with early process of skill acquisition. Additionally, it can even increase frustration if incorporated too early, and that it should be included in the associative stage, once the learners are ready to improve the skill and not initially learn it. The importance of familiarity with the skill and previously acquired knowledge is also confirmed by Wiemeyer and Schneider [WS12]. The combination of auditory and visual feedback is proven to help learners understand what to perform in the cognitive stage, and how to perform in the later stage of learning [MGS21]. It has also been shown to be crucial for skills where coordination and spatial and temporal accuracy are important, such as different motor skills, as found by Moinuddin et al. [MGS21].

It is important to carefully design the multimodal system. If done poorly the cognitive load can rise especially for new learners and become overwhelming. For this reason FL mode was the most successful one. The learners were able to choose from different modes of learning and to combine them in different stages. Furthermore, they were able to adjust some of the options for their own liking within the same amount of time for learning as other modes. This means that participants who learned with video only or with video and audio instructions did not have necessarily the right system to learn with. Some reported that they felt overwhelmed, would like to pause or section the choreography, see it on the paper laid out. For this reason, FL had a faster learning curve and was overall a more stable learning mode in the sense that there was less variance between the participants reporting their performance and perceptions.

This research shows differences among different modes of learning through comparative analysis of test scores throughout the learning as well as reports on perceived effort and enjoyment. What it brings is the opportunity to analyze traditional unimodal visual mode of learning, combination of visual and auditory making multimodal mode and the gamified mode of learning. It gives a unique insight into how different these modes of learning are in motor skills acquisition and how they behave. This gives us an opportunity to make informed decisions when designing learning systems and strategies. Additionally, this study offers the opportunity to have an insight on how participants learn over the course of two days and not just one. The skill itself is complicated enough that it takes certain amount of effort over the course of two days to learn the skill but also simple enough that the task is manageable to learn in two days. The first day is used to get familiar with the task, and on the second to actually learn and remember the skill. This was done with the intention to understand when the gamification should be included and if there is a difference if the inclusion happened from the beginning or after getting familiar with the task. Last but not least, this study offers the software for the learning environment and insights on how to properly design one. What this study shows is that there is no real unified mode or system to learn with that will suit every person. However, it did prove that even if that is true some modes on average give better results even though they may not be the first choice for the individual, such as learning with both visual and auditory cues. Moreover, it proved that the system should be customizable. The modes that were rigid in design, where participants did not have an option to pause and rewind the video, with and without auditory instructions, or to pause and restart the game, or have different levels of conquering the task, performed worse than the one where participants were able to combine what and when they needed. FL mode also gave a great insight that the participants on average will choose exactly both visual and auditory feedback to learn, from video and image maps to video with auditory instructions. The game in this phase of learning was not chosen to learn with by anyone. Through interviews it was clear that participants were interested in gamified learning but later in the process, which aligns with existing literature.

Limitations of this study were present. One of them was the fact that through interviews it was clear that some of the participants were not initially interested or motivated to do the task since most of them were recruited from a university course. This group of people was predominately present in both GAMIFICATION and VG. This was maybe only an excuse to report since some of them were not satisfied with their performance. Furthermore, the game system should have been more tolerant to mistakes or ambiguity in performance. During the learning some of the participants felt as if they were underperforming or that the system didn't acknowledge their proper performance, leading to frustration. As it was found that enjoyment is related to perception of performance, it is better to have the system falsely accept the move and give positive feedback than the opposite. Nevertheless, not everyone experienced this. The important insight gathered from interviews, is that in the gamified modes of learning participants are more focused to satisfy the system and not to adopt new knowledge, highlighting that the game should be included later in the learning process. Also, participants often felt engaged and wanted to keep going which proved that their motivator was the game itself and not the task.

Summarizing, the motor skill learning systems should be flexible enough to be able to section and adapt the task for the learner, choosing the mode, being able to combine them and to separate the task if needed. Additionally, the system should incorporate positive and supportive gamified feedback and elements without making it the central theme, but rather as an additional feature for boosting confidence only. It is advisable to have a game that is separated by levels once the game is implemented. If the task is sequential, dividing it into smaller sequences that add up through levels, or if not, to show simple steps and elements progressing to more complex ones and combining them together. Throughout the game it is important still to focus on positive and supportive feedback, emphasizing it. The game should be available once the person feels confident enough to start practicing the learned skill. For future work a longer study should be implemented with comparing multimodal mode including visual and auditory feedback, in an adaptable system, extending to a longer period, where in later phase of learning the game would be implemented, again starting from the beginning and building up to mastering the skill through levels, with careful consideration of other gamified components that could be implemented.

CHAPTER 7

Conclusion

This thesis explores learning curves of motor skill acquisition between different modes of learning comparatively analyzing test scores of visual learning, visual and auditory combined and gamified mode of learning. Additionally it explores how do enjoyment and effort relate to and impact the learning, and considers when to include or exclude gamification. In order to achieve this understanding and answer the research questions the motor skill task was carefully designed to be complex enough for a two day study, but simple enough to learn in this time period. Additionally, the software was developed for developing the proper learning and testing environment implementing the mentioned modes of learning and utilizing Google's artificial intelligence algorithms for recognizing different hand gestures and rating the participants performance during the game and testing modes. Afterwards user study was designed and implemented where participants were divided into five groups, first being VIDEO where they learned only with visual instructions through video, AUDIO mode of learning utilizing the same video with auditory instructions played over it, GAMIFICATION where the users played a game where they got feedback if the performance is correct or wrong, VG where they on the first day learned with video only and on the second with game only, and FL offering the participants to choose any of the modes with additional options. The study utilized standardized questionnaires for evaluating perceived effort, enjoyment and performance namely NASA-TLX and IMI, additionally to the testing of the skill performance, conducting choreography tests. These quantitative methodologies were supported by interviews. The two day study was conducted at university premises with 56 participants.

The findings show that multimodal learning (FL, AUDIO) was the most effective, while gamified learning was least effective in early stages but valuable later for refining skills. It was found that learning through a game should be offered later once the learner is confident in their knowledge and wants to practice the skill to become better at it.
The system for learning should be adjustable and customizable, with least distracting

elements. It should provide enough structure in the context of time, as well as supportive gamified feedback in the learning process in order to increase perceived performance hence enjoyment, and decreasing negative perception of effort. It should additionally offer certain amount of flexibility in choosing and switching learning modes, as well as to rewind or in other means, section the task.

Future research should explore long-term learning effects, optimize gamification timing, and investigate adaptive feedback mechanisms to enhance engagement and performance. Additionally, the study should be conducted over a longer period of time so that the more detailed insights can be made. Gamified mode of learning, in this case, should be included in later phases of learning evaluating its potential and effectiveness, if and how it positively influences the learning, or not. Additionally, fostering enjoyment should be included in future work, bearing in mind that it is increased through supportive feedback, fostering the feeling and perception of good performance and competence. In this sense, the proper inclusion of supportive gamified feedback during the entire learning process should be investigated, so that the individual still has the realistic insight into their performance but also encouraging.

Overview of Generative AI Tools Used

In this work I have used AI Tools such as ChatGPT and Quillbot for refining the text, paraphrasing and similar. I have afterwords paraphrased it additionally. I have used DaVinci Resolve and it's AI Tool for transcribing the interviews, and for that purpose I have also used ChatGPT. Additionally, I used ChatGPT for gaining initial understanding of certain terms and topics, after which I read and wrote further on my own. Whenever I used these AI Tools, I reworked it, made my own concepts, learned and wrote deeper into it, except when I paraphrased the sentences or made the proper formatting of interviews and references. Last but not least I used ChatGPT to translate required text from English to German, such as this statement and Abstract.

Übersicht verwendeter Hilfsmittel

In dieser Arbeit habe ich KI-Tools wie ChatGPT und Quillbot zur Verfeinerung des Textes, zum Paraphrasieren und Ähnlichem verwendet. Anschließend habe ich die paraphrasierten Texte zusätzlich überarbeitet. Für die Transkription der Interviews habe ich DaVinci Resolve und dessen KI-Tool genutzt, ebenso wie ChatGPT zu diesem Zweck. Darüber hinaus habe ich ChatGPT verwendet, um mir ein erstes Verständnis bestimmter Begriffe und Themen anzueignen, bevor ich selbst weiter recherchierte und schrieb. Wann immer ich diese KI-Tools genutzt habe, habe ich die Inhalte überarbeitet, eigene Konzepte entwickelt, mich tiefer mit den Themen auseinandergesetzt und selbst geschrieben – mit Ausnahme des reinen Paraphrasierens von Sätzen oder der korrekten Formatierung von Interviews und Referenzen. Nicht zuletzt habe ich ChatGPT genutzt, um erforderliche Texte wie diese Erklärung und das Abstract von Englisch ins Deutsche zu übersetzen.

List of Figures

List of Tables

Appendix A: Tables

Table 1: NASA-TLX: Normality Test Results for Different Conditions and Iterations

Condition	Iteration	Statistic	P-Value	Normality Status
VIDEO	First	0.726	0.0029	Not Normal
VIDEO	Second	0.870	0.1244	Normal
VG	First	0.946	0.5901	Normal
VG	Second	0.946	0.5953	Normal
GAMIFICATION	First	0.891	0.0828	Normal
GAMIFICATION	Second	0.855	0.0260	Not Normal
AUDIO	First	0.928	0.3949	Normal
AUDIO	Second	0.942	0.5393	Normal
FL	First	0.945	0.5858	Normal
FL	Second	0.930	0.4107	Normal

Table 2: NASA-TLX: Wilcoxon Signed-Rank Test Results for Within-Group Comparison (First vs. Second Iteration)

Condition	Test	Statistic	P-Value	Interpretation
VIDEO	Wilcoxon	7.0	0.237	No significant difference between iterations.
VG	Wilcoxon	32.5	1.0	No significant difference.
GAMIFICATION	Wilcoxon	32.0	0.345	No significant difference.
AUDIO	Wilcoxon	27.0	0.959	No significant difference.
FL	Wilcoxon	16.0	0.239	No significant difference.

