Auditory, Visual, or Both? Comparing Visual and Auditory Representations of Game Elements in a Gamified Image-Tagging Task

MARC SCHUBHAN, German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus, Germany
MAXIMILIAN ALTMEYER, Saarland University of Applied Sciences (htw saar), Germany
KATJA ROGERS, University of Amsterdam, Netherlands
DONALD DEGRAEN, University of Duisburg-Essen, Germany
PASCAL LESSEL and ANTONIO KRÜGER, German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus, Germany

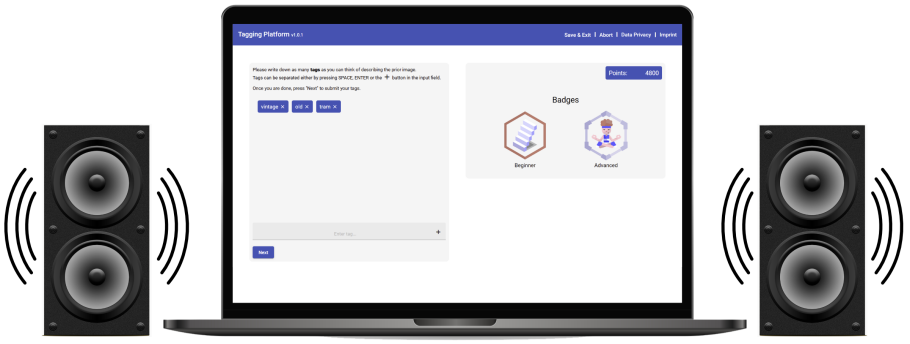

Fig. 1. Our tagging platform showing a tag input field on the left and gamification elements complemented with sound on the right. Depending on the modalities, the speakers are playing sound effects when a user gains points and badges. Laptop and speaker templates taken from Freepik.com

In gamification users show increased motivation and engagement towards tasks. So far, this effect has mostly been empirically tested based on the visual depiction of game elements, while research on the use and addition of auditory aspects is sparse. In this work we investigate the effect of different modalities of game elements (auditory, visual, audiovisual) on user experience, perception and performance in a gamified image-tagging task. We approached this via an online validation survey to find suitable sound effects (N=50), a main quantitative study (N=109) and a qualitative semi-structured interview (N=9). Our results show that while

Authors' Contact Information: Marc Schubhan, marc.schubhan@dfki.de, German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus, Saarbrücken, Germany; Maximilian Altmeyer, maltmeyer@acm.org, Saarland University of Applied Sciences (htw saar), Saarbrücken, Germany; Katja Rogers, k.s.rogers@uva.nl, University of Amsterdam, Amsterdam, Netherlands; Donald Degraen, donald.degraen@uni-due.de, University of Duisburg-Essen, Essen, Germany; Pascal Lessel, pascal.lessel@dfki.de; Antonio Krüger, antonio.krueger@dfki.de, German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus, Saarbrücken, Germany.

ACM 2573-0142/2024/10-ART294
https://doi.org/10.1145/3677059

visual gamification increases performance, auditory and audiovisual gamification had no effect. However, they were shown to have an influence on the user's flow state. Our qualitative follow-up study shed light on underlying reasons and revealed each modality has its own drawbacks and advantages and that combining both visual and auditory aspects was preferred by participants.

CCS Concepts: • **Human-centered computing** → **Auditory feedback**; **Empirical studies in HCI**.

Additional Key Words and Phrases: sound feedback, SFX, gamification, image tagging

ACM Reference Format:
Marc Schubhan, Maximilian Altmeyer, Katja Rogers, Donald Degraen, Pascal Lessel, and Antonio Krüger. 2024. Auditory, Visual, or Both? Comparing Visual and Auditory Representations of Game Elements in a Gamified Image-Tagging Task. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY, Article 294 (October 2024), 28 pages. https://doi.org/10.1145/3677059

1 Introduction

Gamification is defined as "the use of game elements in non-game contexts" [19]. Often these game elements can be found in the form of e.g. points, badges or leaderboards. It has been used in a broad range of domains including health, education, commerce or crowdsourcing to improve user experience and performance [28, 71]. Although past research has shown that a majority of gamified interventions can be considered as successful [28, 37, 71], Koivisto and Hamari [37] concluded in their systematic literature review that "*while the results in general lean towards positive findings about the effectiveness of gamification, the amount of mixed results is remarkable*". Thus, understanding what constitutes an enjoyable, gameful user experience in gamified systems is an ongoing issue in gamification research [36, 55].

Sound effects ("SFX") could play a major role in this regard: In games research, audio has been shown to support flow and immersion as well as to affect a player's emotional state [56]. This can in turn be utilized to support the narrative [70, 83] and enhance the perceived realism of media experiences [42]. Outside of the games and gamification domain, it has been found that musical sounds can be inherently rewarding [68]. These findings are relevant for gamification, which aims to reward users and stimulate the aforementioned psychological states of for example flow, immersion, and emotions in non-game contexts. In gamification, the visual presentation of game elements, such as points, badges, or levels, are often complemented through SFX (e.g. when unlocking a badge [79], when gaining or losing points [38, 54, 82], or when congratulating users upon reaching a certain level or goal [47]). They are thus an integral part of a users' experience in gamified systems. The fact that SFX are themselves considered gamification elements [62] further emphasizes their importance in the domain of gamification.

However, albeit its ubiquity in gamified systems, the literature on actual effects of SFX on user experience and performance in these systems is sparse. Past research has primarily focused on the *visual* modality of gamification, e.g. by investigating the effectiveness of individual game elements [52, 76], the need for customization [43] and personalization [33, 77], or the role of choice [44]. Regarding the *auditory* modality of gamification, past research has for example studied whether different SFX accompanying the game element points make a difference in performance and motivation [2] or vocal gamified feedback for everyday household tasks through Amazon's Alexa [12]. However, as far as we know, auditory and visual gamification has not yet been investigated through a systematic comparison. Thus, it remains unclear if and how SFX affect the user experience and motivation, or if audio-only gamification can engage users on its own, which would make it interesting to consider for contexts where visual game elements are less applicable. In this work, we take a step back and aim to contribute answers to this open question in the course of a

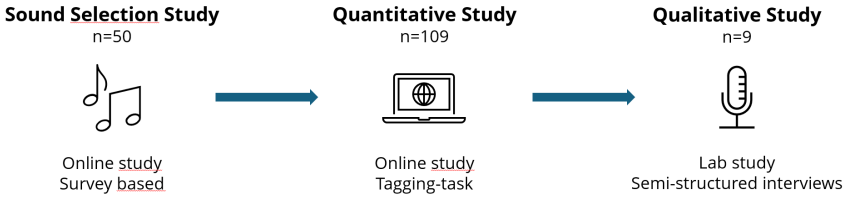

Fig. 2. Overview on our three-step procedure and methodology applied to each step of this work.

validation study and two main studies, leveraging both quantitative and qualitative approaches (see Figure 2).

First, we followed an iterative selection process to identify SFX that suitably represent the employed game elements (badges and points). We selected these game elements because they are among the most frequently used ones [28, 36, 71]. In an online validation study (N=50), we were able to distill the most suitable SFX—one representing badges and one representing points—to provide appropriate stimuli for the remaining studies of this work. Second, similar to past research [52, 53, 69], we developed an image-tagging platform to compare auditory and visual gamification feedback to no feedback or their combination regarding task performance and user experience (N=109). Findings of our online study show that visual gamification significantly increases task performance in this context, replicating previous research on gamified image tagging [44, 52, 53, 69]. Furthermore, while we did not find significant differences between visual, auditory, audiovisual and no feedback regarding intrinsic motivation and affective experience, results revealed significant differences regarding the prevalence of flow experiences. Based on the effects, it seems that audiovisual gamification (i.e. accompaniment of visual game elements by SFX) is advantageous regarding flow when compared to the other modalities. Lastly, to better understand the interplay of visual and auditory gamification, we conducted a qualitative lab study (N=9) with semi-structured interviews. We found that while visual gamification alone is perceived well, it may increase distraction from the task. Moreover, results suggest that SFX alone are also perceived as motivating and could be particularly suitable to increase attention. However, they come at the cost of increased confusion, due to users struggling to understand the semantics behind them. With respect to the quantitative results for audiovisual gamification, these qualitative findings suggest that combining visual gamification and SFX may cancel out each other's disadvantages. Based on the combined knowledge of these user studies, we discuss advantages and drawbacks of using auditory, visual and audiovisual feedback in gamified systems and beyond.

2 Related Work

This work is placed in the context of gamified systems and investigates fundamental aspects in terms of auditory depiction and perception of game elements. Hence, in the following, we will first elaborate on the past and current state of gamification research, and why SFX are a potentially interesting and important aspect to investigate. For this, we will also take a look at the related context of gameful systems, where the impact of audio in general has been investigated to a further extent, and outline sonification literature, which makes use of SFX to represent data.

2.1 Visual Gamification

Gamification research evolved in two waves, as described by Nacke and Deterding [55]. The first wave tried to uncover whether gamification works, i.e., to understand its general effectiveness and the effectiveness of individual game elements. A comprehensive literature review revealed that

gamification in general yields positive effects (Hamari et al. [28]). Mekler et al. [52] for example investigated the effect of points, levels and leaderboards on intrinsic motivation and found that they neither improve nor decrease it, but that user performance still was significantly higher with gamification. They did this by counting tags users created on an image-tagging platform with or without said game elements. The literature review [28] also revealed that the effectiveness of gamification depends on user personality, as well as the context in which it is applied. Investigations in various contexts emerged, such as education [4], or health and well-being [34]. While most studies report positive results [20], some also investigated the "dark side of gamification" (Toda et al. [76]), finding that for example leaderboards are linked to lower performance in educational contexts.

The second, and current wave according to Nacke and Deterding [55] tries to find answers to "how", "when" and "how and when not" gamification works. *Goal-setting theory* [49] for example has been used to explain how leaderboards work (Landers et al. [40]): Users tend to aim near the top of the leaderboard, which turned out to be comparable to giving users difficult goals in their study. Other works conduct more user-centered research. For example, a study investigating how giving users a choice to enable or disable gamification affects their motivation found out that those who disliked the game elements and were able to disable them, showed significantly higher task performance compared to those who disliked them, but had no choice (Lessel et al. [44]). Others (e.g. Jia et al. [33] or Rogers et al. [65]) looked into tailoring gamification, by finding a connection between personality traits, or player types, and game elements like points, levels, and leaderboards among others.

What all of the aforementioned studies or literature reviews have in common is their focus on the visual aspects of gamification elements, neglecting the potential influence of SFX. In the aforementioned examples and in similar research, SFX were often not part of the gamification elements, or investigating the impact of SFX was no substantial part of their research question, reflecting the prominence of visual gamification.

2.2 Audio in Gamified Systems and Games

In the digital world, sound in general not only enhances user experience in many applicable contexts like entertainment media, but also helps making applications more accessible (e.g. [16, 22, 30]). While literature on SFX in gamification is sparse, first approaches in this context exist. Altmeyer et al. investigated the effect of sound feedback on user performance and user experience in a gamified image classification task [2]. To this end, they implemented a task where participants would have to mark specific areas of an image containing specific objects. Points were awarded for each correct area and subtracted for a false choice. In five conditions, they tested the impact on task performance of various sounds played when awarding participants with points and compared it to a control group without sound. Their results indicate no significant difference in terms of user performance, affect, immersion or enjoyment. Nonetheless, differences were found regarding perceived pressure and tension, which led to the recommendation to not use low-valence sounds in gamification as a consequence. While their study did investigate effects based on the presence of certain SFX, they focused primarily on investigating SFX following the Circumplex Model of Affect [61] on a scale from low valence to high valence and low arousal to high arousal. A systematic approach comparing no gamification to visual gamification, sound feedback alone and a combination of both has yet to be done to draw conclusions on the impact of SFX on user engagement and experience in gamified systems.

