Bräuer, Paula; Mazarakis, Athanasios

Article — Published Version

"Alexa, can we Design Gamification Without a Screen?" - Implementing Audio-Gamification for Intelligent Virtual Assistants

Computers in Human Behavior

Computers in Human Behavior 135 (2022) 107362

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

"Alexa, can we design gamification without a screen?" - Implementing cooperative and competitive audio-gamification for intelligent virtual assistants

Paula Bräuer *, Athanasios Mazarakis

Kiel University, Kiel, Germany

ARTICLE INFO

Keywords:
Gamification
Motivation
Performance
Intelligent virtual assistant
Amazon Alexa
Voice user interfaces

ABSTRACT

Intelligent virtual assistants (IVAs) like Amazon Alexa or Google Assistant have become increasingly popular in recent years, and research into the topic is growing accordingly. A major challenge in designing IVA applications is making them appealing. Gamification as a concept might help to boost motivation when using IVAs. Visual representation of progress and feedback is an essential component of gamification. When using IVAs, however, visual information is generally not available. To this end, this article reports the results of a lab experiment with 81 subjects describing how gamification, utilized entirely by audio, can assist subjects to work faster and improve motivation. Game design elements such as points and levels are integrated within an Alexa Skill via audio output to motivate subjects to complete household tasks. The results show a substantial effect on the subjects. Both their attitude and the processing time of the given tasks were positively influenced by the audio-gamification. The outcomes indicate that audio-gamification has a huge potential in the field of voice assistants. Differences in experimental conditions were also considered, but no statistical significance was found between the cooperative and competitive groups. Finally, we discuss how these insights affect IVA design principles and future research questions.

1. Introduction

Since the introduction of Siri for the Apple iPhone in 2011 (Guzman, 2019) intelligent virtual assistants (IVA) like Amazon Alexa, Google Assistant, or Microsoft Cortana have been introduced and are getting more and more popular (Clark et al., 2019; Dunn, 2016). In the literature, the term IVA is used interchangeably with terms like Intelligent Personal Assistant, Conversational Agent, Conversational User Interface, Virtual Personal Assistant, and Voice-Enabled Assistant, to mention a few (Cowan et al., 2017). These IVAs are operated primarily via voice and offer functions such as answering questions, playing music, controlling smart home devices, or creating shopping lists (Hoy, 2018). IVAs carry out about one billion tasks per month (Dellaert et al., 2020). Interaction with IVAs takes place via various interfaces, such as speaker devices, laptops, or smartphones (Guzman, 2019). In addition, wearables such as smartwatches and glasses[1] include an IVA, which can thus be used anytime and anywhere. To be able to react to user input, the speech recognition of the IVA is continuously active and waiting for specific wake words, e.g., "Ok Google" or "Alexa." If such a wake word is recognized, the voice input of the user is recorded and sent to a server for processing. Subsequently, the result of the request is sent back to the IVA, which transmits the (hopefully) desired information to the user via audio output (Hoy, 2018).

Often, speakers used to interact with an IVAs do not have a screen and can only reproduce information via audio output (Kinsella, 2019). This attribute has many advantages in everyday situations, such as driving a car, where the driver's view should remain focused on the road (Hofmann, Tobisch, Ehrlich, Berton, & Mahr, 2014), or following a recipe in the kitchen when handling hot ingredients while cooking (Vtyurina & Fourney, 2018). IVAs are also increasingly being used to control augmented and virtual reality applications. For example, Cortana is one of the essential input tools for the augmented reality glasses HoloLens from Microsoft (Derby, Rickel, Harris, Lovell, & Chaparro, 2020). The IVAs offer a simple way of interacting with a system without the use of classic input tools such as the keyboard or partially unintuitive input via gesture control. This improves the immersive experience of

* Corresponding author. ZBW/Kiel University, Düsternbrooker Weg 120, 24105, Kiel, Germany.
E-mail addresses: p.braeuer@zbw.eu (P. Bräuer), a.mazarakis@zbw.eu (A. Mazarakis).

[1] https://www.amazon.com/Echo-Frames/dp/B07W72XKPJ.

https://doi.org/10.1016/j.chb.2022.107362
Received 17 July 2021; Received in revised form 10 June 2022; Accepted 15 June 2022
Available online 18 June 2022

these applications (Shang & Wu, 2019).