Table 10: Test Scores: Normality Test Results (Shapiro-Wilk)

Condition	Test	Statistic	p-value	Normal
AUDIO	test1	0.830841	0.023947	False
AUDIO	test2	0.737018	0.001408	False
AUDIO	test3	0.782650	0.005563	False
AUDIO	test4	0.768017	0.003575	False
AUDIO	test5	0.777868	0.004814	False
FL	test1	0.811236	0.013224	False
FL	test2	0.862517	0.062072	True
FL	test3	0.904389	0.208982	True
FL	test4	0.856035	0.051146	True
FL	test5	0.823976	0.019454	False
GAMIFICATION	test1	0.890025	0.080898	True
GAMIFICATION	test2	0.771797	0.002283	False
GAMIFICATION	test4	0.872056	0.044849	False

Continued on next page

Table 10: Normality Test Results (Shapiro-Wilk)

Condition	Test	Statistic	p-value	Normal
GAMIFICATION	test5	0.918500	0.209225	True
GAMIFICATION	test6	0.953081	0.609486	True
VG	test1	0.917083	0.295090	True
VG	test2	0.853185	0.046959	False
VG	test3	0.891387	0.144720	True
VG	test4	0.921389	0.330383	True
VG	test5	0.957094	0.734991	True
VIDEO	test1	0.728212	0.003027	False
VIDEO	test2	0.957311	0.769801	True
VIDEO	test3	0.873915	0.135381	True
VIDEO	test4	0.570889	0.000044	False
VIDEO	test5	0.641995	0.000297	False

Table 11: Test Scores: Kruskal-Wallis Test Results

Test	Statistic	p-value	Significant
test1	12.736271	0.012639	True
test2	8.529355	0.074002	False
test3	17.861137	0.001314	True
test4	16.258417	0.002691	True
test5	16.888680	0.002032	True

Table 12: Test Scores: Post-hoc Dunn Test Results for Test1

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
AUDIO	1.000000	1.000000	1.000000	1.000000	0.158790
FL	1.000000	1.000000	1.000000	1.000000	0.059658
GAMIFICATION	1.000000	1.000000	1.000000	1.000000	0.342087
VG	1.000000	1.000000	1.000000	1.000000	0.007368
VIDEO	0.158790	0.059658	0.342087	0.007368	1.000000

Table 13: Test Scores: Post-hoc Dunn Test Results for Test2

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
AUDIO	1.000000	1.000000	1.000000	1.000000	1.000000

Continued on next page

Table 13: Post-hoc Dunn Test Results for Test2

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
FL	1.000000	1.000000	0.060728	1.000000	0.480178
GAMIFICATION	1.000000	0.060728	1.000000	0.988871	1.000000
VG	1.000000	1.000000	0.988871	1.000000	1.000000
VIDEO	1.000000	0.480178	1.000000	1.000000	1.000000

Table 14: Test Scores: Post-hoc Dunn Test Results for Test3

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
AUDIO	1.000000	1.000000	0.361738	0.150078	1.000000
FL	1.000000	1.000000	0.009250	0.003403	0.231260
GAMIFICATION	0.361738	0.009250	1.000000	1.000000	1.000000
VG	0.150078	0.003403	1.000000	1.000000	1.000000
VIDEO	1.000000	0.231260	1.000000	1.000000	1.000000

Table 15: Test Scores: Post-hoc Dunn Test Results for Test4

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
AUDIO	1.000000	1.000000	0.470505	0.411952	1.000000
FL	1.000000	1.000000	0.007441	0.008020	0.875612
GAMIFICATION	0.470505	0.007441	1.000000	1.000000	1.000000
VG	0.411952	0.008020	1.000000	1.000000	1.000000
VIDEO	1.000000	0.875612	1.000000	1.000000	1.000000

Table 16: Test Scores: Post-hoc Dunn Test Results for Test5

index	AUDIO	FL	GAMIFICATION	VG	VIDEO
AUDIO	1.000000	0.647486	1.000000	0.781467	1.000000
FL	0.647486	1.000000	0.005004	0.003079	0.301650
GAMIFICATION	1.000000	0.005004	1.000000	1.000000	1.000000
VG	0.781467	0.003079	1.000000	1.000000	1.000000
VIDEO	1.000000	0.301650	1.000000	1.000000	1.000000

Table 3: NASA-TLX: Pairwise Mann-Whitney U Test Results with Bonferroni Correction for Second Iteration

Condition 1	Condition 2	Statistic	P-Value	Significant?
VIDEO	GAMIFICATION	45.0000	0.2680	No
VIDEO	VG	27.0000	0.0936	No
VIDEO	AUDIO	44.5000	0.7312	No
VIDEO	FL	64.5000	0.2694	No
GAMIFICATION	VG	60.5000	0.3794	No
GAMIFICATION	AUDIO	94.5000	0.3495	No
GAMIFICATION	FL	114.0000	0.0450	No
VG	AUDIO	88.0000	0.0747	No
VG	FL	104.0000	0.0047	Yes
AUDIO	FL	87.0000	0.0871	No

Table 4: NASA-TLX: Group-Level TLX Scores by Condition and Iteration

Condition	**Iteration**	**Mean**	**Std. Dev.**
AUDIO	First	3.83	0.68
AUDIO	Second	3.88	0.63
FL	First	3.53	1.01
FL	Second	3.27	0.87
GAMIFICATION	First	3.86	0.92
GAMIFICATION	Second	4.06	1.03
VG	First	4.21	0.97
VG	Second	4.45	0.77
VIDEO	First	3.57	0.73
VIDEO	Second	3.91	1.13

Table 5: IMI: Shapiro-Wilk Normality Test: Subscale Scores by Condition and Iteration

Subscale	**Condition**	**Iteration**	**Statistic**	**P-Value**	**Normal**
Interest/Enjoyment	AUDIO	First	0.879073	0.101263	True
Interest/Enjoyment	AUDIO	Second	0.926640	0.345851	True
Interest/Enjoyment	FL	First	0.963554	0.815100	True
Interest/Enjoyment	FL	Second	0.917279	0.296623	True
Interest/Enjoyment	GAMIFICATION	First	0.947328	0.520041	True
Interest/Enjoyment	GAMIFICATION	Second	0.893866	0.091915	True
Interest/Enjoyment	VG	First	0.957265	0.737166	True
Interest/Enjoyment	VG	Second	0.949854	0.666777	True
Interest/Enjoyment	VIDEO	First	0.878888	0.183770	True
Interest/Enjoyment	VIDEO	Second	0.880391	0.158514	True
Perceived Competence	AUDIO	First	0.926866	0.380026	True
Perceived Competence	AUDIO	Second	0.856606	0.044339	False
Perceived Competence	FL	First	0.971891	0.904955	True
Perceived Competence	FL	Second	0.901572	0.193190	True
Perceived Competence	GAMIFICATION	First	0.900682	0.115383	True
Perceived Competence	GAMIFICATION	Second	0.842805	0.017762	False
Perceived Competence	VG	First	0.935703	0.471368	True
Perceived Competence	VG	Second	0.794309	0.012363	False
Perceived Competence	VIDEO	First	0.960189	0.811899	True
Perceived Competence	VIDEO	Second	0.953991	0.733826	True
Effort/Importance	AUDIO	First	0.926310	0.374741	True
Effort/Importance	AUDIO	Second	0.763602	0.003720	False
Effort/Importance	FL	First	0.892243	0.148314	True
Effort/Importance	FL	Second	0.905384	0.214831	True
Effort/Importance	GAMIFICATION	First	0.942390	0.449806	True
Effort/Importance	GAMIFICATION	Second	0.895489	0.097021	True
Effort/Importance	VG	First	0.924715	0.359882	True
Effort/Importance	VG	Second	0.958663	0.770529	True
Effort/Importance	VIDEO	First	0.717601	0.003557	False
Effort/Importance	VIDEO	Second	0.948614	0.674851	True
Pressure/Tension	AUDIO	First	0.962100	0.797586	True
Pressure/Tension	AUDIO	Second	0.928355	0.363004	True
Pressure/Tension	FL	First	0.924466	0.357607	True
Pressure/Tension	FL	Second	0.861409	0.060055	True
Pressure/Tension	GAMIFICATION	First	0.875250	0.049757	False
Pressure/Tension	GAMIFICATION	Second	0.881684	0.061418	True
Pressure/Tension	VG	First	0.908665	0.235153	True
Pressure/Tension	VG	Second	0.935743	0.506685	True
Pressure/Tension	VIDEO	First	0.992115	0.997693	True
Pressure/Tension	VIDEO	Second	0.908241	0.303760	True

Table 6: IMI: Kruskal-Wallis Test: Subscale Test Results by Iteration

Subscale	Iteration	Statistic	P-Value	Significant
Interest/Enjoyment	First	5.181501	0.269176	False
Interest/Enjoyment	Second	11.759292	0.019234	True
Perceived Competence	First	10.480854	0.033062	True
Perceived Competence	Second	18.136122	0.001161	True
Effort/Importance	First	15.275021	0.004164	True
Effort/Importance	Second	8.131435	0.086881	False
Pressure/Tension	First	2.092016	0.718839	False
Pressure/Tension	Second	4.713622	0.317963	False

Table 7: IMI: Mann-Whitney U Test with Bonferroni Correction: Significant Pairwise Comparisons

Subscale	Iteration	Comparison	Corrected P-Value	Significant
Perceived Competence	Second	(GAMIFICATION, FL)	0.0297	True
Perceived Competence	Second	(VG, FL)	0.0131	True
Effort/Importance	First	(VG, FL)	0.0351	True
Effort/Importance	First	(GAMIFICATION, FL)	0.0249	True

Table 8: IMI: Wilcoxon Test Results for Iterations

Condition	Subscale	Statistic	P-Value	Significant
GAMIFICATION	Interest/Enjoyment	39.5	0.6748	False
GAMIFICATION	Perceived Competence	40.5	0.4631	False
GAMIFICATION	Effort/Importance	28.0	0.2196	False
GAMIFICATION	Pressure/Tension	24.0	0.0785	False
FL	Interest/Enjoyment	10.0	0.0739	False
FL	Perceived Competence	3.0	0.0125	True
FL	Effort/Importance	13.0	0.8653	False
FL	Pressure/Tension	13.5	0.1530	False

Table 9: IMI: Cronbach's Alpha for IMI Subscales

Subscale	Cronbach's Alpha
Interest/Enjoyment	1.02
Perceived Competence	0.97
Effort/Importance	0.96
Pressure/Tension	0.91