Auditory gamification has also been investigated using virtual assistants like Amazon's Alexa to provide vocal gamified feedback for everyday household tasks (Bräuer and Mazarakis [12]). They tested their system against a control group without gamification in a competitive, as well as

in a cooperative setting. Both gamified conditions showed significantly better task performance compared to the control group, indicating that auditory feedback alone can already be used to increase user motivation. Yet they did not compare their results to a visual implementation of their gamified system, hence, a comparison to established gamification systems is missing. Furthermore, their work focuses on spoken feedback and does not look at SFX specifically.

Miranda and Palmer [54] investigated whether gamification and SFX can increase intrinsic motivation. They implemented a system using several gamification elements, such as points, a highscore and streaks in a visual search task. Besides neutral sound feedback for correct and incorrect responses, additional sounds for multiplicator rewards were played. Three experiments were conducted to investigate the impact of the gamification elements and various sound effects. The results showed that sound effects without gamification led to attentional capture but resulted in the lowest interest and enjoyment scores, and thus were less motivating. Gamification without sound increased intrinsic motivation but felt not rewarding enough to capture attention. The combination of gamification elements and sound was the most effective option for both capturing attention and increased intrinsic motivation. However, it needs to be considered that different sounds were played in the sound condition and several gamification elements were used which were not directly mapped to specific visual counterparts. Both make it hard to compare auditory versus visual feedback mechanisms. Moreover, the question which underlying reasons were deciding causes for the effects remains open.

Although the influence of SFX on user perception and engagement can not be fully grasped yet when looking at the field of gamification, it is worthwhile to look at related contexts like for example games, where studies on auditory impact have been conducted to a greater extent. Here, sound can for example be found in the form of background music (BGM) setting the mood for a scene, SFX giving a player auditory feedback, or voice lines providing dialogue, explanations etc. [6, 48] and has been shown to impact user performance and experience in the past. In literature there is evidence that background music can impact the perception of elapsed time (Cassidy and Macdonald [15]). An effect on perceived flow when playing with or without music has also been found in a study testing the influence of different levels of beats per minute (BPM) on player performance and game perception (Levy et al. [45]). While the BPM itself did not yield significant results, the addition of background music generally significantly improved the players' flow perception, including effects on player concentration and the autotelic experience. A study on the effects of background music on risk-taking behavior, found that in the presence of music, players decrease their will to take risks, as long as they had no prior knowledge about the game (Rogers et al. [63]). Further studies found that player performance can be affected through game-unrelated music [74, 85] and established the connection between background music and engagement and immersion in the gaming context [24, 56]

In general, related work has shown that the addition of audio alone can already influence players in various ways. Next to the motivational factors that are prevalent in gamification research, especially the findings on perceived flow and player performance are interesting here, as a similar effect could benefit gamification elements often applied to repetitive or even tedious tasks. While these works underlined the positive effects of music, they did not investigate different audio features influencing the experience beyond the addition of the music itself. According to Levy et al.'s study [45] the BPM were not decisive for their results, hence, the effect could be caused either by another facet of music, or the addition of audio in general and might consequently apply to SFX in gamification elements as well.

2.3 Sonification

Sonification describes "the use of nonspeech audio to convey information" [39]. It is an established method which has been applied to and researched in various fields, such as physical activity or physiotherapy to represent movement [57, 80], astronomy for multi-sensory data representations [86], or bioinformatics to represent e.g. strands of DNA [60]. Plaisier et al. [60] successfully use sonification not only to represent data, but also to increase public engagement with their work.

In the past, various concepts have emerged in sound design targeting SFX for sonification. One of the first concepts are so called *Auditory Icons* introduced and described by Gaver as "caricatures of naturally-occuring sounds" [26, 27]. Auditory Icons can be used to convey information while using devices, e.g. when dragging, dropping or selecting objects, or informing users about occurring events, improving navigation, collaboration, inform about ongoing process, as well as previous and possible interactions [27].

Blattner et al. later introduced the concept of *Earcons* [7, 11]. These are rhythmic sequences of pitches varying in rhythm, intensity, timbre, register and dynamics. Due to their complexity and structure, they have been successfully used to represent for example hierarchical structures often found in menu-like interfaces. Earcons can be further divided into one-element Earcons, which are e.g. single notes, compount Earcons, which are a combination of two or more elements, and inherited Earcons, which represent a "family", i.e. Earcons with similar attributes.

Lastly, Walker et al. [81] introduce the concept of *Spearcons*. Those are spoken phrases, which are sped up to the point they are not recognized as speech anymore. In a first study, Walker et al. showed that their Spearcons outperformed Auditory Icons and Earcons in the context of menu navigation. Participants were generally faster and more accurate. This was verified by in a study comparing Auditory Icons, Earcons, Spearcons and speech based feedback representing environmental features (Dingler et al. [21]). Results indicate that Spearcons excel in terms of learnability compared to Earcons and Auditory Icons. Yet, according to Li et al. [46] they come with certain drawbacks, such as the need to reproduce them for several languages, the need for a priori knowledge about potential contents, and lower recognition rate in verbal tasks, making them generally less applicable in the context of gamification.

Nonetheless, the area of sonification generally shows that SFX can be specifically used to convey information to users, which reinforces the aim to make use of fitting SFX in the context of gamification, where conveying information is fundamental. Given tasks such as regular running, users can not always rely on visual displays to communicate the required information, which is where SFX could accomodate or function as their own gamification element. Additionally, SFX can help to capture a user's attention as indicated in literature (Miranda and Palmer [54]).

3 Sound Selection Process and Evaluation

In this section, we provide details on our SFX selection process. This was done through an initial screening of 100 SFX, followed by an online study in which participants were asked to rate the suitability of various SFX to represent the game elements points and badges. We chose these game elements as they are frequently used in gamification research [28, 36, 71].

3.1 Initial Screening

In order to investigate a reasonable set of sounds, we started our pre-selection by collecting a set of 100 different samples. With sample diversity in mind, we chose sounds from two different providers, Pixabay[1] and Mixkit[2]. During this process, we took samples from the categories *Notification*, *Win*

[1] https://pixabay.com/, last accessed on August 15, 2024
[2] https://mixkit.co/, last accessed on August 15, 2024

and *Technology* on each site. As we required a set of rather short sound effects that should not resemble recognizable things to reduce bias, we applied the following criteria to preemptively exclude unfitting sounds:

- SFX should have a duration shorter than 5 seconds
- SFX should not be technical sounds (e.g. phone ringing or keyboard strokes)
- SFX should not be animal sounds (e.g. roaring or steps)
- SFX should not be human sounds (e.g. voices)

While especially vocal feedback would aid in transferring information via speech, we chose to not include it in our sample as past research has shown that persuasive messages or persuasive dialogues can influence motivation and user experience [5, 72]. Consequently, we would not be able to pinpoint whether potential effects on our dependent variables originate from the SFX or from the spoken messages. Hence, by excluding vocal feedback, we excluded a confounding factor and focused on the effects of sound only.

On this first set, the first and second author independently rated each SFX as either *suitable* or *not suitable* for applying the sound to either points or badges. The raters ended up with an agreement rate of 71%. Sounds that were not deemed suitable by both raters were excluded from the sample set. In a second round, both raters discussed all sounds with diverging opinions until they reached an agreement that classified them as either suitable or not suitable. The final SFX set consisted of 17 sounds, which can be found in the appendix of this work. For detailed information on creator, source and license see appendix Table 6. Each SFX was furthermore edited to ensure equal loudness among the whole set.

3.2 Sound Validation Survey

To validate our selection of SFX and to find one representative for each game element from our set, we conducted a validation survey in the form of an online survey, in which participants could rate the set of sounds in terms of how fitting they thought each was for either points or badges.

3.2.1 Method & Procedure. During the survey, participants first performed a sound calibration task. For this, participants were asked to note down six consecutive numbers that were listed by a voice recording in ascending volume, similar to the process described by Altmeyer et al. [2]. This ensured that participants' audio device was turned on and set to a similar volume level across participants. Afterwards, an introduction to the concepts of "points" and "badges" was made through examples from Duolingo[3], a gamified app for learning languages, and Relevo[4], a gamified app for renting sustainable food containers. After completing this step, participants were asked to rate each of the 17 sounds on a Likert scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*) for the statements "*The sound played matches the element points*", "*The sound played matches the element badges*" and "*I associate the sound played with a positive event*". Each sound evaluation was followed by an assignment task; participants were asked to assign the sound to the element points or badges. If they were unable to assign a specific element, participants could respond with the option "Cannot tell". This last question was intended to directly compare sound suitability for points and badges by asking participants to make a clear choice in either direction, in case the prior items resulted in the same rating for both. At the end, participants completed a demographic questionnaire asking for their age, gender and nationality. The study was approved by the Ethical Review Board of the Faculty of Mathematics and Computer Science at Saarland University (No. 22-03-2).

[3] https://en.duolingo.com/, last accessed on August 15, 2024
[4] https://relevo.de/, last accessed on August 15, 2024

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Points	Mean	5.94	4.44	5.30	4.16	4.02	5.84	3.80	4.08	5.30	4.68	4.88	4.58	5.48	4.50	5.06	4.56	5.74
	SD	1.33	1.57	1.76	1.68	1.82	1.15	1.56	1.95	1.36	1.57	1.62	1.65	1.18	1.54	1.52	1.79	1.26
	Q1	5.00	3.00	4.00	3.00	2.25	6.00	2.00	2.25	5.00	4.00	4.00	3.25	5.00	4.00	4.00	3.00	5.00
	Q2	6.00	5.00	6.00	4.00	4.00	6.00	4.00	4.00	6.00	5.00	5.00	5.00	6.00	5.00	5.50	5.00	6.00
	Q3	7.00	6.00	7.00	5.00	5.00	7.00	5.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	7.00
Badges	Mean	3.36	4.92	3.34	5.32	4.70	4.18	5.96	3.18	4.52	5.78	5.18	5.88	3.54	5.26	4.82	3.16	4.24
	SD	1.61	1.66	1.38	1.61	1.64	1.57	1.16	1.65	1.62	1.39	1.37	1.04	1.46	1.43	1.60	1.36	1.57
	Q1	2.00	4.00	2.00	5.00	4.00	3.00	6.00	2.00	3.00	6.00	5.00	6.00	2.00	4.00	4.00	2.00	3.00
	Q2	3.00	5.50	3.00	6.00	5.00	5.00	6.00	3.00	5.00	6.00	5.00	6.00	4.00	6.00	5.00	3.00	4.50
	Q3	5.00	6.00	5.00	6.00	6.00	5.00	7.00	4.00	6.00	7.00	6.00	6.75	5.00	6.00	6.00	4.00	5.75
Choice	Points	46	20	38	8	13	42	5	20	29	8	20	8	36	11	18	29	34
	Badges	2	25	3	34	24	6	42	8	12	34	23	37	5	29	21	3	9
	-	2	5	9	8	13	2	3	22	9	8	7	5	9	10	11	18	7

Table 1. Mean, Standard Deviation (SD), 25th (Q1), 50th (Q2) and 75th (Q3) percentile of all 17 sound (A–Q) ratings on a scale from 1 to 7 regarding their fit to Points and Badges, as well as the opposing choices made for both elements.