IVAs also offer enormous potential in the healthcare sector. Interacting contactless with a computer can significantly reduce the risk of transmitting pathogens. For example, in sterile rooms, settings on devices can no longer be made at the touch of a button but simply by voice input (Mewes, Hensen, Wacker, & Hansen, 2017). It is also assumed that IVAs have enormous potential for e-marketing. Through AI-supported communication, for example, personalized consulting and shopping experiences can be created (Dellaert et al., 2020). However, the implementation of IVAs has not yet reached the point where users tend to personify them (Lopatovska & Williams, 2018). A study by Trajkova and Martin-Hammond (2020) showed that IVAs are currently rather poorly accepted by older adults for several reasons, including a lack of meaningful use cases or use in shared spaces, like, e.g., via cell phone while shopping.

Despite the increasing prevalence of IVAs and continued optimistic forecasts for the market (Shepherd & Liu, 2019), IVAs are often used by consumers only for a very limited set of tasks. Most of the time, IVAs are only used to play music or to ask about the weather (Lopatovska et al., 2019; Newman, 2018). In order to inspire users to use IVAs more widely, concepts are needed on how to make their use more interesting and motivating (Ji et al., 2017). As visual interaction via a screen is often missing, it is rather challenging to design motivational aids for IVAs because one can only rely on sound. In addition, a study by Montalvo et al. (2021) shows that interaction with an IVA is not considered easier when it has a screen, and usability is not improved compared to a device without a screen. Also, the effectiveness of performing tasks on IVAs with and without a screen does not seem to differ, so Montalvo et al. assume that an IVA without a screen is sufficient to meet the needs of most users. To motivate users to use an IVA and to make the applications more interesting, new approaches must be sought (Brandtzaeg & Følstad, 2018).

A common concept to engage and motivate users is gamification (Warmelink, Koivisto, Mayer, Vesa, & Hamari, 2018). Deterding, Dixon, Khaled, and Nacke (2011) define gamification as "the use of game design elements in non-game contexts." To investigate the effect of game design elements on motivation and performance, among other aspects, studies have been conducted in various non-game contexts, such as work (Sailer, Hense, Mandl, & Klevers, 2017, pp. 795–818; Stanculescu, Bozzon, Sips, & Houben, 2016), health (Allam, Kostova, Nakamoto, & Schulz, 2015; Cechetti et al., 2019) and education (Dindar, Ren, & Järvenoja, 2021; Geelan et al., 2015). Most gamification studies also show a positive effect in many different contexts (Johnson et al., 2016; Majuri, Koivisto, & Hamari, 2018; Warmelink et al., 2018).

In Human-Computer Interaction (HCI), information has long been conveyed visually via graphical user interfaces. Acoustic output was only used to convey simple information such as alerts or in the area of multimedia applications (Sciarretta & Alimenti, 2021). Gamified applications also tend to present elements of game design visually. For example, the progress in a level is often displayed in the form of a progress bar (Zichermann & Cunningham, 2011), badges are explicitly distinguished by a visual representation (Antin & Churchill, 2011), and even just feedback (e.g., if an answer is right or wrong) is usually conveyed via visual elements such as checkmarks and crosses or colors, often using the colors red and green (Baneres, Rodriguez, & Serra, 2019; Czeszumski, Ehinger, Wahn, & König, 2019; Mazarakis & Bräuer, 2018, 2020, 2022). Implementing gamification for an IVA without visual support presents a vital challenge to overcome.

Not long ago, video games provided excitement and overcame the limitations of the graphics processing units, and visual representations using elaborated sound effects. Sound is used to highlight and support progress, feedback, and other game design elements (Collins, 2008). One prominent example is the game "Super Mario." Each time the player collects a coin, i.e., scores a point, a characteristic sound is heard.

There are also plenty of purely audio-based computer games. These games offer, in particular for visually impaired people, the chance to play video games, which is difficult to impossible with classic video games, as they put a strong focus on visual elements (Garcia & de Almeida Neris, 2013). However, since both the control (such as fast addressing correct items in a menu) and overview elements (e.g., a mini-map or the level display), as well as the actual activity and story, must be represented by sound, finding a suitable relationship between functionality and aesthetics poses a fundamental challenge in the development of audio games (Friberg & Gärdenfors, 2004). Similar challenges can also be found in the design of applications for IVAs, where it is currently often assumed that those aesthetics play a minor role in design (Harris, 2004).