Appendix B: Figures

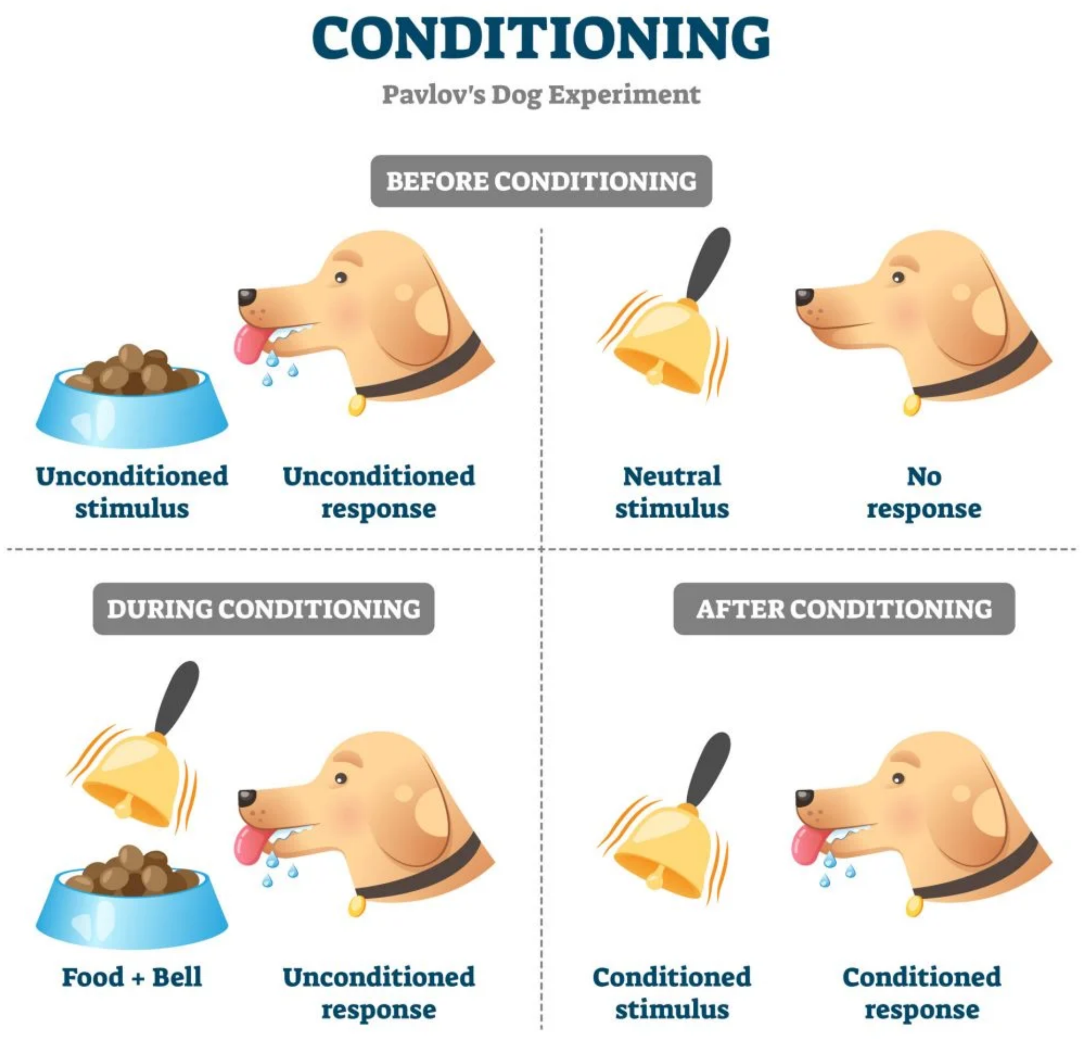


Figure 1: Pavlov's experiment description[Mcl23]

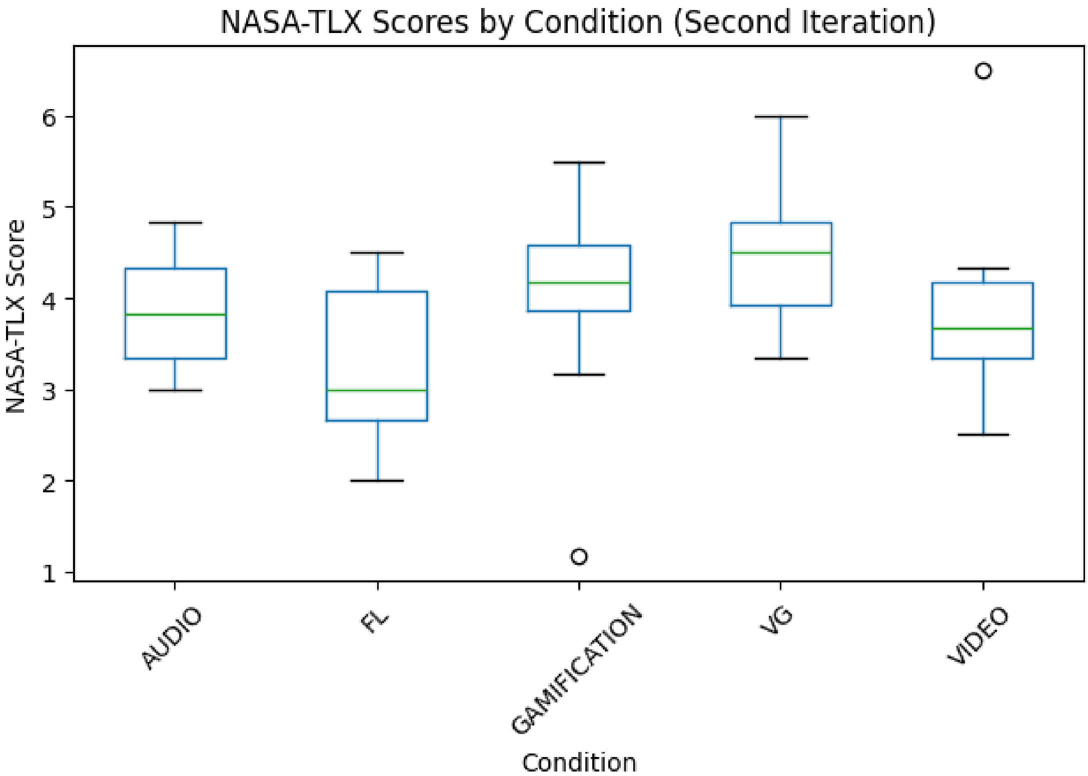


Figure 2: NASA-TLX: Graphical Representation of Pairwise Comparison Results for Second Iteration