3.2.2 Participants. In total, 50 participants were acquired via Prolific[5], each receiving the equivalent of 1.25£ for a study lasting approximately 10 minutes. When asked about their gender, 32 participants self-identified as female, and 18 as male. In terms of age, six were between 18–24, sixteen between 25–31 years old, eight between 32–38, six between 39–45, three between 46–52, five between 53–59 and six were 60 or older. Regarding nationality, the bulk of our participants were British (37), with other participants being French (2), Turkish (2), American (2), Nigerian (1), Nepalese (1), Malaysian (1), Irish (1), Italian (1), Polish (1), or Greek (1).

3.2.3 Results. In terms of positivity, the median rating for each sound was 4 out of 7. The median matching rate for each sound was 4 out of 7 for either points or badges, indicating that none of the SFX from our set were unsuited to be assigned to one of the two game elements. To select the most representative sound for each element, we looked at their matching ratings and the element participants would assign them to. Table 1 shows the descriptives for each sound effect (labeled A through Q) and each matching rating (points, badges, positivity).

For points, participants assigned the highest average matching rating to sound A: With a mean agreeableness of 5.94, a total of 22 participants strongly agreed that this sound matched the game element points, while 14 agreed, 9 slightly agreed, 2 neither agreed nor disagreed and 3 disagreed. For badges, participants assigned the highest matching rating to sound G. Here, the mean agreeableness was 5.96 out of 7, and a total of 17 participants strongly agreed that this sound matched, while 23 agreed, 5 slightly agreed, 3 neither agreed nor disagreed and 2 disagreed. Following these results, we selected sound A for points and sound G for badges and used them subsequently from here on.

4 Quantitative Study

Based on open questions outlined in the summary of related works introduced in Section 2, we aimed to systematically investigate the influence of SFX on user experience and performance in gamification. To this end, we used an image-tagging task to compare no gamification, visual gamification, visual gamification with sound effects (audiovisual) and sound effects only (audio). We pose the following hypotheses (H):

H1 Participants with visual gamification tag more than participants without any gamification.
H2 The presence of an auditory modality has an influence on the number of tags.

[5] https://www.prolific.co/, last accessed on August 15, 2024

H3 The gamification modality (audio, visual, audiovisual) influences user experience (SAM, IMI, AFSS).

Since we are not aware of past research comparing visual and auditory representations of game elements in the context of gamification, we could not derive a priori assumptions about the direction of our expected effects regarding modalities beyond visual. Hence, to operationalize this uncertainty, **H2** and **H3** were established as non-directional hypotheses. **H1** would be a replication of prior gamification research, which showed that the presence of visual game elements led to an increase in motivation. In prior image-tagging studies (e.g. [52] or [69]), this was shown by participants submitting significantly more tags in the respective gamified conditions compared to the control groups. With our study, we expect to replicate those results. **H2** follows past research, where audio was shown to influence user performance in gameful contexts [12, 75]. This influence of audio on performance cannot only be motivated by past empirical findings, but also by existing theoretical work. For instance, the Arousal-Mood Hypothesis states that audio affects arousal and mood, which in turn influence performance [32]. Thus, by replacing visual gamification elements through auditory representations, we expect to find effects on performance. Lastly, **H3** is based on empirical works in the context of gameful systems, wherein audio was shown to influence user experience and perception in terms of flow and enjoyment [14, 29, 45]. **H3** is further supported through Kahneman's Capacity Model of Attention [35]. According to this model, attentional capacity is a limited cognitive resource which is dynamically allocated depending on the task. Attention and the affective state of arousal are thereby tightly coupled. As we expect the different combinations of visual and auditory stimuli to affect the attention of participants, we also expect to find effects on the affective measures of the user experience. To measure potential impacts on affective experience, we use the Self-Assessment Manikin (SAM) [9]. SAM has an own factor for valence and arousal, which seem particularly interesting to consider, based on Kahneman's Capacity Model of Attention. Furthermore, to measure effects on intrinsic motivation, we use the 22-item version of the Intrinsic Motivation Inventory (IMI) [51, 66] The IMI was used since past research has found that gamification affects motivation [53]. This is explainable through Self-Determination Theory [67]: gamification elements (such as Points or Badges) may act as rewards which in turn can affect different types of motivation. Since we investigate gamification elements represented through auditory and visual modalities, we expected to find effects on motivation, according to past research and Self-Determination Theory. Thus, we consider motivation an important part of the user experience. Lastly, we use the Activity Flow State Scale (AFSS) [58, 59] to measure the effects on flow experience induced by different gamification settings. Flow—the experience of being fully absorbed in a task [17] was chosen as past research has empirically demonstrated that audio can enhance flow experiences in games [45, 56]. Also, feedback more generally has been described as a central antecedent of flow [41]. Thus, given that we both have a game-like context and provide feedback through auditory and visual modalities, flow seems an important aspect to consider when investigating the user experience. The exact procedure as well as details on the scales and subscales used are described in the following.

4.1 Apparatus

Similar to prior research (for example, [44], [52] or [69]), we use an image-tagging platform to measure performance quantitatively through the amount of tags submitted by users. We implemented the platform as a web application in a way that allowed us to define several conditions with the game elements points and badges, with or without SFX, or no game elements at all. Figure 3 shows a screenshot of the platform displaying an image to the user and Figure 1 shows the tagging process, with the tag input field on the left, and the game elements points and badges on the right.

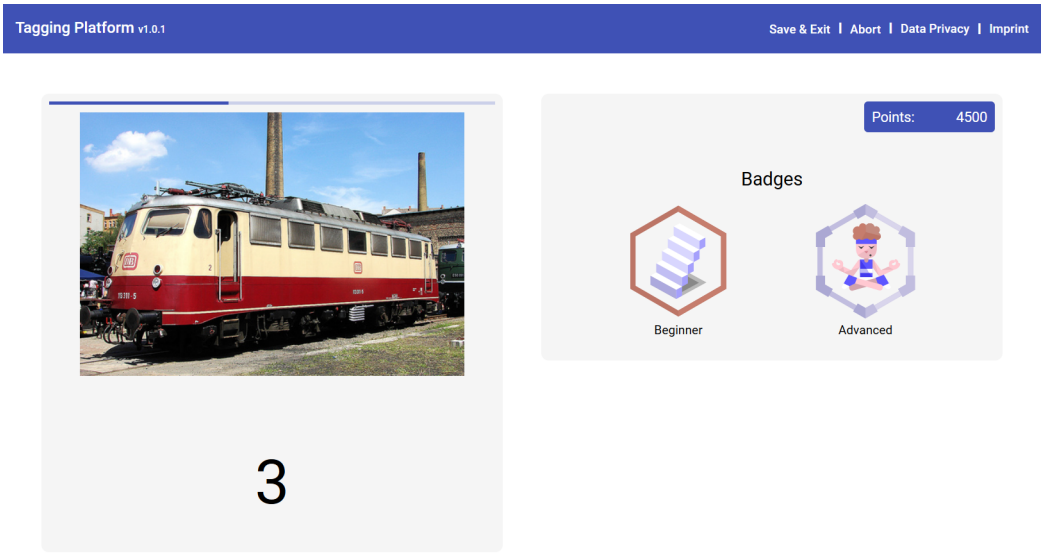

Fig. 3. The tagging platform showing an image on the left, and the gamification elements points and badges on the right.

Similar to [3] the images presented to the users were taken from the MIRFLICKR-25000 image collection [31]. While tagging, a badge was unlocked after entering 15, 45, 75 and 105 tags. This was done in accordance to work by Schubhan et al. [69] in which the game elements aimed to motivate approximately 100 tags over 15 images. In addition, we integrated a tutorial that explained the tagging task, as well as the game elements if applicable. Questions could be prompted via the questionnaire view, such that we were able to define questions at any given point during the study. Upon entering the platform, a welcome page greeted the user, explaining the study and giving an overview on data privacy, which had to be accepted before the study would start.

4.2 Method & Procedure

In this study, we designed four between-participants conditions for the image-tagging platform: One with no gamification at all (*Baseline*), in which participants would only tag images without seeing game elements nor hearing sounds; one with visual gamification but no sound (*Visual*), i.e. participants would see a visualization of points and badges next to the tagging task; one with sound only (*Audio*), wherein participants would hear the sound for points and badges without a visual representation; and, lastly, one where they would both see and hear the gamification elements (*Audiovisual*).

Similar to our sound validation survey and analogous to Altmeyer et al.'s [2] study on sounds in gamification, we opted for an online-study in order to secure a reasonable sample size. After giving their consent for participation on the landing page, participants in the sound conditions would first go through a volume calibration process, as explained in Section 3 and then continue with a tutorial. The *Baseline* and *Visual* conditions skipped the volume calibration and started with the tutorial right away. The tutorial explained the functionality of the platform, how to add and edit tags, and that they would see each image for five seconds, before it vanishes and the tagging starts, similar to prior work [52]. For each of the gamified conditions, participants would additionally receive a tutorial explaining the game elements, or the purpose of the sounds in the auditory condition

respectively. The tutorial contained three trial images to tag. After that, a notification informed the participants that the main run was about to start, wherein they would tag another 15 images in their respective condition. Once they finished the main task, they were asked to answer the SAM [9] items *Valence* and *Arousal*, each on a scale from 1 to 9, the IMI [51, 66] subscale items for *Choice*, *Competence*, *Enjoyment* and *Pressure*, rating each item on a seven point Likert-scale ranging from *Not at all true* (1) to *Very true* (7), as well as the AFSS [58, 59], measuring *Merging Action and Awareness*, *Goal Clarity*, *Concentration on Task at Hand*, *Unambiguous Feedback*, *Challenge-Skill Balance*, *Transformation of Time*, *Sense of Control*, *Loss of Self-Consciousness* and *Autotelic Experience*. Here all items had to be rated on a five point Likert-scale ranging from *Strongly disagree* (1) to *Strongly agree* (5). At the end, a short questionnaire on demographics similar to that described in section 3 including age range, gender and nationality concluded the study. The study was approved by the Ethical Review Board of the Faculty of Mathematics and Computer Science at Saarland University (No. 22-06-5).

4.3 Participants

Overall, 120 participants took part in the study, equally distributed among all conditions. As in Section 3, they were hired via Prolific, receiving 2.25£ for a study duration of approximately 18 minutes. Similar to Schubhan et al. [69], we applied Tukey Fences [78] for outlier detection on the amount of tags submitted in order to exclude data points with a noticeably higher or lower tag count per condition. In these cases, the motivation, or lack of motivation respectively, might have been influenced by factors outside of the scope of the presented task and gamification elements. This led to three exclusions each in the *Baseline*, *Audio* and *Audiovisual* conditions, as well as two in the *Visual* condition, leaving 27 participants in *Baseline*, *Audio* and *Audiovisual*, and 28 in *Visual* respectively, thus 109 in total. 60 of them identified as female, 46 as male, 2 as non-binary and 1 preferred to not state their gender. 9 participants were 18-24 years old, 23 were 25-31, 22 were 32-38, 21 were 39-45, 13 were 46-52, 20 were 53-59 and 1 participant was 60 years or older. Most of the participants were British (92), Scottish (4) or American (3) with the remaining 10 participants each representing another individual nationality.