However, this is a major challenge. So far, there is no scientific research on whether and how purely acoustically implemented game design elements can influence motivation to use IVAs. This article addresses this fundamental research gap and presents the results of an experiment on the use of game design elements presented only via audio output, which we define as *audio-gamification* (Mazarakis, 2021). Our pioneering study aims to provide insight into how to design and implement gamification, which relies only on audio feedback and without any visual representations.

The following section summarizes related work of virtual assistants and gamification as well as aspects regarding motivation, performance, cooperation, and competition. It is followed by the methods section, which presents our experiment. Chapter four shows the results as descriptive and inferential statistics. After discussing and interpreting our results, we conclude and present outlines for future work in the final chapter.

2. Theoretical background

Some of the basic principles on which this article is based include research on audio feedback, gamification, and IVAs. In the following, these areas are examined in more detail, particularly gamification and IVAs concerning the motivational context, performance aspects, and state of the art according to cooperation and competition. Finally, the hypotheses are derived based on this theoretical background.

2.1. Gamification and intelligent virtual assistants

Games for IVAs are a research area that has been almost ignored so far and where there is a need for further research (de Barcelos Silva et al., 2020). Though gamification is closer to game design than games (Landers, Auer, Collmus, & Armstrong, 2018; Mazarakis, 2021), there are many similarities, and contextual research is being performed on this paring (Leaning, 2015). By conducting research on gamification for IVAs, we can also contribute to addressing this research gap. Recently, the effect of game design elements in combination with an IVA got some attention, e.g., when learning English as a foreign language, which was investigated in an experiment by Tejedor-Garcia, Escudero-Mancebo, Cardenoso-Payo, and Gonzalez-Ferreras (2020). The authors describe the design and testing of a gamified application that can be used to practice the pronunciation of English vocabulary. The application was developed for smartphones and works with Google Assistant for voice recognition. To promote motivation, challenges were created in which users could compete. To investigate the effect of the game design elements on motivation and performance in pronunciation practice, a comparison was made with results from an earlier study (Tejedor-Garcia, Escudero-Mancebo, Gonzalez-Ferreras, Cámara-Arenas, & Cardenoso-Payo, 2016). The comparison showed positive trends in both learning success and motivation. For the implementation of the game design elements (challenge, leaderboard, points, avatar, and badges), visual representation was used. The results suggest that gamification is helpful in this context, but the question remains open whether gamification in combination with IVAs also works without visual components.

A very simple and, at the same time, very important element of gamification is feedback (McGonigal, 2011; Zichermann &

Cunningham, 2011). Feedback informs the user about their performance or the status of their actions, enabling behavior change (Kapp, 2012; Ramaprasad, 1983). Feedback can provide many functions like information, timing, error analyses, or in general for motivation (Mory, 2004). For example, participation can be increased by adding simple right/wrong feedback to a specific task (Mazarakis, 2015). So feedback can have different definitions and it is an overly complex concept (Ramaprasad, 1983). Therefore, to investigate the effect of gamification in the context of IVAs, audio feedback can be considered particularly important.

Thiebes, Lins, and Basten (2014) identified a comprehensive collection of 31 different game design elements through a systematic review of the game design elements proposed in the gamification literature. In their article, the authors list “audible feedback” as one of the game design elements they identified, which they separate from classic feedback. Audio feedback offers the possibility to create emotions and transmit information. This can be done via sound effects, other accompanying non-verbal effects, or music. Then again, it is also possible to use oral sound (or the spoken word), where information is given meaning through emphasis or rate of speech (Clark, 2003).

Speech is often used in audio games to illustrate and help in complex situations that are easier to explain verbally (Röber & Masuch, 2005). For example, in story-based adventure games, a narrator can develop the story and summarize scenes or events at certain points during the game. In audio games, speech is better suited for transmitting large amounts of information, whereas beacons and auditory textures perform better for simple and short messages (Röber & Masuch, 2005). In order for the user to know what the result of the input command is or whether the input was accepted, immediate feedback is required in purely acoustic applications (Garcia & de Almeida Neris, 2013). For example, sound effects, such as footsteps while walking, show the player that a motion input was successful.

Research by Schmidmaier, Hußmann, and Runge (2020) deals with the effects of audio feedback on the formation of trust. The authors were able to experimentally demonstrate a significant effect of melody and mode on the formation of trust. This example shows how even simple adjustments can influence the effect of audio applications, such as skills for speech assistants.