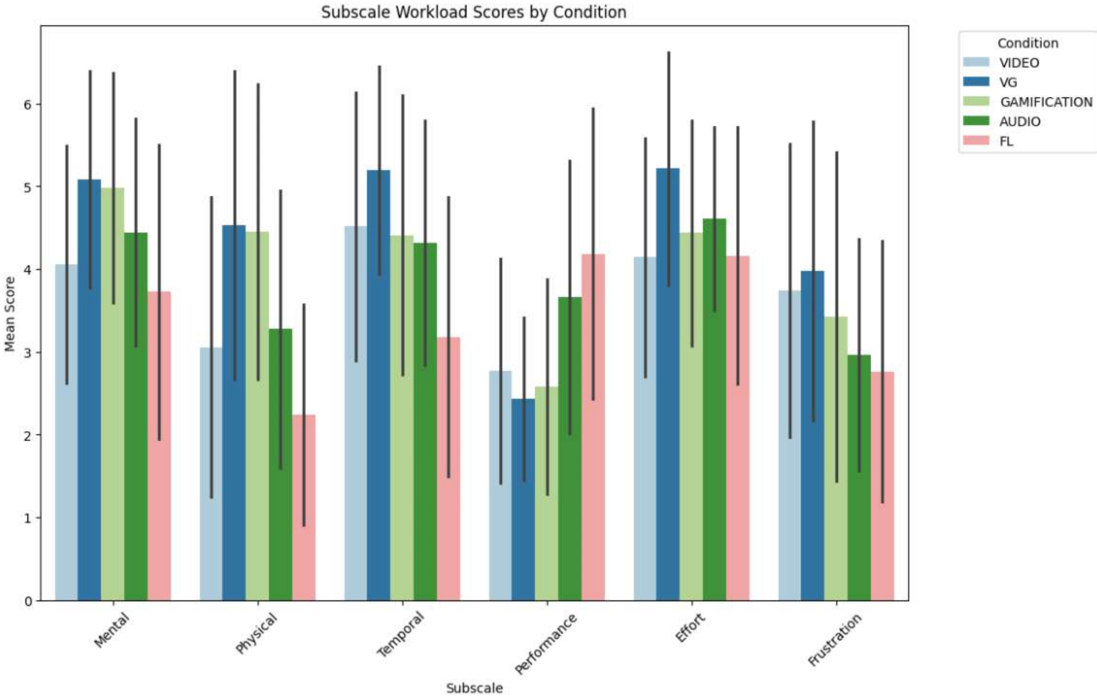


Figure 3: NASA-TLX: Subscale Mean Workload Scores by Condition

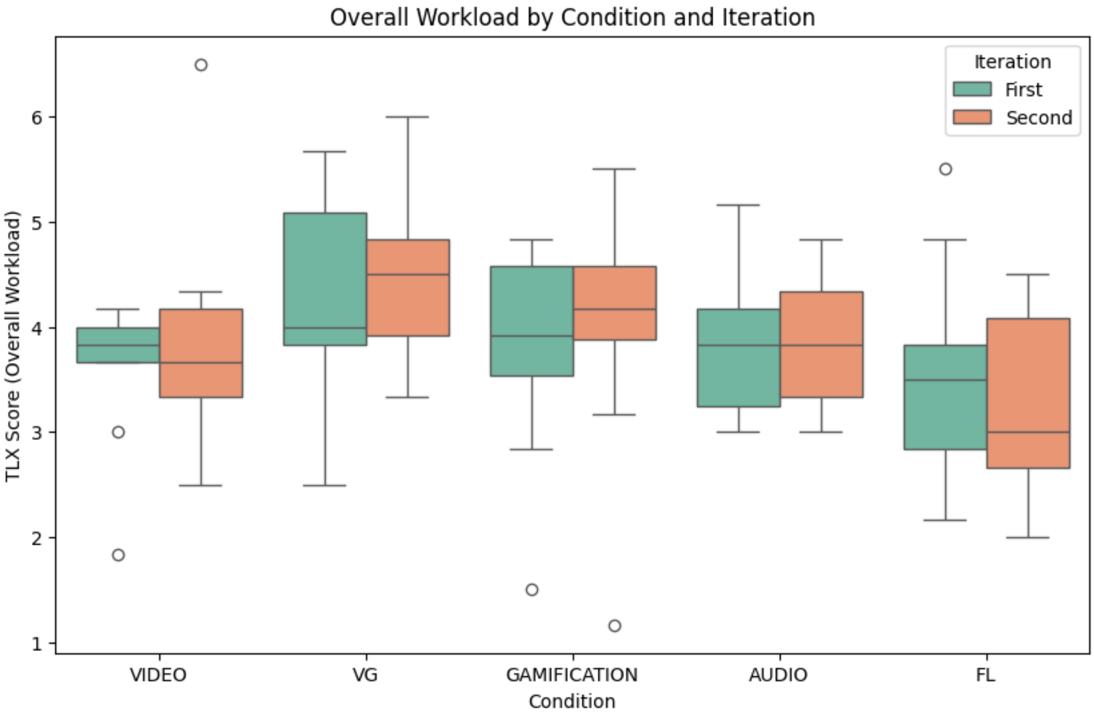


Figure 4: NASA-TLX: Overall Workload By Condition and Iteration

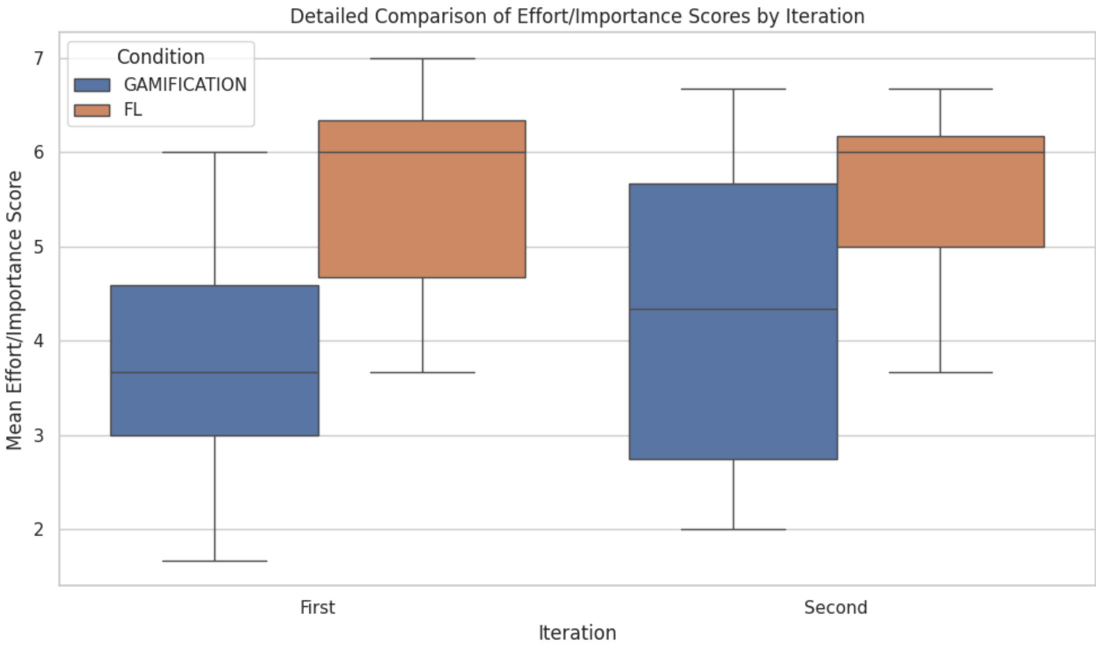


Figure 5: Pairwise Comparison Effort/Importance per Iteration

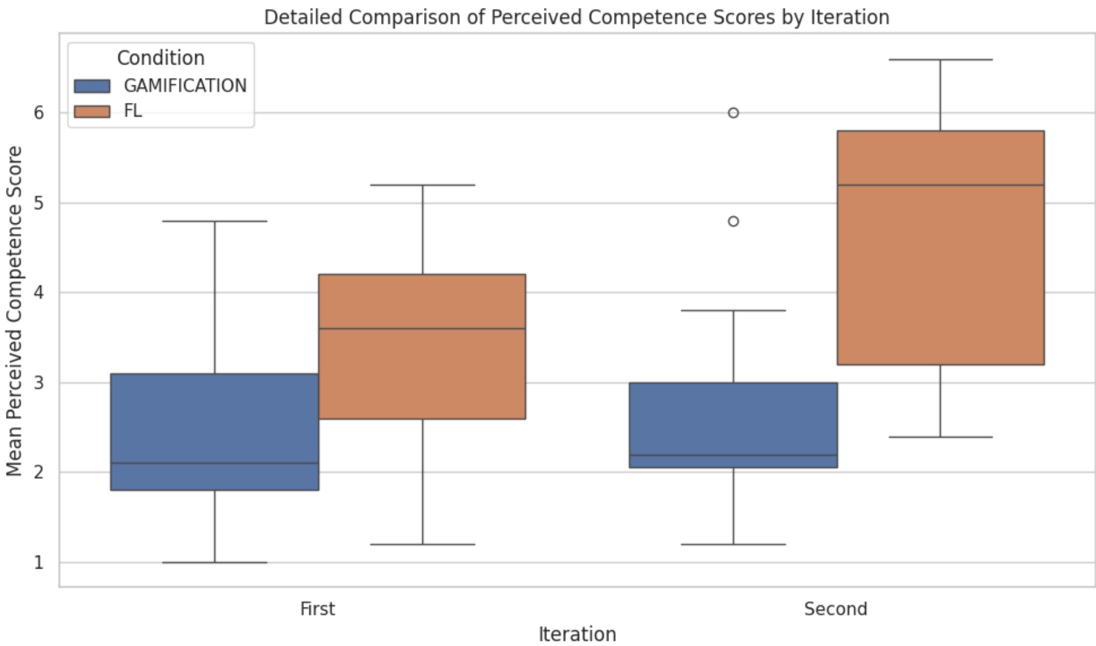


Figure 6: Pairwise Comparison Perceived Competence per Iteration

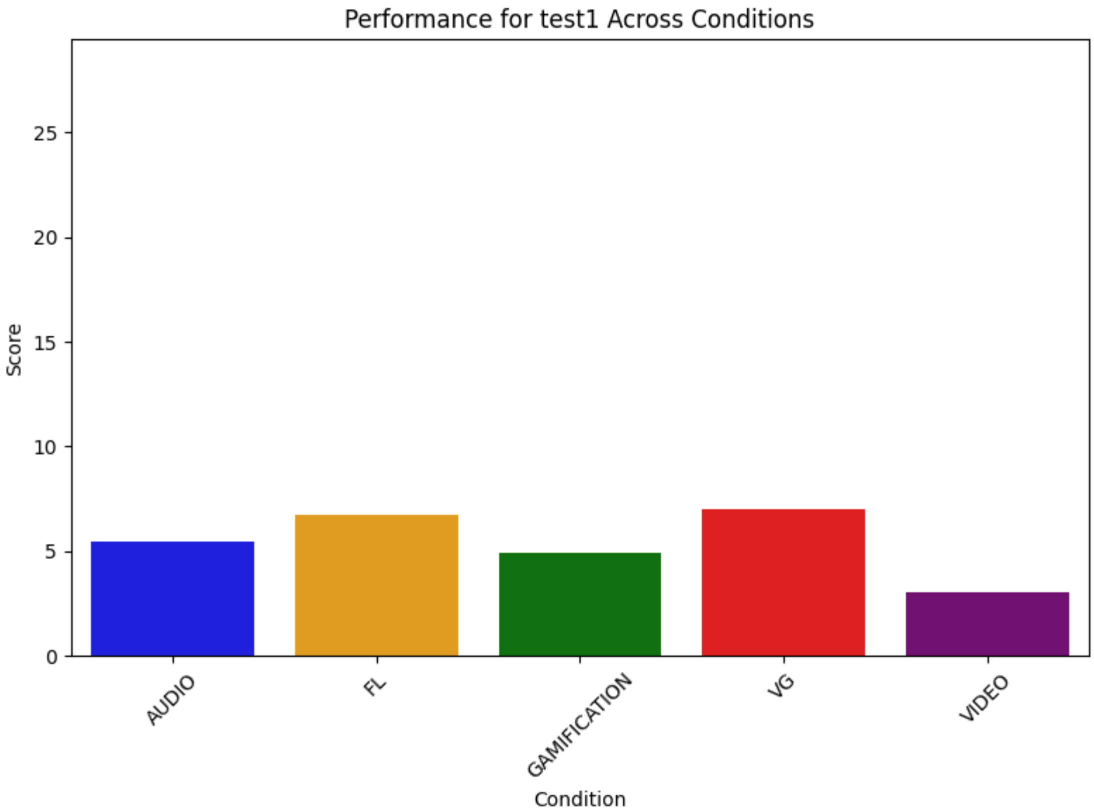


Figure 7: Test Scores: Results of Each Condition For Test1

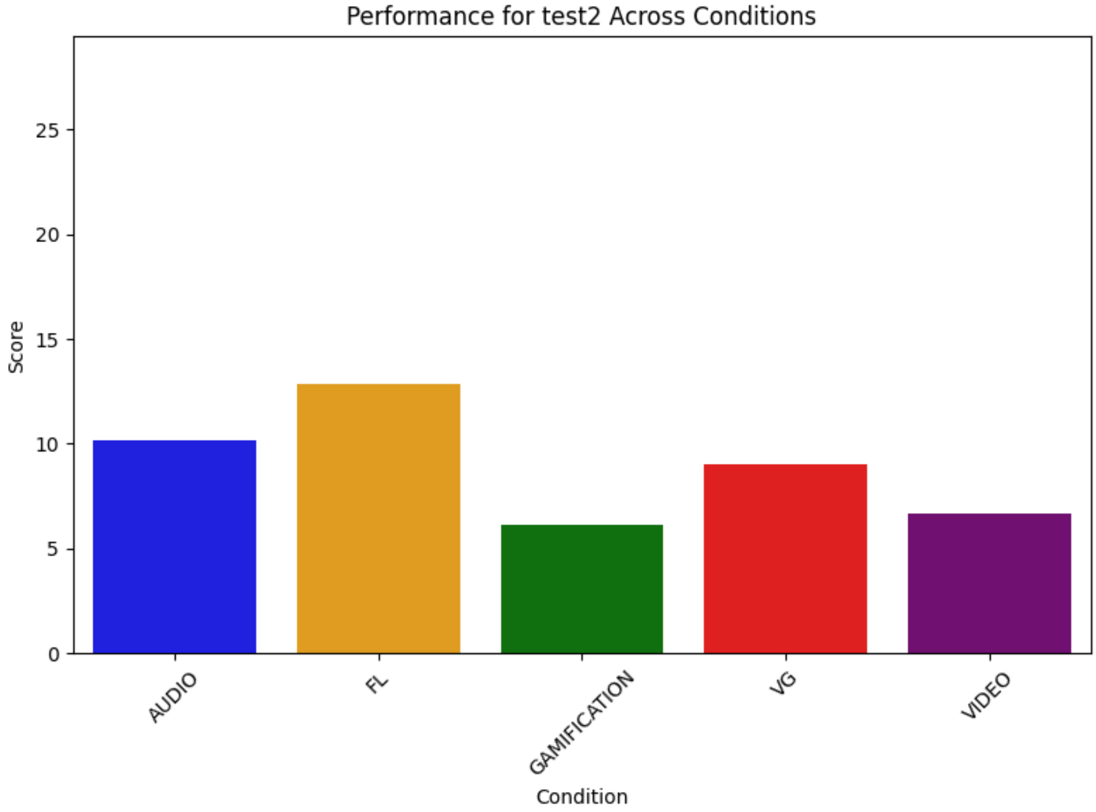


Figure 8: Test Scores: Results of Each Condition For Test2

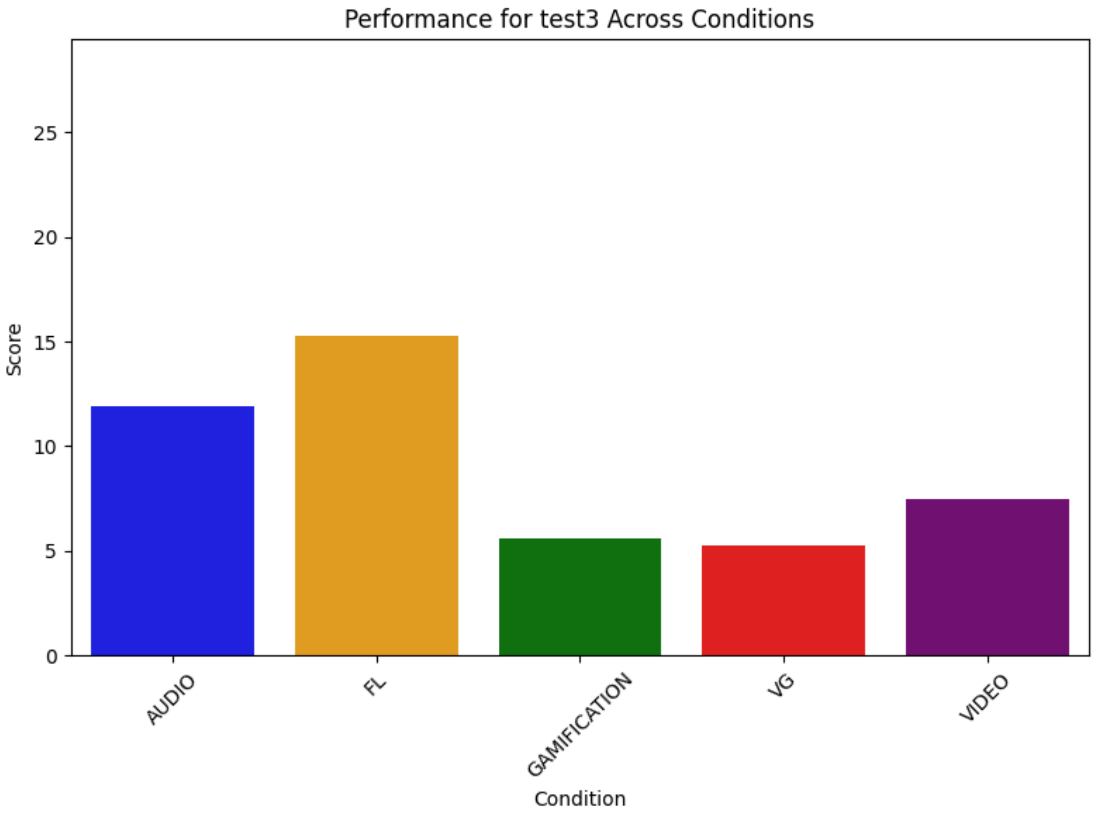


Figure 9: Test Scores: Results of Each Condition For Test3

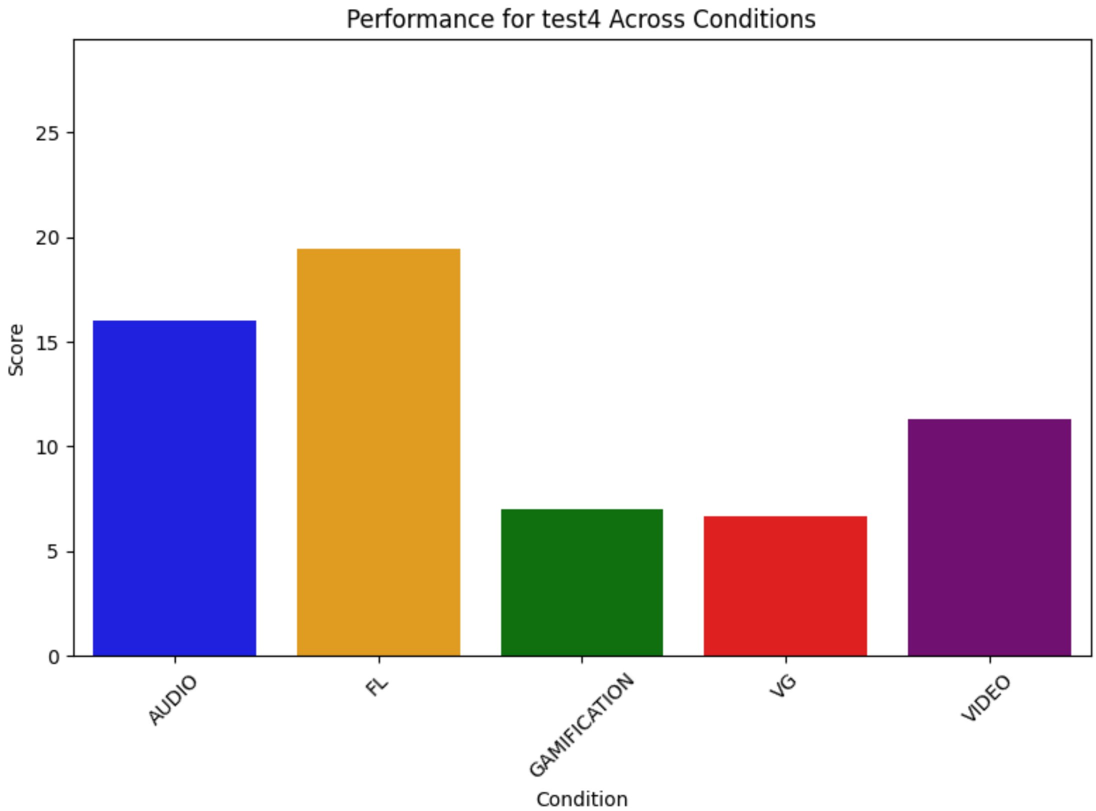


Figure 10: Test Scores: Results of Each Condition For Test4

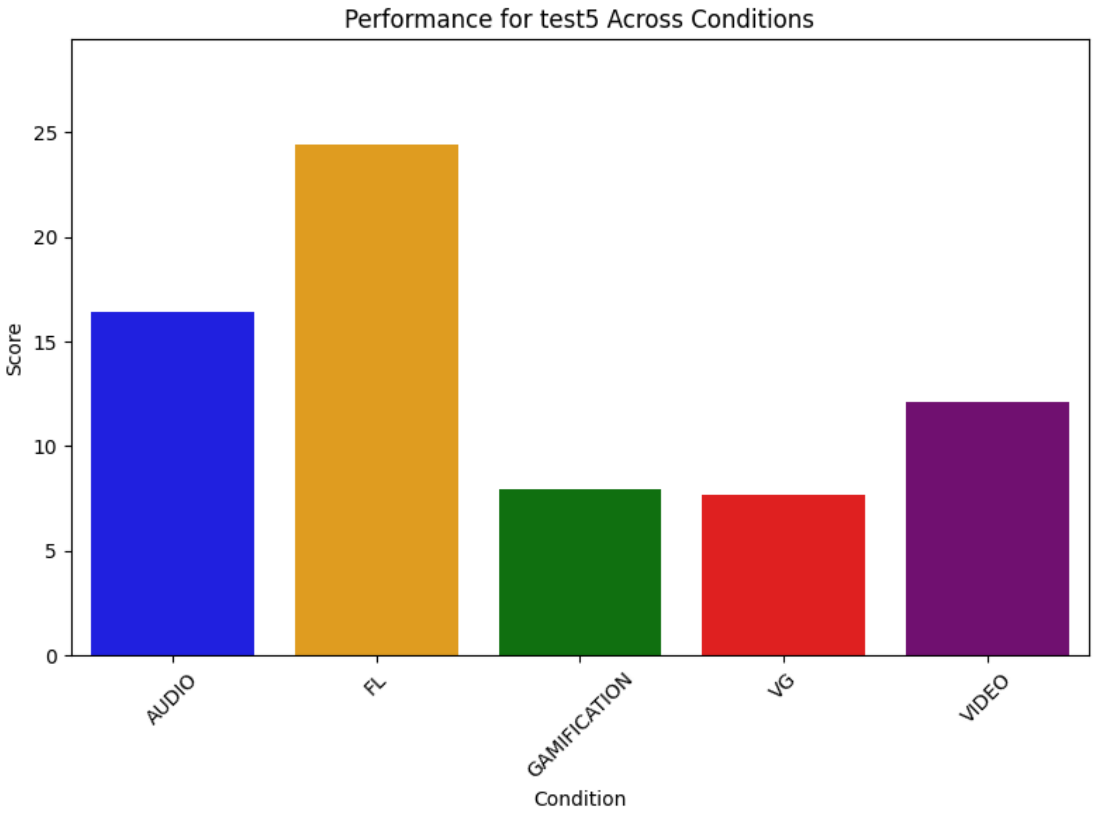


Figure 11: Test Scores: Results of Each Condition For Test5

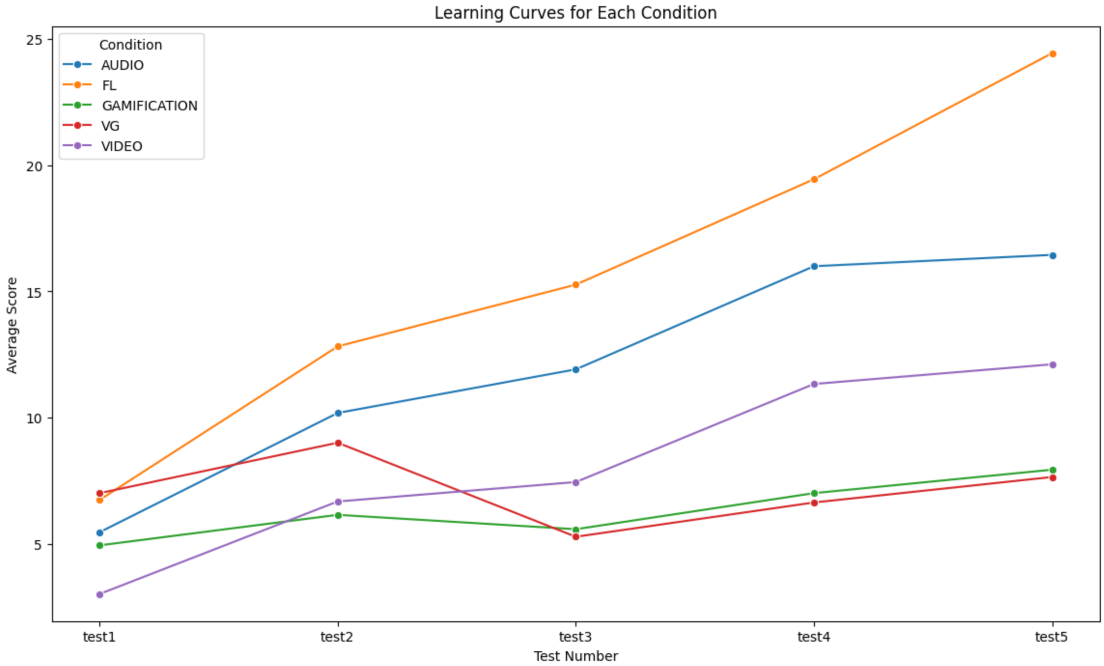


Figure 12: Learning Curve: Results of Each Condition For Each Test

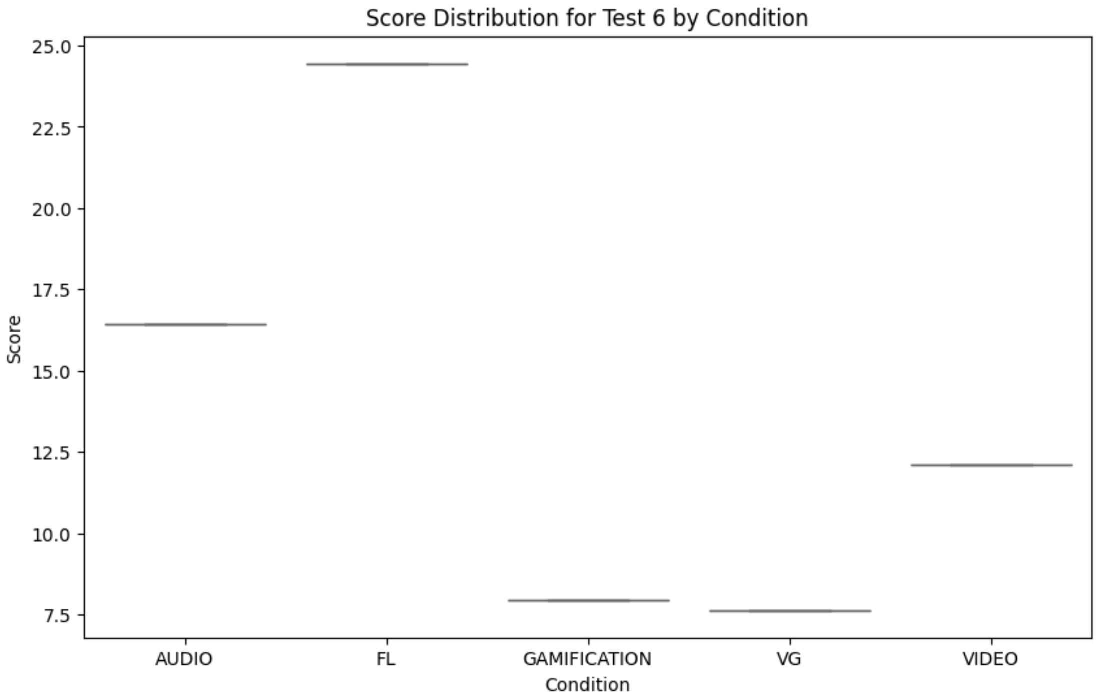


Figure 13: Score Distribution for Test5 by Condition

Appendix C: User Study

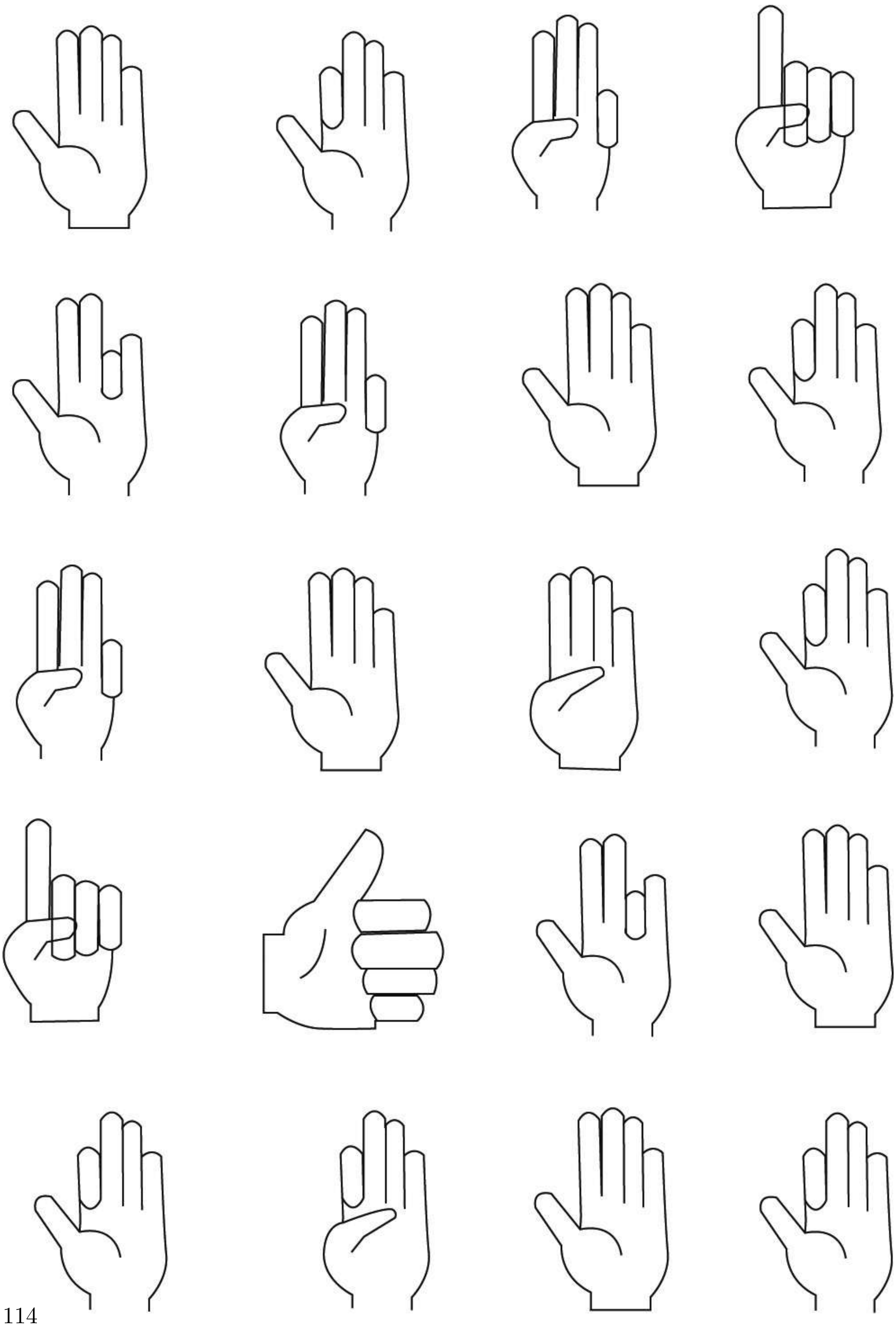


Figure 14: Pictogram map showing choreography steps

ARTIFACT-BASED COMPUTING & USER RESEARCH

Consent Form

Project title