4.4 Study Results

During our analysis, we ran ANOVAs on our data to compare all conditions overall, and pairwise comparisons if significant differences were found. We performed a Kruskal-Wallis ANOVA on the IMI subscales Choice (χ^2=0.91, p=0.82), Enjoyment (χ^2=5.49, p=0.14) and Pressure (χ^2=2.02, p=0.57), as the data was not normally distributed. Competence was normally distributed with homogenic variances, but here Fisher's ANOVA as well did not yield significant results (F=2.16, p=0.10). Thus, we were not able to find a significant difference regarding intrinsic motivation across the conditions. For the SAM subscales, we ran two more non-parametric Kruskal-Wallis ANOVAs. Both Valence (χ^2=3.05, p=0.38) and Arousal (χ^2=1.74, p=0.63) did not pass the significance threshold. In the following, we outline our results R1 through R7 on tag quantity and flow perception.

4.4.1 Tag Quantity. The average number of tags submitted can be found in Table 2. Participants in the *Visual* (V) condition submitted the most tags with an average of 107 tags, followed by *Audiovisual* (AV), *Audio* (A), and, lastly, *Baseline* (BL). A Shapiro-Wilk test showed that the assumption of normality was met for tag quantity data (W=0.99, p=0.28), while Levene's test showed a significant difference regarding homogeneity of variances (F=5.17, p<0.01). Hence, we ran Welch's One-Way ANOVA, which indicated significant differences between the conditions (F=3.3, p=0.03). The results of the Tukey corrected Games-Howell post-hoc tests are shown in Table 4. This shows a significant difference between the *Baseline* and *Visual* conditions, replicating existing gamification research,

	BL	V	A	AV	
n	27	28	27	27	
Number of tags	79.2 / 24.5	107.0 / 45.8	80.9 / 28.7	93.4 / 33.9	
Choice	5.77 / 1.11	5.79 / 0.93	5.69 / 0.95	5.96 / 0.80	IMI
Competence	4.10 / 1.59	4.65 / 1.52	4.07 / 1.08	4.80 / 1.00	IMI
Interest / Enjoyment	4.60 / 1.59	5.22 / 1.15	4.49 / 1.23	5.02 / 1.65	IMI
Pressure / Tension	2.41 / 1.61	2.34 / 1.34	2.62 / 1.25	2.41 / 1.22	IMI
Merging Action and Awareness (MMA)	2.88 / 0.92	3.25 / 0.94	3.05 / 0.90	3.36 / 0.84	AFSS
Clear Goals (CG)	3.91 / 0.75	3.86 / 0.64	3.64 / 0.69	4.19 / 0.68	AFSS
Concentration on Task at Hand (CO)	4.44 / 0.55	4.19 / 0.74	3.97 / 0.65	4.42 / 0.54	AFSS
Unambiguous Feedback (UF)	2.94 / 1.05	3.63 / 0.88	3.00 / 0.83	3.81 / 0.61	AFSS
Challenge Skill Balance (CS)	3.40 / 0.93	3.67 / 0.83	3.36 / 0.59	3.78 / 0.44	AFSS
Transformation of Time (TT)	2.58 / 1.16	2.88 / 1.07	2.62 / 1.05	2.79 / 1.12	AFSS
Sense of Control (CN)	3.67 / 0.97	4.04 / 0.93	3.54 / 0.84	4.19 / 0.68	AFSS
Loss of Self-Consciousness (SC)	3.22 / 1.01	3.74 / 1.13	3.43 / 0.87	3.85 / 0.85	AFSS
Autotelic Experience (AE)	3.33 / 1.04	3.54 / 0.81	3.01 / 0.75	3.53 / 0.93	AFSS
Arousal	4.07 / 2.18	4.71 / 2.16	4.26 / 2.10	4.15 / 1.81	SAM
Valence	6.67 / 1.59	7.07 / 1.51	6.52 / 1.34	6.89 / 1.42	SAM

Table 2. Means and standard deviations of the amount of tags, the Intrinsic Motivation Inventory (IMI), Affective Flow State Scale (AFSS) and Self-Assessment Manikin (SAM) over the Baseline (BL), Visual (V), Audio (A) and Audiovisual (AV) conditions.

	IMI				AFSS									SAM	
	Choice	Competence	Interest	Pressure	MAA	CG	CO	UF	CS	TT	CN	SC	AE	Arousal	Valence
F	-	2.16	-	-	1.52	-	-	-	3.37	-	-	-	-	-	-
χ^2	0.91	-	5.49	2.02	-	9.19	10.11	19.13	-	1.07	12.47	7.26	6.97	1.74	3.05
p	0.82	0.10	0.14	0.57	0.21	**0.03***	**0.02***	**<0.01***	**0.03***	0.78	**<0.01***	0.06	0.07	0.63	0.38

Table 3. ANOVA results of the Intrinsic Motivation Inventory (IMI), Affective Flow State Scale (AFSS) and the Self-Assessment Manikin (SAM) subscales comparing all conditions. P-values ≤ 0.05 are marked with a *.

Tags

	BL	V	A	AV
BL	-	**0.03***	1.00	0.31
V		-	0.07	0.59
A			-	0.47

CG

	BL	V	A	AV
BL	-	0.78	0.56	0.56
V		-	0.78	0.18
A			-	**0.02***

CO

	BL	V	A	AV
BL	-	0.58	**0.03***	0.80
V		-	0.45	0.59
A			-	**0.05***

UF

	BL	V	A	AV
BL	-	**0.02***	1.00	**<0.01***
V		-	**0.02***	1.00
A			-	**<0.01***

CS

	BL	V	A	AV
BL	-	0.66	1.00	0.23
V		-	0.39	0.92
A			-	**0.02***

CN

	BL	V	A	AV
BL	-	0.19	0.94	0.09
V		-	**0.05***	0.94
A			-	**0.02***

Table 4. Post-hoc pairwise comparisons of the significant ANOVAs and the respective (adjusted) p-values.

although none of the other pairwise comparisons indicate a significant difference. Thus, we posit result ***R1: Participants with visual gamification produced significantly more tags than***

those who tagged without gamification and ***R2: Participants with SFX did not produce significantly more tags than those without SFX***.

4.4.2 Flow. The descriptive statistics for the AFSS subscales are also shown in Table 2. We ran ANOVAs on all subscales, using a Fisher's ANOVA on normal data with equal variances (reported with an F-statistic), Welch's ANOVA on normal data with unequal variances (reported with an F-statistic), and a non-parametric Kruskal-Wallis ANOVA on non-normal data (reported with a χ^2-statistic), as shown in Table 3. The ANOVAs on *Merging Action and Awareness* (MAA), *Transformation of Time* (TT), *Loss of Self-Consciousness* (SC) and *Autotelic Experience* (AE) were not significant, while those on *Clear Goals* (CG), *Concentration on Task at Hand* (CO), *Unambigous Feedback* (UF), *Challenge Skill Balance* (CS) and *Sense of Control* (CN) all showed significant results with $p \leq 0.05$. The post-hoc pairwise comparisons for Welch's Anova (CS) were calculated using the Games-Howell method with Tukey correction and the Dunn method with Bonferroni-Holm adjustment for Kruskal-Wallis ANOVA (CG, CO, UF and CN). Looking at the pairwise results of CG, we found a significant difference between *Audiovisual* and *Audio* (z=2.93, p=0.02), leading us to derive ***R3: Participants who tagged with audiovisual gamification found goals to be clearer than those who tagged with SFX only***. In terms of CO, there is a significant difference between *Baseline* and *Audio* (z=2.86, p=0.03), as well as *Audiovisual* and *Audio* (z=2.61, p=0.05), indicating that ***R4: Participants who tagged without gamification or with audiovisual gamification felt more concentrated than those who tagged with SFX only***. UF showed significant differences between *Baseline* and *Visual* (z=−2.68, p=0.02), *Baseline* and *Audiovisual* (z=−3.32, p<0.01), *Visual* and *Audio* (z=2.80, p=0.02), as well as *Audiovisual* and *Audio* (z=3.44, p<0.01). Here, we can conclude ***R5: Participants with auditory or no gamification perceived the feedback as less unambiguous than those with visual and audiovisual gamification***. The CS subscale showed a significant difference between *Audio* and *Audiovisual* (t=−2.96, p=0.02), leading to ***R6: Participants who tagged with audiovisual gamification found the Challenge-Skill-Balance to be better than those who tagged with SFX only***. Lastly, CN showed significant differences between *Visual* and *Audio* (z=2.58, p=0.05) as well as *Audiovisual* and *Audio* (z=3.00, p=0.02), resulting in ***R7: Participants who tagged with audiovisual gamification had a greater Sense of Control compared to those who tagged with SFX only***.

5 Qualitative Study

To better understand the underlying reasons for our results in the quantitative study (and the absence of effects), we complemented our statistical results from Section 4 by conducting one more study following a qualitative approach. In the following, we will describe the procedure, methodology, and results.

5.1 Method & Procedure

The conditions remained the same as before, i.e. there was a *Baseline*, *Visual*, *Audio* and *Audiovisual* condition. Unlike in the previous study, participants faced all conditions in a within-participants design, and each condition was followed by a semi-structured interview. This design allowed us to ask questions not only about each condition individually, but also to gain information on comparisons between them. A translation of the interview script can be found in Appendix B of this work. In order to not overwhelm participants with four times the amount of pictures to tag, we reduced this to three per condition, allowing participants to still tag several pictures and also to grasp visual and auditory changes. All participants started with the *Baseline* condition and a tutorial, followed by the other three conditions in Latin Square randomized order to make sure that the counter-balancing led to as much diversity in the condition succession as possible. The

interviews were audio recorded and later transcribed for further analysis. In the end, participants concluded this study with a questionnaire on demographic data, i.e.: age range, nationality and gender.

5.2 Participants

The study got approval from the Ethical Review Board of the Faculty of Mathematics and Computer Science at Saarland University (No. 22-08-1). This time, we recruited a convenience sample from the social network of the authors. None of the participants took part in any of the earlier studies. A total of nine participants took part in this study, whereof three identified as female and six as male. Seven were German, one Belgian and one Pakistani. One was between 18-24 years old, six between 25-31 and two between 32-38. Participants were not compensated for participation and the study took approximately 45 minutes to complete. Of this, the average interview lasted 28.03 minutes with a transcription word count of 3273.56. Seven interviews were held in German, two in English to adhere to the participants' language preference.

5.3 Coding Process

To prepare the interviews for the analysis, we followed a hybrid thematic analysis approach using a codebook, combining both inductive and deductive coding [25]. The coding process was guided by the underlying question why users prefer visual, auditory or audiovisual depictions of gamification, representing the deductive aspect of the analysis. However, we also allowed new codes to be created when analyzing transcripts, adding an inductive component to the process. We proceeded with the following coding procedure: The first and second author both independently read through the same four interviews first, tried to find reasons for why users preferred a certain feedback modality and identified any further aspects they deemed relevant to answer the underlying question mentioned before or noticed to reoccur throughout the transcriptions. Transcripts were processed in their original language as both coders were fluent in English and German. Afterwards, both authors compared their codes, discussed differences and transferred them into a codebook, including explanations when a code was used. From here on, this codebook was used as a foundation and general guide for the remaining interviews. After all interviews were coded, both authors ended up with an inter-rater agreement of $\kappa = 0.72$ (Cohen's κ). Emerging differences were again discussed, resulting in the final set of codes and code counts. Table 5 shows an overview over our final codebook. Finally, following Fereday and Muir-Cochrane [25] both authors collaboratively identified common themes that cluster the codes into coherent groups. To this end, we followed an iterative approach, where both, the first and second author, repeatedly grouped the codes and labeled them, until consensus was reached and all codes were assigned a group. With each iteration we refined the grouping and labels until we could fit all codes into coherent groups and match all groups with a label, or theme respectively. While some of the themes hold contradictory codes (e.g. "SFX motivating" vs. "SFX not motivating"), it is important to note that themes in thematic analysis can hold contradictory codes [10, 13]. The emerged themes will be discussed in the following.