Nykänen, Lopez, and Toulson (2016) investigated the effect of audio feedback on driving behavior. By using different forms of audio feedback, drivers should be motivated to drive as environmentally friendly as possible. Furthermore, the use of audio feedback should avoid the distraction of drivers that can occur when using a screen, for example. In a lab experiment, the authors were able to determine that signal tones and beep noises were received very differently while music or speech output was much better accepted. The results show that audio feedback is perceived very differently. It is already obvious that successful gamification with different audio game design elements is, therefore, a real challenge but at the same time also essential.

Silva-Coira, Cortiñas, and Pedreira (2016) establish a concept of how IVAs can be used in gamified environments. In the scenario presented, the IVA serves to help the user in their tasks by explaining the task or describing the use of tools. The goal of using IVAs in gamified environments is to detect the user’s mood using sentiment analysis and thus provide them with adapted answers. This should improve the experience of the application. Unfortunately, the authors describe the concept very superficial, and an evaluation of the concept is missing.

Ji et al. (2017) conducted a study with more than 20.000 users of an Alexa skill to identify ways to maintain the user’s motivation to use the IVA. To test the effect of different design variants, the IVA was extended with different modules during the test period. Subsequently, the trend of the rating of the skill on the Amazon platform was set in relation to the adjustments made. The authors found that, in general, a variety of conversational activities is desirable, and users tend to prefer natural conversations over menu-based conversations. It was also shown that the length of interactions with the conversation-based modules corresponded positively with user ratings. The user evaluation was also positively influenced by the inclusion of narratives.

Stories or narratives are used in many games, and gamification research has already considered this element (Grobelny, Smierzchalska, & Czapkowski, 2018; Mazarakis & Bräuer, 2018, 2022; Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818). This game design element is particularly suitable for gamifying applications for IVAs, as it can be implemented well without a visual component. Narratives are often used in audio games for the same reason (Röber & Masuch, 2005). Based on a detailed literature review, Keusch and Zhang (2017) point out that it is important for the impact of a narrative integrated into an application to be adapted to the context. Therefore, despite the promising potential of narratives in the context of IVAs, attention should be paid to the design of the story for the respective application.

Landesberger, Ehrlich, and Minker (2020) also recognized the potential of IVAs in cars. They conducted a gamified study to investigate how subjects react to urgent acoustic messages during an activity, such as driving. Unfortunately, the study did not use a purely acoustic application for this purpose but developed a kind of visual game that was interrupted by acoustic subtasks. Information on progress and feedback on the completed tasks was reflected in the study by visual elements. Also, Fadhil and Villafiorita (2017) use a visual representation of game design elements in their study, in which they investigate the effect of gamification on learning with the help of conversational user interfaces. Since the chatbot in their study is a text-based language assistant, a visual representation is appropriate, but this does not apply to purely language-based agents. However, to obtain clear insights into the effect of audio-gamification in situations where no screen is available and therefore is not influenced by visual elements, it is necessary to conduct pioneering research.

2.1.1. Motivation aspects

The relevance of user motivation in human–artificial intelligence interactions is becoming increasingly apparent as the usefulness of today’s IVAs grows (Li & Yanagisawa, 2021). An initial work addresses what positive and negative experiences can be identified when using IVAs and how they affect the motivation to use them (Weber & Ludwig, 2020). The goal of gamification is usually to increase the motivation of the users for a given task or activity (Warmelink, Koivisto, Mayer, Vesa, & Hamari, 2020). Motivation in this context is leaned on Deci and Ryan’s self-determination theory (SDT) (Ryan & Deci, 2000). This motivational theory is the most popular in gamification research (Seaborn & Fels, 2015).

According to SDT, motivation can range between extrinsic and intrinsic motivation. Intrinsic motivation is defined as doing something because it is inherently interesting or fun. On the other hand, extrinsic motivation refers to doing something because it leads to a definable result (Ryan & Deci, 2000, 2017). A derived theory of SDT is the theory of basic psychological needs (Deci & Ryan, 2000). This theory argues that there are three basic psychological needs: experience of competence, sense of autonomy, and social relatedness. Experience of competence relates to a need for challenge and feelings of effectance (Deci & Ryan, 2000). Autonomy describes a sense of volition or willingness when doing a task (Deci & Ryan, 2000), and social relatedness is said to occur when a person feels connected to others (Guardia, J. G., Ryan, Couchman, & Deci, 2000). Fulfilling these three needs can give rise to intrinsic motivation. The theory assumes that the three basic psychological needs are influenced by an individual’s environment. This influence can be both positive and negative and can contribute to satisfying the needs or inhibiting their fulfillment (Vansteenkiste & Ryan, 2013).