We are interested in understanding the steps one person takes to learn and master a new set of gestures or skills, and what can help them in doing so. We are curious about the steps you take to develop the skill, the method you use, and the procedure you employ, the difficulties you encounter, what you find simple to pick up, and what steps technology can support in the future.

Procedure

The study is split into two days. You will be given a set of hand gestures on day one, and you must learn it by watching the video, playing it again, and imitating it. You get to put the skill even more into practice on the second day. Over the course of two days, your data will be recorded as you practice and learn the skill. You will be questioned further about your experiences and learning process. A moderator will be in the room during these two sessions observing your learning and taking notes without distracting you.

This study involves no risks.

<u>Primary investigators</u>

- Ana Vesic
  - Email: ana.vesic@tuwien.ac.at
- Ambika Shahu
  - Email: ambika.shahu@tuwien.ac.at

I have been asked to take part in the TU Wien research project specified above. I hereby consent to participate in this project.

Figure 15: Consent Form First Page

consent to and understand the following	Yes
• Share images, text, audio and videos where applicable during the study with the experimenters of this study. • Video and audio are captured during the interview. • Tracking position of both hands during the experiment.	O
• The data provided during the study and the recordings of the interviews may be used by the experimenters of this project for related future research work and publications. • I also understand that I could be de-identified when the research team uses still images, audio recordings while reporting this study via research publications and articles	O

Name of participant:

____

Participant signature:

____

Participant number:

____

Experiment mode:

____

Experimenter's signature:

____

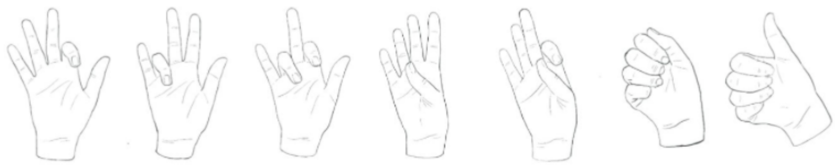
Date: ____

Figure 16: Consent Form Second Page

Experiment Guide

Introduction into this topic Done?___

For this study, we have prepared a small dance routine, only using your right hand. This style of dance is called Hasta Mudra, and originates in India. Our dance consists of following set of hand gestures, which appear in a random order. Your goal is to learn it as quickly and correctly as possible. To support your learning process, you will see a video with the choreography on the screen. We want to know how you learn, what steps you take, and overall process. The goal is to understand the challenges and gaps in which technology can effortlessly fit in and aid in your learning process. Important for you is, that whatever happens, you can't do anything wrong – your job would be to try to learn challenge as accurately and as fast as possible. However, it is strictly **forbidden** to practice this movements at home in any way. This could temper our data and compromise our hard-won findings.