5.4 Interview Results

In the following sections, we focus on specific themes that we shaped in our analysis regarding the perceptional and motivational aspects of the visual game elements, the auditory game elements, as well as the combination of both. As the majority of interviews were not held in English, quotes are translated here for comprehensibility.

5.4.1 Theme 1: Visual gamification alone is perceived well and motivating, but may increase distraction, lacks meaning and receives low attention. Regarding the visual game elements, i.e. feedback from

	Code	Description	Count
Theme 1	Aspects of visual gamification perceived well	Participant mentions that they liked some aspect the visual gamification	7
	Visual gamification motivating	Participant mentions that the visual gamification was motivating (e.g. that they added more tags)	4
	Visual gamification important	Participant mentions that the visual gamification is important to them, e.g. to get to know their progress or in general	2
	Visual gamification distracting	Participant mentions that the visual feedback distracted them from the main task	4
	Visual gamification low attention	Participant mentions that they noticed visual gamification, but either did not really care or mentioned that they payed more attention when having SFX	4
	Indifference towards visual gamification	Participant mentions that the do not really care about visual gamification	2
	Additional visual feedback wanted	Participant mentions that visual gamification could be enhanced by adding more /other feedback	6
	Badge motivating	Participant mentions that the visual badge was motivating	4
	Points not motivating	Participant mentions that the points are not motivating/they did not like them/did not matter	2
Theme 2	Aspects of SFX perceived well	Participant mentions that they liked some aspect of SFX, without providing further reasons/details	6
	SFX motivating	Participant mentions the the SFX feels rewarding to them	5
	SFX not motivating	Participant mentions that the SFX had no or little effect on their motivation (e.g. to tag images)	1
	SFX high attention	Participant mentions that the SFX captured their attention	6
	SFX low attention	Participant mentions that they did not pay much attention to the SFX	1
	SFX irritating	Participant mentions that the SFX were irritating/that they did not understand the semantics for them	6
	SFX distracting	Participant mentions that the SFX distracted them from the main task	2
	SFX less distracting	Participant mentions that the SFX is less distracting/minimalistic/less overwhelming	2
	Subconscious SFX awareness	Participant mentions that the SFX played in their head even after it was gone	1
	Indifference towards SFX	Participant mentions that they do not really care about SFX	1
	SFX does not convey information	Participant mentions that the SFX did not convey information, e.g. on progress	3
	Negative SFX wanted	Participant mentions that they would like to have a negative SFX when removing tags	2
	SFX for Badge motivating	Participant mentions that the SFX for Badges was motivating, e.g. made them add more tags	1
	SFX increase flow	Participant mentions that the SFX increased their flow experience	1
Theme 3	SFX increases pressure	Participant mentions that SFX increased pressure	1
	Context matters for SFX	Participant mentions that the context is important to judge suitability of SFX	1
	SFX for Points unpleasant	Participant mentions that the SFX for points was too sharp/high/loud	7
	Dislike repetitive SFX for Points	Participants mentions that they do not like that the points SFX is played so often/think the Points SFX is repetitive	4
	SFX for Points annoying	Participant mentions that SFX are annoying	3
	SFX for Badge not well received	Participant mentions that they did not like the Badge sound	1
	SFX for Badge well received	Participant appreciates the SFX used for badges	6
Theme 4	Lack of visual gamification	Participant mentions that they were missing visual feedback that the gamification provided	3
	Lack of SFX	Participant mentions they were missing SFX	3
	Visual and SFX complement each other	Participant mentions that the SFX supported the visual feedback or vice-versa	5
	SFX+Visual gamification overwhelming	Participant mentions that the combination of both is overwhelming	1
	Visual dominance	Participant mentions that the SFX were less salient and that they payed more attention to the visual gamification	1

Table 5. Codebook with semantic codes and their counts based on the interviews.

the *Visual* and *Audiovisual* conditions, most participants (7) mentioned that they liked aspects of the visual game elements like their presentation, animations etc. Two found the visual elements to be important for the tagging task. We can conclude from this that overall they were perceived well and four participants specifically mentioned that they found the game elements to be motivating for the task, especially the badges (4), as for example P1 said: "*I noticed that I tried to enter a lot to unlock the badge [...]*". In contrast, two participants did not find the points element to be motivating, as they missed a purpose and found the amount of points rewarded arbitrary. On the downside, four

participants found the game elements to be visually distracting since they were already performing a visually demanding task and paying attention to the elements would distract them from it. P3 explains that "*all the badges just distract from your task and stuff. The thing is you're visually tagging something. So I look at the picture, I have to remember what the picture is about and have to come up with whatever is in this picture, right? If I see a badge up here, I get other visual information and I'm distracted by this [...]*". Hence, visual gamification in combination with a visually demanding task could lead to unwanted distractions. Another four interviewees highlighted that they did not pay a lot of attention to the visual elements as they either did not care enough about them or were already familiar with them from a prior condition. Two participants even felt indifferent towards the game elements. Lastly, six interviewees expressed wishes to extend the game elements further. Here, they referred to either adding more celebratory effects, such as confetti or fireworks, or adding more game elements, such as progress bars.

5.4.2 Theme 2: SFX alone may increase attention and are perceived as motivating, but can cause confusion. Similar to the visual game elements, six participants mentioned that they liked aspects of the SFX conditions *Audio* and *Audiovisual* with five highlighting that they found the SFX specifically to be motivating. P9 emphasized the badge sound being motivating and that the SFX for points "*might be a thing with rhythm. I'd argue that you come into a state of flow. You enter something and immediately receive feedback, enter another thing, receive feedback. The system is like a loop*". We could conclude from this that the SFX create a sense of flow (which aligns well to the results from the quantitative study presented before). Only P4 pointed out that SFX were not motivating for them. Unlike the visual elements, six interviewees reported that the SFX had a high impact on their attention, while only one mentioned not paying much attention to the SFX, meaning that sound effects could be specifically used in scenarios, where high attention matters. Compared to the visual elements, only two pointed out that they felt distracted by the SFX, which is opposed by two other interviewees mentioning that they felt less distracted from the main task, because of the SFX. Consequently, there seems to be no clear indication of the SFX' distractiveness, but in comparison to visual game elements, distraction seems to be less of a drawback and could even be reduced for some users. Only P2 felt indifferent towards the SFX, as they found them not to be relevant.

Another important factor are the semantics of a sound. Six participants were irritated by the sound, mostly in the *Audio* condition, without a visual component, as they were not sure what the sound implied while performing the tagging task. Three furthermore supported this by pointing out that they were missing information being inherently conveyed by the sounds. P2 for example mentioned that "*there's no information coming with the sound.*". P8 further confirms this: "*I think probably it was a sound to give the feedback that four tags are optimal for this image, which is nice, but I don't know what it was meant to be*". While P8 in this case already expresses uncertainty towards the meaning of the SFX, their possible explanation is also incorrect. Another point directed towards the meaning of the SFX, was brought up by two participants missing a specifically negatively connoted sound when removing a tag. We conclude that it is important to provide semantics with the sounds themselves. As several participants compared the points SFX to collecting coins in Super Mario, one could for example similarly increase the pitch of the sound for every tag entered, indicating an increase in points.

5.4.3 Theme 3: How frequent SFX are played as well as attributes such as loudness and pitch play a role in how SFX are perceived. Overall, the SFX choice for badges was well received by most participants (6). In contrast only one participant did not like it. The sound for points on the other hand received critiques, such as being too sharp or high (7) and being repeated too frequently during the task (4). Three participants specifically mentioned that they perceived the points sound

as annoying (e.g. P3: "*I noticed a very annoying bling, which was very sharp*"). While this contradicts our findings from the sound validation survey regarding the sound choice, the critique about the sound repetition could already hint at the reason for this feedback: While the online survey aimed at finding a generally fitting sound for points, it was ultimately implemented in a more specific way and context, which could have influenced the fit of the sound for points. This was also highlighted by one of the participants, who specifically mentioned that the sound was not fitting for the context it was used in. One participant also felt more pressured due to the SFX, which might be explicable by the statement of P1, who said: "*[...] whenever I was typing, sometimes when I was thinking, I still had the [points SFX] in my head and payed more attention to it subconsciously than I thought*". This subconscious awareness might have also been caused by the high frequency with which the SFX for points were played.

5.4.4 Theme 4: Visual gamification and SFX complement each other and may cancel out each other's disadvantages. Regarding the combination of both SFX and visual game elements, five participants highlighted a synergy between both approaches, i.e. the sound effect supporting the visuals and vice-versa. P7 for example, highlighted that "*whenever I heard the longer, positive sound, the badge appeared and that... that supported visually what I already perceived auditory. And I liked that*". P8 additionally mentioned that "*every tag is accompanied by some sound. It's always good to have... for the feedback you don't have to focus your attention and look right. So, it's always you would hear it.*". When taking Theme 1 and Theme 2 into account, the quotes from P7 and P8, alongside the opinion of the other three interviewees, hint at the benefits of combining visuals and audio. In Theme 1, we learned that visual game elements can be distracting, while in Theme 2, we learned that SFX alone can increase attention, but they can also have an unclear meaning and be misinterpreted. In combination, the visual game elements might implicitly explain the SFX to the user, while the SFX allow them to focus more on the visual task. Interestingly, although especially the points sound was criticized by most interviewees, three missed the sound effects when experiencing the *Visual* condition after *Audio* or *Audiovisual*, and another three mentioned the same about the visual components when experiencing *Audio* after *Visual* or *Audiovisual*. Only one participant felt overwhelmed by the combination of both, and one experienced a visual dominance effect, i.e. the SFX were less noticeable when combined with the visual elements.

5.4.5 General User Preferences. At the end of the interviews, participants were asked about their most preferred and least preferred condition. They could name multiple if they were not able to make a singular choice. Seven out of nine participants reported *Audiovisual* to be their favorite condition. Two favored *Audio* and *Visual* respectively and only one favored *Baseline* with no gamification element at all. On the contrary, five participants least preferred *Baseline*, four *Audio*, one *Audiovisual* and no one answered with *Visual*. These answers match the themes found above. The symbiosis hypothesized through Theme 4 might explain why *Audiovisual* was favored by most participants, while the lack of any game elements explains why *Baseline* was least favored. The disfavor towards *Audio* can be explained by Theme 3 and the general perception of the SFX for points.

When asked which condition they payed the most attention to, six named *Audiovisual* and four *Audio*, while *Baseline* and *Visual* were not mentioned at all, supporting Theme 2 that the addition of sounds increases user attention.

Lastly, we asked them which of the two factors, SFX or visuals, they find to be most dispensable for the task given. Five participants found the visual elements to be dispensable, four the SFX component. Hence, no clear direction can be given for contexts where only one of either is applicable. Yet considering Theme 4 and that *Audiovisual* was favored by most participants, a combination of both seems desirable, if applicable.

6 Discussion

In this section, we discuss our insights from three user studies on the impact of auditory and visual modalities in gamification. The studies include a *sound validation survey* to identify appropriate SFX for points and badges, a *quantitative* study comparing the effects of auditory and visual gamification on task performance and user experience, and a *qualitative* study delving deeper into participants' perceptions of each modality.

6.1 SFX as Gamification Element

After screening potential sound effects to represent points and badges, we found that two SFX were particularly well-received by participants, one representing points and one representing badges. This indicates that SFX alone could serve as a rewarding feedback to users, adding further support to the assumption made by Robinson et al. [62] that SFX themselves can be considered gamification elements. From a sonification perspective, both SFX could be seen as earcons. The SFX for points falls into the category of *single-pitch earcons*, which "can be used to represent simple, basic, or commonly occurring user-interface entities" [7], and the SFX for badges can be seen as a more complex *compound earcon* as two or more audio elements are placed in succession [7], adding further theoretical ground to their capability to transfer information to users. Outside of games and gamification, Salimpoor et al. [68] found that auditory pleasure is tied to the release of dopamine, which in turn means that SFX can become rewarding to users. This further supports the potential of SFX in representing gamification elements. On a more general level, this aligns well to ongoing gamification research finding that gamification elements are not necessarily tied to visual stimuli, but can also be rewarding when mediated through different modalities, such as haptic feedback or a combination of several modalities targeted by physical representations of game elements [1, 18].

6.2 Visual Gamification Increases Task Performance: Replication of Past Research

But how do auditory representations of game elements compare to the much more frequently used visual ones? To contribute answers to this question, we performed an online study quantifying these differences. We found that participants with visual gamification provided significantly more tags than those without gamification (**R1**). This finding aligns well to existing gamification research in the context of image tagging, which has found the same effect of visual gamification on the number of generated tags [2, 44, 52, 53, 69]. Thus, our study replicates previous research and adds support to the effectiveness of (visual) gamification in this context, supporting **H1: Participants with visual gamification tag more than participants without any gamification**.

6.3 Absence of Effects on Task Performance in Auditory and Audiovisual Gamification

However, for participants in auditory gamification or audiovisual gamification conditions, we could not find such effects (**R2**) within our sample. Thus, it seems like sound effects alone do not affect participants' tagging performance to a large degree. Surprisingly, we also could not find significant differences in conditions where participants received both, sound effects and visual stimuli, since participants in the audiovisual condition also did not provide significantly more tags compared to no gamification at all. When considering past research, the absence of effects of auditory stimuli in the context of gamification has been reported by others as well: Both Altmeyer et al. [2] and Hicks et al. [29] could not identify a correlation between SFX and performance. While Altmeyer et al. [2] discussed visual dominance [73] as a potential underlying reason, Hicks et al. [29] assumed that the application context had an effect on participants performing well. Both visual dominance and the application context could play a role in our study as well. Since the task of image tagging (context) is inherently a visual one, the principle of visual dominance might have increased the semantic

relevance of the task. In addition, we found that visual gamification was perceived as motivating and, in contrast to auditory feedback, did not cause confusion (**Theme 1** and **Theme 2**). Thus, it seems visual gamification introduced and communicated goals and progress more clearly than auditory gamification to participants. According to *goal-setting theory*, establishing clear goals is directly linked to action and performance [50], potentially explaining why participants provided more tags in visual gamification, but not in auditory conditions. We also learned that the sound effects we used were partially perceived as annoying, due to them being played for every single tag (**Theme 3**), which might have affected the performance especially of participants receiving audio-only feedback. This matches similar complaints about repetitiveness of audio reported in other work [64]. Together with the results from our sound validation survey which selected this specific SFX for points, these results raise questions about the interplay of context and frequency of use in a gamified setting, as the results contradict the generalizability of the selected sound effects. The importance of context has already been highlighted in Hamari et al.'s [28] literature review, albeit for the effectiveness of gamification in general and not specifically for the use of SFX for this purpose. With our outcome, we can further extend this and recommend considering the use and context for the choice of SFX as well. For frequent SFX occurrence as in our context of image tagging for example, a softer and lower pitched sound was recommended for points by some of our interviewees, which also falls in line with effects found in sonification literature, such as [8], where high pitched auditory icons (environmental sounds associated to a virtual object) were perceived as annoying.

That participants in the audiovisual gamification condition, i.e. those receiving visual and auditory feedback, did also not provide significantly more tags than those without gamification seems less intuitive and might not be readily explainable by goal-setting theory. In fact, regarding user experience measures (**R3–R7**) and feedback from the interviews (**General User Preferences**), it seemed as participants liked the combination of both auditory and visual feedback most. Why this did not affect participants' tagging performance to a measurable degree in our study remains unclear. Similar to Altmeyer et al. [2], reasons might be related to the effect size being smaller than we could detect with our sample size. This might be supported by the fact that the tag count in the audiovisual condition is descriptively the second-highest among all conditions. Also, Miranda and Palmer [54] found that increased enjoyment appeared to be independent of overall levels of performance, in a similar study design. Our results support this, since albeit the positive user experience and perception in the audiovisual condition, no positive effects on performance were found. Consequently, we have to conclude that **H2: The presence of an auditory modality has an influence on the number of tags** is not supported by our data.

6.4 Effects on User Experience: Audiovisual Gamification Enhances Flow Experiences

Regarding the user experience of participants (**H3**), we were not able to find effects on intrinsic motivation nor affect, but on the prevalence of flow experiences. Regarding the former, finding no effects on intrinsic motivation is in line with previous research in the context of image tagging [53]. Mekler et al. [53] stated that the motivational appeal of many games lies in their ability to provide players with challenges to master. However, the image annotation task did not offer much of a challenge since participants were free to create as many tags as they wanted. Considering that moderate performance targets do not motivate people to put in much effort, Mekler et al. assumed that gamified feedback does not further encourage participants to achieve more challenging targets and is less likely to fulfill basic psychological needs. A related factor might be that tag quality did not count, i.e. participants were not receiving feedback on whether their tags were a good fit, which might have further reduced their perceived challenge. Potentially, gamified feedback might have a

stronger effect on affect and intrinsic motivation in a different context, as past research has quite frequently demonstrated [36, 38, 84].

However, we found that the different gamification modalities affected flow experience (R3 – R7). First, we found a significant difference between *Audiovisual* and *Audio* (**R3**) regarding the clear goals factor of AFSS. Goals were perceived as more clearly by participants exposed to audiovisual gamification than those receiving auditory gamification only. This difference fits the interview analysis well, which demonstrates some interviewees felt irritated in the *Audio* condition. They were not necessarily aware of the SFX's meaning and its semantics (**Theme 2**), which may have worsened goal clarity for them. This falls in line with a study from the sonification and sports context, where SFX were used to represent basketball player movement, yet, participants in their user study misinterpreted the SFX despite them being explained earlier [23]. Thus, we can derive that making sure users understand the semantic meaning of sound effects, especially when using sound effects without further visual stimuli, is important. One way of achieving this has been suggested in the interviews, i.e., by increasing the pitch of the sound effect for every tag entered to indicate an increase in points.

With **R4**, we found that participants with auditory gamification reported decreased concentration on the task at hand (according to the respective factor of the AFSS) compared to *Baseline* and *Audiovisual*. This could be due to participants being irritated by the SFX, as they had problems interpreting their meaning (**Theme 2**). Thus, it seems the semantic dissonance between a sound effect and its meaning is responsible for the negative effect on concentration on the task. To explain why we found no significant effect regarding concentration on task at hand in *Visual*, we need to first take a look at why *Baseline* and *Audiovisual* performed better than *Audio* on this measure. For *Baseline*, showing no game elements at all might have increased concentration as there was no potential source of distraction. For *Audiovisual*, the synergy effect between SFX and visuals (**Theme 4**) could be the cause: While some participants felt irritated by auditory gamification (**Theme 2**) and distracted by visual gamification (**Theme 1**), the combination of both may have been able to compensate each others weaknesses to a certain extent. Visuals might give the sounds a clear meaning on the one hand (**R3**), while sounds themselves are perceived as less distracting from the task on the other hand, due to addressing a different modality than the task (**Theme 2**). This would be in line with findings from Hicks et al. [29], stating that "elements need to be chosen in a way that they do not act as a confound. For example, if a task is predominantly visual, graphical effects can be problematic, and alternatives such as audio feedback need to be considered". To conclude, instead of amplifying the negative effects, having both auditory and visual feedback complement each other's weaknesses in an already visually demanding task like image tagging, significantly improving perceived concentration of the users compared to auditory gamification alone.

Also in the range of flow effects, with **R5**, we found an increased lack of ambiguity of feedback for conditions with visual feedback (*Visual*, *Audiovisual*) compared to those without it (*Baseline*, *Audio*). For no gamification, this effect can again be explained with the lack of game elements, thus the lack of specific feedback mechanisms in the first place, while *Audio* was perceived as irritating, as the meaning of the chosen SFX was not entirely clear (**Theme 2**). This once more highlights the importance of adding feedback in general, but also adding semantics to the sound and ensuring a high semantic cohesion to enhance their inherent meaning.

Lastly, the effects of **R6** and **R7** indicate that audiovisual gamification resulted in a better perception of the challenge-skill balance, as well as a greater sense of control in comparison to *Audio* during the tagging task. Both results could again be attributed to the uncertainty that came with the *Audio* condition (**Theme 2**) as mentioned in the interviews. Thus it might be harder for participants to estimate their own skill, or feel of being in control, if they cannot interpret the

provided feedback in the form of SFX correctly. Taking **R3–R7** together and considering the absence of effects on SAM, IMI and parts of the FSS subscales, we conclude that **H3: The gamification modality (audio, visual, audiovisual) has an influence on the user experience** is partially supported.

6.5 Implications

What we can takeaway from our results is that the usage of SFX in gamification can benefit user experience. In all of our results on flow, *Audiovisual* outperformed at least one other condition. While descriptively *Audiovisual* generated the second-most number of tags after *Visual*, this could not be quantified with a statistical significance towards any of the other conditions. Hence, in contexts for which performance is the utmost goal, visual gamification alone might be best suited. In contexts in which user experience plays an important role as well, the addition of SFX can improve user experience, without sacrificing factors like concentration. Yet these recommendations are thus far limited to the context of this study.

Audio alone on the other hand could not show benefits with the SFX used in the context of image tagging. It was on par with the other conditions in terms of task performance. The same applies to the user experience. Based on the feedback we got in the interviews, auditory-only gamification might still show more potential in contexts with a task that is less visually demanding, and with a sound chosen specifically for the respective context and use.

7 Limitations

During the course of our studies we faced some limitations, which we would like to emphasize and discuss. With the exception of the interviews in the qualitative study, we conducted these studies (validation survey and quantitative study) online. While this was in line with [2] and allowed us to recruit a reasonable number of participants, we could not control for differences in loudness or quality of speakers, and headsets etc. used for media consumption. We did try to counteract this issue by adding the volume calibration task and a validation item to both online studies, requiring participants to activate their output devices at a reasonable level. Yet there was no control whether they changed their settings afterwards.