In the end, it doesn't matter how good you perform during the experiments. The only requirement is to try to learn this skill actively.

Flow of the Experiment Explanation Done?___

This is a 2 day study, as previously said, we want to understand the learning process.

We will start with some introductory questions just to understand your background that is related to this topic, we will show you the task and ask some questions regarding your thoughts on it.

In each day there will be three rounds of learning and a follow-up testing, two times. Hence, there will be six rounds of learning and two rounds of testing. It is not important to score high or not make mistakes, we are just curious about how you learn and affects your learning.

At the end of each day there will be a survey to fill out and some follow-up questions.

Figure 17: Experiment Guide First Page

Short Interview after the experiment Done?___

1. How did you find it?
2. What was hard in this challenge? Why was it hard?
3. Did you derive any strategy to learn the skill?
4. Regarding discomfort, is there any, if so can you tell me what it is and how strong it is on the scale 1-10?
5. Would you like it to be slower or faster?
6. Was it confusing? What are your impressions on your performance? How do you feel? If something was confusing what was it, and why was it confusing?
7. Did you enjoy it?
8. Was something too much effort, discomfort?
9. When do you think learning happened?
10. What helped you in learning?
11. How did you improve performance? What helped you in it?
12. Would more rounds help if so how many?
13. Do you think you learned all of the gestures, if so when?

Closing Note Done?___

Thanks for your valuable feedback. It was very helpful to us. If you don't mind answering, we have some follow-up questions based on our observations. Do you have any questions for me? Thank you again for the time you have given to me again.

Figure 18: Experient Guide Example Questions

Bibliography

[Ada71] Jack A. Adams. A closed-loop theory of motor learning. *Journal of Motor Behavior*, 3(2):111–150, 1971.

[AKUF23] Cláuvin Almeida, Marcos Kalinowski, Anderson Uchôa, and Bruno Feijó. Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. *Information and Software Technology*, 156:107142, April 2023.

[ALDK07] Michael D. Matthews Angela L. Duckworth, Christopher Peterson and Dennis R. Kelly. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087–1101, 2007.

[Art08] Anthony Artino. Cognitive load theory and the role of learner experience: An abbreviated review for educational practitioners. *AACE Journal*, 16:425–439, 01 2008.

[AS14] Bruce K. Alexander and Curtis P. Shelton. Empiricism: John locke, david hume, and experience as reality. In *A History of Psychology in Western Civilization*, pages 168–195. Cambridge University Press, 2014.

[Asc15] Christie Aschwanden. Not even scientists can easily explain p-values, 2015. Archived from the original on 25 September 2019. Retrieved 11 October 2019.

[Ban71] Albert Bandura. *Social Learning Theory*. General Learning Corporation, 1971. Archived PDF accessed on December 14, 2024.

[Ban77] Albert Bandura. *Social Learning Theory*. Prentice-Hall, Oxford, England, 1977.

[BGRS23] Charles Baah, Irene Govender, and Prabhakar Rontala Subramaniam. Exploring the role of gamification in motivating students to learn. *Cogent Education*, 10, 05 2023.

[BKC17] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[BPL10] Y-Lan Boureau, Jean Ponce, and Yann LeCun. A theoretical analysis of feature pooling in visual recognition. pages 111–118, 2010.

[BPS10] Bettina Bläsing, Martin Puttke, and Thomas Schack, editors. *The Neurocognition of Dance: Mind, Movement and Motor Skills*. Psychology Press, New York, 2010.

[BW63] Albert Bandura and Richard H. Walters. *Social Learning and Personality Development*. Holt, Rinehart and Winston, New York, 1963.

[Cha88] Paul Chance. *Learning and Behavior*. Wadsworth Publishing Company, 2nd edition, 1988.

[CHQK13] Lewis L. Chen, Leroy Y. Hung, Julie Quinet, and Kevin Kosek. Cognitive regulation of saccadic velocity by reward prospect. *The European Journal of Neuroscience*, 38(3):2434–2444, 2013.

[CMB09] Neil R. Carlson, G. Neil Martin, and William Buskist. *Psychology: The Science of Behavior*. Pearson Education Canada, U.S., 4th edition, 2009.

[Cor24] Microsoft Corporation. *TypeScript Handbook*. 2024. Accessed: 2024-12-14.

[Csi90] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York, NY, USA, 1990.

[CSM04] Raymond C. Chang, Steven Stout, and Ralph R. Miller. Comparing excitatory backward and forward conditioning. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, 57(1):1–23, January 2004.

[DDKN11] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining gamification. volume 11, pages 9–15, 09 2011.

[DEPL94] Edward L. Deci, Hossein Eghrari, Brian C. Patrick, and Dean R. Leone. Facilitating internalization: the self-determination theory perspective. *Journal of Personality*, 62(1):119–142, March 1994.

[Des24] Descript. Descript: Audio/video editing, transcription, and screen recording, 2024. Accessed: 2024-12-14.

[Ebb13] Hermann Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Columbia University. Teachers College. Educational reprints. Teachers College, Columbia University, 1913.

[EFS+16] Alfred O. Effenberg, Ursula Fehse, Gerd Schmitz, Bjoern Krueger, and Heinz Mechling. Movement sonification: Effects on motor learning beyond rhythmic adjustments. *Frontiers in Neuroscience*, 10:219, May 2016.

[EK80] Johannes Engelkamp and Horst Krumnacker. Image- and motor-processes in the retention of verbal materials. *Zeitschrift für experimentelle und angewandte Psychologie*, 1980.

[EKTR93] Anders K. Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406, 1993.

[FP67] Paul Morris Fitts and Michael I. Posner. *Human Performance*. Basic Concepts in Psychology Series, Open University Set Book. Brooks/Cole Publishing Company, Belmont, CA, 2, illustrated, revised edition, 1967. Original from University of Michigan, digitized on 2 Nov. 2007.

[FS57] Charles B. Ferster and Burrhus F. Skinner. *Schedules of Reinforcement*. Appleton-Century-Crofts, 1957.

[fSaWdK+19] European Agency for Safety, Health at Work, Jan de Kok, Paul Vroonhof, Jacqueline Snijders, Georgios Roullis, Martin Clarke, Kees Peereboom, Pim van Dorst, and Iñigo Isusi. Work-related musculoskeletal disorders – prevalence, costs and demographics in the eu, 2019. European Agency for Safety and Health at Work (EU-OSHA), Published by Publications Office.

[GG98] Vittorio Gallese and Alvin I. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501, December 1998.

[Goo24a] Google. Mediapipe solutions guide, 2024. Accessed: 2024-11-02.

[Goo24b] Google. Mediapipe vision: Hand landmarker, 2024. Accessed: 2024-11-02.

[Goo24c] Google AI. Mediapipe: Cross-platform machine learning solutions for edge ai applications, 2024. Accessed: 2024-12-14.

[Goo24d] Google AI. Mediapipe palm detection: A high-fidelity palm detection model, 2024. Accessed: 2024-12-14.

[GPC13] Oonagh M. Giggins, Urlik McCarty Persson, and Brian Caulfield. Biofeedback in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 10(60), 2013.

[HBDH10] Simon Holland, Anders Bouwer, Mathew Dalgelish, and Topi Hurtig. Feeling the beat where it counts: Fostering multi-limb rhythm skills with the haptic drum kit. pages 21–28, 01 2010.

[HBT24] Madeleine E. Hackney, Agnieszka Z. Burzynska, and Lena H. Ting. The cognitive neuroscience and neurocognitive rehabilitation of dance. *BMC Neuroscience*, 25:58, 2024.

[Hey11] Cecilia Heyes. Automatic imitation. *Psychological Bulletin*, 137(3):463–483, May 2011.

[Hic24] Ian Hickson. *The WebSocket API*. 2024. Accessed: 2024-12-14.

[HKS14] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?–a literature review of empirical studies on gamification. *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, pages 3025–3034, 2014.

[HR18] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *Proceedings of the Association for Computational Linguistics (ACL)*, pages 328–339, 2018.

[HS88] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Advances in Psychology*, volume 52, pages 139–183. North-Holland, Amsterdam, 1988.

[HSD+10] Kevin Huang, Thad Starner, Ellen Do, Gil Weinberg, Daniel Kohlsdorf, Claas Ahlrichs, and Ruediger Leibrandt. Mobile music touch: mobile tactile stimulation for passive learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 791–800, New York, NY, USA, 2010. Association for Computing Machinery.

[HTD+03] Granville S. Hall, Edward B. Titchener, Karl M. Dallenbach, Madison Bentley, Edwin G. Boring, and Margaret F. Washburn. *The American Journal of Psychology*. Number v. 14. University of Illinois Press, 1903.

[Hul30] Clark L. Hull. Simple trial and error learning: A study in psychological theory. *Psychological Review*, 37(3):241–256, 1930.

[Hum86] Human Performance Research Group. Nasa task load index (tlx) v.1.0 paper and pencil package. Technical report, NASA Ames Research Center, Moffett Field, California, 1986. Phone: (415)694-6072.

[HZC+17] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[IGC16] Yoshua Bengio Ian Goodfellow and Aaron Courville. *Deep Learning*. MIT Press, 2016. Accessed: 2024-12-14.

[JC99] Susan A. Jackson and Mihaly Csikszentmihalyi. *Flow in Sports: The Keys to Optimal Experiences and Performances*. Human Kinetics, 1999.

[JKCP15] Ankur Joshi, Saket Kale, Satish Chandel, and Dinesh Pal. Likert scale: Explored and explained. *British Journal of Applied Science Technology*, 7:396–403, 01 2015.

[JKG09] Jan Jastorff, Zoe Kourtzi, and Martin A. Giese. Visual learning shapes the processing of complex movement stimuli in the human brain. *Journal of Neuroscience*, 29(44):14026–14038, 2009.

[JYL14] Yoshua Bengio Jason Yosinski, Jeff Clune and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 2014.

[KC09] Donna H. Krasnow and Steven J. Chatfield. Development of the "performance competence evaluation measure": Assessing qualitative aspects of dance performance. *Journal of Dance Medicine Science*, 13(3):101–107, 2009.

[KHX+19] John W. Krakauer, Alkis M. Hadjiosif, Jing Xu, Aaron L. Wong, and Adrian M. Haith. Motor learning. *Comprehensive Physiology*, 9(2):613–663, Mar 14 2019. Erratum in: Comprehensive Physiology. 2019 Jun 12;9(3):1279. doi: 10.1002/cphy.cv09i03corr.