Furthermore, the context is limited to the image-tagging context, and the sounds are limited to those that were results of our validation survey. This selection of context and sound effects inevitably affects the generalizability of our findings. Thus, future work should further investigate the interplay between audio, visual and audiovisual gamification in different contexts. As became apparent in the interviews, the sound choice for points was perceived as too sharp or too repetitive by some participants when presented so many times, despite being favored in the validation survey. Both, specific sound design and context, could lead to different results if they were chosen differently.

Lastly, our sample in the qualitative study was predominantly German, or European respectively, meaning that we cannot account for cultural differences in visual and auditory perception. While this limits generalizability of the interviews to a certain extend, the themes derived in section 5 fit our quantitative results from section 4 well, where we evaluated a different and more diverse sample.

8 Conclusion

In this work, we systematically compared visual, auditory and audiovisual gamification regarding performance, user experience and perception in a gamified image-tagging context. With our findings we replicate existing gamification research: Visual gamification significantly increased task performance compared to using no gamification. Yet we could not find effects on user performance

and user experience related factors such as intrinsic motivation, valence or arousal when taking auditory elements into account. Perceived flow on the other hand was affected by the study condition, with the combination of visual and auditory gamification often exceeding others. Audiovisual gamification was also the most favored one among our participants.

Thus, we recommend the use of audiovisual gamification whenever user experience matters, especially since it can improve users' concentration in visually demanding tasks. However, in contexts where performance is important, one might consider using visual gamification only, as it showed the highest tag count in our study.

Regarding the use of sound effects in general, it is important to be aware of context and frequency of use, as repetitiveness, loudness and pitch may become an issue. Hence, the ultimate usage scenario should be considered and taken into account when choosing SFX. Furthermore, providing a semantic meaning with each sound, for example a gradually increasing pitch for increasing points, can decrease confusion about the SFX' meaning. This specifically applies to scenarios where gamification is provided solely via auditory channels.

For future work, we are interested in conducting a replication study in which the SFX are chosen for and validated within the specific task and context that they will be used in. Furthermore, we see two potentially interesting directions building on this work. First, investigating the role of context or the underlying task could yield interesting results. Some participants mentioned that visual gamification distracted them from the already visually demanding task of image tagging, which is why they preferred an auditory component to it. Hence, using auditory gamification in a context like sports (e.g. running) could be interesting, where visual game elements may be less applicable and the dependence on auditory input may be higher. Second, an investigation towards SFX with an inherent semantic meaning could be interesting to see whether they potentially reduce the confusion reported in the audio-only condition.

References

[1] Maximilian Altmeyer, Donald Degraen, Tobias Sander, Felix Kosmalla, and Antonio Krüger. 2021. Does Physicality Enhance the Meaningfulness of Gamification?Transforming Gamification Elements to their Physical Counterparts. In *33rd Australian Conference on Human-Computer Interaction*. ACM. https://doi.org/10.1145/3520495.3520500

[2] Maximilian Altmeyer, Vladislav Hnatovskiy, Katja Rogers, Pascal Lessel, and Lennart E. Nacke. 2022. Here Comes No Boom! The Lack of Sound Feedback Effects on Performance and User Experience in a Gamified Image Classification Task. In *CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3491102.3517581

[3] Maximilian Altmeyer, Berina Zenuni, Hanne Spelt, Thierry Jegen, Pascal Lessel, and Antonio Krüger. 2022. Do Hexad User Types Matter? Effects of (Non-) Personalized Gamification on Task Performance and User Experience in an Image Tagging Task. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–27.

[4] Gabriel Barata, Sandra Gama, Joaquim Jorge, and Daniel Gonçalves. 2017. Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior* 71 (jun 2017), 550–585. https://doi.org/10.1016/j.chb.2016.08.049

[5] William L Benoit and Pamela J Benoit. 2008. *Persuasive Messages: The Process of Influence.* Blackwell Publishing.

[6] Axel Berndt and Knut Hartmann. 2008. The functions of music in interactive media. In *Interactive Storytelling: First Joint International Conference on Interactive Digital Storytelling, ICIDS 2008 Erfurt, Germany, November 26-29, 2008 Proceedings 1*. Springer, 126–131.

[7] Meera Blattner, Denise Sumikawa, and Robert Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction* 4, 1 (mar 1989), 11–44. https://doi.org/10.1207/s15327051hci0401_1

[8] Terry L Bonebright and Michael A Nees. 2007. Memory for Auditory Icons and Earcons with Localization Cues. Georgia Institute of Technology. https://www.researchgate.net/profile/Michael-Nees/publication/228817420_Memory_for_auditory_icons_and_earcons_with_localization_cues/links/0fcfd5100a69c95b41000000/Memory-for-auditory-icons-and-earcons-with-localization-cues.pdf

[9] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (mar 1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[10] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[11] Stephen A Brewster, Peter C Wright, and Alastair DN Edwards. 1994. A detailed investigation into the effectiveness of earcons. In *Santa Fe Institute Studies In The Sciences Of Complexity*, Vol. 18. Addison-Wesley Publishing CO, 471–471.
[12] Paula Bräuer and Athanasios Mazarakis. 2022. "Alexa, can we design gamification without a screen?" - Implementing cooperative and competitive audio-gamification for intelligent virtual assistants. *Computers in Human Behavior* 135 (oct 2022), 107362. https://doi.org/10.1016/j.chb.2022.107362
[13] David Byrne. 2021. A Worked Example of Braun and Clarke's Approach to Reflexive Thematic Analysis. *Quality & Quantity* 56, 3 (June 2021), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y
[14] Gianna Cassidy and Raymond Macdonald. 2009. The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae* 13, 2 (sep 2009), 357–386. https://doi.org/10.1177/102986490901300207
[15] G.G. Cassidy and R.A.R. Macdonald. 2010. The effects of music on time perception and performance of a driving game. *Scandinavian Journal of Psychology* 51, 6 (jun 2010), 455–464. https://doi.org/10.1111/j.1467-9450.2010.00830.x
[16] Ruei-Che Chang, Chia-Sheng Hung, Dhruv Jain, and Anhong Guo. 2023. SoundBlender: Manipulating Sounds for Accessible Mixed-Reality Awareness. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (<conf-loc>, <city>San Francisco</city>, <state>CA</state>, <country>USA</country>, </conf-loc>) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 83, 4 pages. https://doi.org/10.1145/3586182.3615787
[17] Mihaly Csikszentmihalyi. 1975. *Beyond Boredom and Anxiety*. Jossey-Bass.
[18] Donald Degraen, Marc Schubhan, Maximilian Altmeyer, and Antonio Krüger. 2021. Hakoniwa: Enhancing Physical Gamification using Miniature Garden Elements. In *Academic Mindtrek 2021*. Association for Computing Machinery, New York, NY, USA, 117–127. https://doi.org/10.1145/3464327.3464362
[19] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11* (Tampere, Finland) *(MindTrek '11)*. Association for Computing Machinery, New York, NY, USA, 9–15. https://doi.org/10.1145/2181037.2181040
[20] Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. Gamification in education: A systematic mapping study. *Journal of educational technology & society* 18, 3 (2015), 75–88.
[21] Tilman Dingler, Jeffrey Lindsay, and Bruce N Walker. 2008. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. International Community for Auditory Display.
[22] Konstantinos Drossos, Nikolaos Zormpas, George Giannakopoulos, and Andreas Floros. 2015. Accessible Games for Blind Children, Empowered by Binaural Sound. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '15)*. Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/2769493.2769546
[23] Meng Du, Jia-Kai Chou, Chen Ma, Senthil Chandrasegaran, and Kwan-Liu Ma. 2018. Exploring the Role of Sound in Augmenting Visualization to Enhance User Engagement. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE. https://doi.org/10.1109/pacificvis.2018.00036
[24] Inger Ekman. 2013. On the desire to not kill your players: Rethinking sound in pervasive and mixed reality games.. In *Foundations of Digital Games (FDG)*. 142–149.
[25] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods* 5, 1 (mar 2006), 80–92. https://doi.org/10.1177/160940690600500107
[26] William W. Gaver. 1987. Auditory Icons. *ACM SIGCHI Bulletin* 19, 1 (jul 1987), 74. https://doi.org/10.1145/28189.1044809
[27] William W. Gaver. 1993. Synthesizing auditory icons. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*. ACM Press. https://doi.org/10.1145/169059.169184
[28] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*. IEEE. https://doi.org/10.1109/hicss.2014.377
[29] Kieran Hicks, Kathrin Gerling, Graham Richardson, Tom Pike, Oliver Burman, and Patrick Dickinson. 2019. Understanding the Effects of Gamification and Juiciness on Players. In *2019 IEEE Conference on Games (CoG)*. IEEE. https://doi.org/10.1109/cig.2019.8848105
[30] Md Naimul Hoque, Md Ehtesham-Ul-Haque, Niklas Elmqvist, and Syed Masum Billah. 2023. Accessible Data Representation with Natural Sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 826, 19 pages. https://doi.org/10.1145/3544548.3581087

[31] Mark J. Huiskes and Michael S. Lew. 2008. The MIR flickr retrieval evaluation. In *Proceeding of the 1st ACM international conference on Multimedia information retrieval - MIR '08*. ACM Press. https://doi.org/10.1145/1460096.1460104

[32] Gabriela Husain, William Forde Thompson, and E Glenn Schellenberg. 2002. Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music perception* 20, 2 (2002), 151–171.

[33] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. 2016. Personality-targeted Gamification: A Survey Study on Personality Traits and Motivational Affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2001–2013. https://doi.org/10.1145/2858036.2858515

[34] Daniel Johnson, Sebastian Deterding, Kerri-Ann Kuhn, Aleksandra Staneva, Stoyan Stoyanov, and Leanne Hides. 2016. Gamification for Health and Wellbeing: A Systematic Review of the Literature. *Internet Interventions* 6 (2016), 89–106. https://doi.org/10.1016/j.invent.2016.10.002

[35] Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Citeseer.

[36] Ana Carolina Tomé Klock, Isabela Gasparini, Marcelo Soares Pimenta, and Juho Hamari. 2020. Tailored gamification: A review of literature. *International Journal of Human-Computer Studies* 144 (dec 2020), 102495. https://doi.org/10.1016/j.ijhcs.2020.102495

[37] Jonna Koivisto and Juho Hamari. 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management* 45 (apr 2019), 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

[38] Oliver Korn and Adrian Rees. 2019. Affective effects of gamification. Using biosignals to measure the effects on working and learning users. *ACM International Conference Proceeding Series* (2019), 1–10. https://doi.org/10.1145/3316782.3316783

[39] Gregory Kramer, Bruce Walker, Terri L. Bonebright, Perry R. Cook, John H. Flowers, Nadine E. Miner, John G. Neuhoff, Robin Bargar, Stephen Barrass, Jonathan Berger, Grigori E. Evreinov, W. Tecumseh Fitch, Matti T. Gröhn, Steve Handel, Hans G. Kaper, Haim Levkowitz, Suresh K. Lodha, Barbara G. Shinn-Cunningham, Mary Simoni, and Sever Tipei. 1999. Sonification Report: Status of the Field and Research Agenda Prepared for the National Science Foundation by members of the International Community for Auditory Display. https://api.semanticscholar.org/CorpusID:2927531

[40] Richard N. Landers, Kristina N. Bauer, and Rachel C. Callan. 2017. Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior* 71 (jun 2017), 508–515. https://doi.org/10.1016/j.chb.2015.08.008

[41] Anne Landhäußer and Johannes Keller. 2012. Flow and its affective, cognitive, and performance-related consequences. *Advances in flow research* (2012), 65–85.

[42] Birger Langkjær. 2009. Making fictions sound real-On film sound, perceptual realism and genre. *MedieKultur: Journal of media and communication research* 26, 48 (2009), 13–p.

[43] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. Don't Whip Me With Your Games: Investigating Bottom-Up Gamification. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2026–2037. https://doi.org/10.1145/2858036.2858463

[44] Pascal Lessel, Maximilian Altmeyer, Lea Verena Schmeer, and Antonio Krüger. 2019. "Enable or Disable Gamification?": Analyzing the Impact of Choice in a Gamified Image Tagging Task. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3290605.3300380

[45] Laura Levy, Rob Solomon, Maribeth Gandy, and Richard Catrambone. 2015. The Rhythm's Going to Get You: Music's Effects on Gameplay and Experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (London, United Kingdom) *(CHI PLAY '15)*. Association for Computing Machinery, New York, NY, USA, 607–612. https://doi.org/10.1145/2793107.2810329

[46] Simon Y. W. Li, Chun-Wan Yeung, Thomas Davidson, Younji Ryu, Monika Srbinovska, Isaac Salisbury, Robert G. Loeb, and Penelope M. Sanderson. 2019. Spearcons for Patient Monitoring: Program of Laboratory-Based Feasibility Studies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (nov 2019), 663–667. https://doi.org/10.1177/1071181319631258

[47] Wei Li, Tovi Grossman, and George Fitzmaurice. 2012. GamiCAD: A Gamified Tutorial System For First Time AutoCAD Users. *ACM Symposium on User Interface Software and Technology* (2012), 103. https://doi.org/10.1145/2380116.2380131

[48] Mats Liljedahl. 2010. Sound for Fantasy and Freedom.

[49] Edwin A. Locke. 1968. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance* 3, 2 (may 1968), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4

[50] Edwin A. Locke and Gary P. Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist* 57, 9 (2002), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

[51] Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60, 1 (mar 1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413

[52] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Toronto, Ontario, Canada) *(Gamification '13)*.

ACM, New York, NY, USA, 66–73. https://doi.org/10.1145/2583008.2583017
[53] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. 2017. Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior* 71 (2017), 525–534. https://doi.org/10.1016/j.chb.2015.08.048
[54] Andrew T. Miranda and Evan M. Palmer. 2014. Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods* 46, 1 (2014), 159–172. https://doi.org/10.3758/s13428-013-0357-7
[55] Lennart E. Nacke and Sebastian Deterding. 2017. The Maturing of Gamification Research. *Computers in Human Behavior* 71 (2017), 450–454. https://doi.org/10.1016/j.chb.2016.11.062
[56] Lennart E. Nacke and Mark Grimshaw. 2011. Player-Game Interaction Through Affective Sound. In *Game Sound Technology and Player Interaction*. IGI Global, Hershey, PA, US, 264–285. https://doi.org/10.4018/978-1-61692-828-5.ch013
[57] Joseph W. Newbold, Nadia Bianchi-Berthouze, and Nicolas E. Gold. 2017. Musical Expectancy in Squat Sonification for People Who Struggle with Physical Activity. In *Proceedings of the 23rd International Conference on Auditory Display - ICAD 2017 (ICAD 2017)*. The International Community for Auditory Display. https://doi.org/10.21785/icad2017.008
[58] Brennan R. Payne, Joshua J. Jackson, Soo Rim Noh, and Elizabeth A. L. Stine-Morrow. 2011. Activity Flow State Scale. https://doi.org/10.1037/t06855-000
[59] Brennan R. Payne, Joshua J. Jackson, Soo Rim Noh, and Elizabeth A. L. Stine-Morrow. 2011. In the zone: Flow state and cognition in older adults. *Psychology and Aging* 26, 3 (sep 2011), 738–743. https://doi.org/10.1037/a0022359
[60] Heleen Plaisier, Thomas R. Meagher, and Daniel Barker. 2021. DNA Sonification for Public Engagement in Bioinformatics. *BMC Research Notes* 14, 1 (July 2021). https://doi.org/10.1186/s13104-021-05685-7
[61] Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development and Psychopathology* 17, 03 (sep 2005). https://doi.org/10.1017/s0954579405050340
[62] David Robinson and Victoria Belotti. 2013. A Preliminary Taxonomy of Gamification Elements for Varying Anticipated Commitment. *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (2013).
[63] Katja Rogers, Matthias Jörg, and Michael Weber. 2019. Effects of Background Music on Risk-Taking and General Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347158
[64] Katja Rogers and Michael Weber. 2019. Audio Habits and Motivations in Video Game Players. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound* (Nottingham, United Kingdom) *(AM'19)*. Association for Computing Machinery, New York, NY, USA, 45–52. https://doi.org/10.1145/3356590.3356599
[65] Mitchell Rogers, Wendy Yao, Andrew Luxton-Reilly, Juho Leinonen, Danielle Lottridge, and Paul Denny. 2021. Exploring Personalization of Gamification in an Introductory Programming Course. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. ACM. https://doi.org/10.1145/3408877.3432402
[66] Richard M. Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology* 43, 3 (sep 1982), 450–461. https://doi.org/10.1037/0022-3514.43.3.450
[67] Richard M. Ryan and Edward L Deci. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press, New York.
[68] Valorie N Salimpoor, David H Zald, Robert J Zatorre, Alain Dagher, and Anthony Randal McIntosh. 2015. Predictions and the brain: How musical sounds become rewarding. *Trends in cognitive sciences* 19, 2 (2015), 86–91.
[69] Marc Schubhan, Maximilian Altmeyer, Dominic Buchheit, and Pascal Lessel. 2020. Investigating User-Created Gamification in an Image Tagging Task. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3313831.3376360
[70] Leigh Schwartz. 2006. Fantasy, realism, and the other in recent video games. *Space and culture* 9, 3 (2006), 313–325.
[71] Katie Seaborn and Deborah I. Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74 (feb 2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006
[72] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376843
[73] Scott Sinnett, Charles Spence, and Salvador Soto-Faraco. 2007. Visual dominance and attention: The Colavita effect revisited. *Perception and Psychophysics* 69, 5 (2007), 673–686. https://doi.org/10.3758/BF03193770
[74] Siu-Lan Tan, John Baxa, and Matthew P Spackman. 2010. Effects of built-in audio versus unrelated background music on performance in an adventure role-playing game. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 2, 3 (2010), 1–23.
[75] Siu-Lan Tan, John Baxa, and Matthew P. Spackman. 2010. Effects of Built-in Audio versus Unrelated Background Music on Performance In an Adventure Role-Playing Game. *International Journal of Gaming and Computer-Mediated*

Simulations 2, 3 (jul 2010), 1–23. https://doi.org/10.4018/jgcms.2010070101

[76] Armando M. Toda, Pedro H. D. Valle, and Seiji Isotani. 2018. The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education. *Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions* (2018), 143–156. https://doi.org/10.1007/978-3-319-97934-2

[77] Gustavo F Tondello, Rina R Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-human Interaction in Play*. ACM, 229–243. https://doi.org/10.1145/2967934.2968082

[78] John W Tukey et al. 1977. *Exploratory data analysis*. Vol. 2. Reading, MA.

[79] L. M. Van Der Lubbe and M. C.A. Klein. 2020. Integrating gamification into a system to improve diet compliance for elderly users. *EAI International Conference on Smart Objects and Technologies for Social Good* (2020), 150–155. https://doi.org/10.1145/3411170.3411250

[80] Katharina Vogt, David Pirrò, Ingo Kobenz, Robert Höldrich, and Gerhard Eckel. 2010. *PhysioSonic - Evaluated Movement Sonification as Auditory Feedback in Physiotherapy*. Springer Berlin Heidelberg, 103–120. https://doi.org/10.1007/978-3-642-12439-6_6

[81] Bruce N Walker, Amanda Nance, and Jeffrey Lindsay. 2006. Spearcons: Speech-based earcons improve navigation performance in auditory menus. Georgia Institute of Technology.

[82] Alf Inge Wang and Andreas Lieberoth. 2016. The effect of points and audio on concentration, engagement, enjoyment, learning, motivation, and classroom dynamics using Kahoot. In *European conference on games based learning*, Vol. 20. Academic Conferences International Limited.

[83] Richard TA Wood, Mark D Griffiths, Darren Chappell, and Mark NO Davies. 2004. The structural characteristics of video games: A psycho-structural analysis. *CyberPsychology & behavior* 7, 1 (2004), 1–10.

[84] Nannan Xi and Juho Hamari. 2019. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management* 46 (2019), 210–221.

[85] Masashi Yamada. 2001. The effect of music on the performance and impression in a video racing game. *Journal of Music Perception and Cognition* 7 (2001), 65–76.

[86] A. Zanella, C. M. Harrison, S. Lenzi, J. Cooke, P. Damsma, and S. W. Fleming. 2022. Sonification and Sound Design for Astronomy Research, Education and Public Engagement. *Nature Astronomy* 6, 11 (Aug. 2022), 1241–1248. https://doi.org/10.1038/s41550-022-01721-z

A Sound Survey

A.1 Sound Source, Creator & Licence

Sound	Creator	Source	Licence
A	MattLeschuck	pixabay.com	Pixabay License
B	Wagna	pixabay.com	Pixabay License
C	bradwesson	pixabay.com	Pixabay License
D	plasterbrain	pixabay.com	Pixabay License
E	syseQ	pixabay.com	Pixabay License
F	Mixkit	mixkit.co	Mixkit Sound Effects Free License
G	Mixkit	mixkit.co	Mixkit Sound Effects Free License
H	Mixkit	mixkit.co	Mixkit Sound Effects Free License
I	Mixkit	mixkit.co	Mixkit Sound Effects Free License
J	Mixkit	mixkit.co	Mixkit Sound Effects Free License
K	Mixkit	mixkit.co	Mixkit Sound Effects Free License
L	Mixkit	mixkit.co	Mixkit Sound Effects Free License
M	Mixkit	mixkit.co	Mixkit Sound Effects Free License
M	Mixkit	mixkit.co	Mixkit Sound Effects Free License
O	Mixkit	mixkit.co	Mixkit Sound Effects Free License
P	BeezleFM	pixabay.com	Pixabay License
Q	FunWithSound	pixabay.com	Pixabay License

Table 6. Sounds (A–Q), their creator, source and licence.

B Interview Guide

For improved readability, the following transcript from the semi-structured interview guide has been translated from German to English for this appendix.

Interview 1 [Baseline]

- Which feedback did you notice when adding a tag?
- How did you like the feedback?
- What did the feedback feel like to you?
- What would you like to change about the feedback that you received when adding a tag?
- How did you like the image tagging? What was good or bad?

Interview 2-4 [Visual | Audio | Audiovisual]

- Which feedback did you notice when adding a tag?
- How did you like the feedback?
- What did the feedback feel like to you?
- What would you like to change about the feedback that you received when adding a tag?
- How did you like the image tagging? What was good or bad?
- As how fitting would you rate the [sounds | visual feedback] for the task?
- How much attention did you pay to the [sounds / visual feedback]?

Interview 5 [General]

- You received variations of feedback. Which one did you like best? Why?
- Which one did you like least? Why?
- To which variation did you pay the most attention? Why?
- How much did you notice the sounds?
- Would you rather go without the sounds or the visual feedback? Why?

Received February 2024; revised June 2024; accepted July 2024