[Kra06] John W. Krakauer. Motor learning: Its relevance to stroke recovery and neurorehabilitation. *Current Opinion in Neurology*, 19(1):84–90, 2006.

[Kre83] John R. Krebs. Animal behaviour. from skinner box to the field. *Nature*, 304(5922):117, 1983.

[LBBH98] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[LBU24] Dorothy Laubscher, Chantelle Bosch, and Est Unm. Validity and reliability of the intrinsic motivation inventory subscales within a self-directed blended learning environment. 10:22–28, 04 2024.

[LGG+17] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, , and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[Lik32] Rensis Likert. *A Technique for the Measurement of Attitudes*. Number 140 in Archives of Psychology. The Science Press, New York, 1932. Originally presented as a Ph.D. dissertation at Columbia University.

[MA00] Todd Maurer and Kimberly Andrews. Traditional, likert, and simplified measures of self-efficacy. *Educational and Psychological Measurement - EDUC PSYCHOL MEAS*, 60:965–973, 12 2000.

[MA13] Richard A. Magill and David I. Anderson. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill Education, New York, NY, USA, 10th edition, 2013.

[Mag07] Richard A. Magill. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill Education, Boston, MA, 8th edition, 2007.

[May09] Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, 2nd edition, 2009.

[Mcl23] Saul Mcleod. Little albert experiment (watson & rayner). Simply Psychology, 2023. Accessed: 31 July 2024.

[MGS21] Arsalan Moinuddin, Ashish Goel, and Yashendra Sethi. The role of augmented feedback on motor learning: A systematic review. *Cureus*, 13(11):e19695, November 2021.

[MWW21] Moritz Mödinger, Alexander Woll, and Ingo Wagner. Video-based visual feedback to enhance motor learning in physical education—a systematic reviewvideobasiertes visuelles feedback zur verbesserung des motorischen lernens im sportunterricht – ein systematisches review. *German Journal of Exercise and Sport Research*, 52, 12 2021.

[NAS24] NASA Human Systems Integration Division. Nasa task load index (nasa-tlx). `https://humansystems.arc.nasa.gov/groups/tlx/`, 2024. Accessed: 2024-11-21.

[NH10] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[NL12] Karl M. Newell and Yeou-Teh Liu. Functions of learning and the acquisition of motor skills (with reference to sport). *The Open Sports Sciences Journal*, 5:17–25, 2012.

[OGS01] James Onate, Kevin Guskiewicz, and Robert Sullivan. Augmented feedback reduces jump landing forces. *The Journal of orthopaedic and sports physical therapy*, 31:511–7, 10 2001.

[OO23] Daniel Odebiyi and Chris Okafor. *Musculoskeletal Disorders, Workplace Ergonomics and Injury Prevention*. InTech, 02 2023.

[Ope24a] OpenJS Foundation. *Express: Fast, unopinionated, minimalist web framework for Node.js*. 2024. Accessed: 2024-12-14.

[Ope24b] OpenJS Foundation. *Node.js Documentation*. 2024. Accessed: 2024-12-14.

[Pav27] Ivan P. Pavlov. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press, London, UK, 1927.

[PJVT24] Marija Pranjić, Thenille Braun Janzen, Nikolina Vukšić, and Michael Thaut. From sound to movement: Mapping the neural mechanisms of auditory–motor entrainment and synchronization. *Brain Sciences*, 14(11):1063, 2024.

[PMMK24] Carmen Pârvu, George Danut Mocanu, Petronel-Cristian Moisescu, and Zukhro Bahadirovna Khamraeva. The optimisation of education through feedback and advanced technology in motor learning and the correction of technical sports errors. *Research and Education*, 2024.

[Pre01] Marc Prensky. Digital game-based learning. *McGraw-Hill, New York*, 1, 01 2001.

[PVH$^+$18] Franck Potdevin, Olivier Vors, Aurore Huchez, Michel Lamour, Keith Davids, and Christophe Schnitzler. How can video feedback be used in physical education to support novice learning in gymnastics? effects on motor learning, self-assessment and motivation. *Physical Education and Sport Pedagogy*, 23(6):559–574, 2018.

[RBJ$^+$24] Ghazi Rekik, Yosra Belkhir, Ghada Jouira, et al. Using video modelling to teach motor knowledge in a physical education context: A morning-afternoon differences investigation. *Education and Information Technologies*, 2024.

[RC04] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004.

[RCM12] Claire Renzetti, Daniel Curran, and Shana Maier. *Women, Men, and Society*. Pearson, 6th edition, 2012.

[RCP90] Richard M. Ryan, James P. Connell, and Robert W. Plant. Emotions in non-directed text learning. *Learning and Individual Differences*, 2(1):1–17, 1990.

[RD00a] Richard M. Ryan and Edward L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67, 2000.

[RD00b] Richard M. Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78, 2000.

[RFGF96] Giacomo Rizzolatti, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2):131–141, 1996.

[RLGS15] Thomas R Reppert, Karolina M Lempert, Paul W Glimcher, and Reza Shadmehr. Modulation of saccade vigor during value-based decision making. *The Journal of Neuroscience*, 35(46):15369–15378, 2015.

[RMK83] Richard M. Ryan, Valerie Mims, and Richard Koestner. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4):736–750, 1983.

[Rya82] Richard M. Ryan. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3):450–461, 1982.

[Sch86] Steven Schwartz. *Classic Studies in Psychology*. Mayfield Publishing, Palo Alto, CA, 1986. Print.

[SH20] Michael Sailer and Lisa Homner. The gamification of learning: a meta-analysis. *Educational Psychology Review*, 32:77–112, 03 2020.

[SHA16] Reza Shadmehr, Helen J Huang, and Alaa A Ahmed. A representation of effort in decision-making and motor control. *Current Biology*, 26(14):1929–1934, 2016.

[SJHB09] Daniel Spelmezan, Mareike Jacobs, Anke Hilgers, and Jan Borchers. Tactile motion instructions for physical activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 2243–2252, New York, NY, USA, 2009. Association for Computing Machinery.

[Ski38] Burrhus F. Skinner. *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century-Crofts, New York, 1938.

[Ski53] Burrhus F. Skinner. *Science and Human Behavior*. Macmillan, New York, 1953.

[Ski57] Burrhus F. Skinner. *Verbal Behavior*. Century Psychology Series. Prentice-Hall, New York, 6 edition, 1957. Bobbs-Merrill reprint series in the social sciences.

[Ski68] Burrhus F. Skinner. *The Technology of Teaching*. Appleton-Century-Crofts, New York, 1968.

[SL11] Richard A. Schmidt and Timothy D. Lee. *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics, Champaign, IL, 5th edition, 2011.

[SL19] Richard A. Schmidt and Timothy D. Lee. *Motor Learning and Performance: From Principles to Application*. Human Kinetics, Champaign, IL, USA, 2019.

[SLW+19] Richard Schmidt, Timothy Lee, Carolee Winstein, Gabriele Wulf, and Howard Zelaznik. *Motor Control and Motor Learning: A Behavioral Emphasis*. Human Kinetics, Champaign, IL, 6th edition, 2019.

[SRM+20] Rodrigo Smiderle, Sandro Rigo, Leonardo Marques, Jorge Coelho, and Patricia Jaques. The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, 7, 01 2020.

[SRRW13] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin Review*, 20:21–53, February 2013.

[SS08] Ladan Shams and Aaron R. Seitz. Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11):411–417, November 2008.

[SSW14] Yongqiang Sun, Xiao-Liang Shen, and Nan Wang. Understanding the role of consistency during web–mobile service transition: Dimensions and boundary conditions. *International Journal of Information Management*, 34(4):465–473, 2014.

[SW65] Samuel S. Shapiro and Martin B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[Swe88] John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988.

[TA16] Caitlin Tenison and John R. Anderson. Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5):749–767, 2016.

[TS14] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1653–1660, 2014.

[TSO19] Mirko Thalmann, Alessandra S. Souza, and Klaus Oberauer. How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1):37–55, January 2019.

[Ubi24] Ubisoft Entertainment. Just dance now: Dance to your favorite hits, 2024. Accessed: 2024-12-14.

[VDD+21] Pierre Vassiliadis, Gerard Derosiere, Cecile Dubuc, Aegryan Lete, Frederic Crevecoeur, Friedhelm C. Hummel, and Julie Duque. Reward boosts reinforcement-based motor learning. *iScience*, 24(7):102821, 2021.

[vdKvDvV+19] Katinka van der Kooij, Rosanne van Dijsseldonk, Milou van Veen, Frans Steenbrink, Coen de Weerd, and Krista E Overvliet. Gamification as a sustainable source of enjoyment during balance and gait exercises. *Frontiers in Psychology*, 10:294, 2019.

[vdLSBJ11] Janet van der Linden, Erwin Schoonderwaldt, Jon Bird, and Rose Johnson. Musicjacket—combining motion capture and vibrotactile feedback to teach violin bowing. *Instrumentation and Measurement, IEEE Transactions on*, 60:104 – 113, 02 2011.

[VLDD22] Pierre Vassiliadis, Aegryan Lete, Julie Duque, and Gerard Derosiere. Reward timing matters in motor learning. *iScience*, 25(5):104290, 2022.

[VVAvdK13] Valentijn Visch, Niko Vegt, Hester Anderiesen, and Katinka van der Kooij. Persuasive game design: A model and its definitions. In *CHI 2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences*, Paris, France, 2013.

[Wat13] John B. Watson. Psychology as the behaviorist views it. *Psychological Review*, 20(2):158–177, 1913.

[WH12] Kevin Werbach and Dan Hunter. *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press, Philadelphia, PA, 2012.

[WKC09] Virginia Wilmerding, Donna Krasnow, and IADMS Dance Educators' Committee. Motor learning and teaching dance. `https://iadms.org/media/3588/iadms-resource-papers-motor-learning-and-teaching-in-dance.pdf`, 2009.

[WL16] Ronald L. Wasserstein and Nicole A. Lazar. The asa's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016.

[WNOS13] Pieter Wouters, Christof Nimwegen, Herre Oostendorp, and Erik Spek. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105:249, 02 2013.

[WP01] Gabriele Wulf and Wolfgang Prinz. Directing attention to movement effects enhances learning: A review. *Psychonomic Bulletin Review*, 8(4):648–660, 2001.

[WS12] Josef Wiemeyer and Philipp Schneider. Applying serious games to motor learning in sport. *Int. J. Game-Based Learn.*, 2(4):61–73, October 2012.

[ZBV+20] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. Mediapipe hands: On-device real-time hand tracking, 2020.