Computer game enjoyment has also been positively related to fulfilling those needs (Pe-Than, Goh, & Lee, 2014; Przybylski, Rigby, & Ryan, 2010). With multiple experiments, Ryan, Rigby, and Przybylski (2006) demonstrated that when playing alone, a relationship between autonomy and competence satisfaction could be demonstrated, as well as a relationship between all three needs in multiplayer contexts.

Many studies have shown that gamification influences motivation (Hamari, Koivisto, & Sarsa, 2014; Seaborn & Fels, 2015). Sailer, Hense, Mayr, and Mandl (2017), for example, looked at the effect of different game design elements on the three basic needs. The results of their study show that perceptions of competence can be positively influenced by badges, leaderboards, and performance graphs. Furthermore, avatars and narratives have shown a positive effect on the perception of social inclusion. Although most studies investigating the impact of gamification on motivation have yielded positive results, the context in which the game design elements are used has been shown to influence the effect (Hamari et al., 2014).

Research on IVAs has also recognized SDT as a relevant research concept. For example, Li and Yanagisawa (2021) showed that reducing uncertainty can lead to more interactions with an IVA, thus shifting the motivation for the interactions from non-intrinsic to intrinsic. Here, the authors assume that intrinsic motivation is present in the context of IVAs when activities are performed to pass the time and for play activities. Extrinsically motivated interactions, on the other hand, are pragmatically motivated to obtain information or to have certain tasks done by the assistants to save time.

Another important theory in terms of motivation and gamification is the flow theory. Flow theory should be considered in the context of intrinsic motivation, as the theory consists of aspects like immediate feedback and pursuing an activity for its own sake (Csikszentmihalyi, 2009). Feedback is crucial because an individual can get the necessary information on whether the activity in question is performed as intended. Furthermore, immediate and constant feedback helps to maintain the flow experience (Kapp, 2012). Based on the current state of research on the motivational effects of gamification and the lack of studies examining this very effect through audio-gamification, the following general research question is posed as a basis for upcoming hypotheses:

RQ1. How does audio-gamification affect motivation?

2.1.2. Performance aspects

In addition to the effect of gamification on motivation, studies also examine how gamification affects performance (Groening & Binnewies, 2019; Landers, Bauer, & Callan, 2015; Mekler, Bruehlmann, Tuch, & Opwis, 2017). Depending on the setting, a change in performance can address various metrics, such as quality, quantity, speed, or accuracy. For example, this can be beneficial in the work environment (Warmelink et al., 2018, 2020) when workers perform gamified tasks faster or more accurately or in learning (Bai, Hew, & Huang, 2020; Sailer & Homner, 2020), when students spend more time on a task and better results are achieved.

Groening and Binnewies (2019) show that individual game design elements can influence performance using the example of achievements. In a study with a within-subjects design, in which images were annotated, it could be shown that performance could be increased by adding achievements. Likewise, Mekler et al. (2017) demonstrated positive results by adding points, levels, and leaderboards in a similar setting. In a field experiment, Grobelny et al. (2018) demonstrated a positive effect on sales performance with the help of a narrative-based gamification approach. For this purpose, the sales performance of two franchisees was recorded and compared over two months. Moreover, in logistics, Sailer, Hense, Mandl, and Klevers (2017) were also able to show in an experimental study that performance in order picking can be increased through gamification. Bräuer and Mazarakis (2019) could support these findings in a different logistics study and show that some game design elements might be unsuitable in a specific context to enhance performance.

Based on this body of research, we assume that performance can also be influenced by purely audio-based game design elements, and the following research question was posed:

RQ2. How does audio-gamification affect performance?

2.1.3. Aspects on cooperation and competition

Since numerous studies have already proven the effect of gamification (Hamari et al., 2014; Mekler et al., 2017; Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818), current research is also concerned with investigating which different design approaches are particularly effective in the implementation of gamification (Rapp, Hopfgartner, Hamari, Linehan, & Cena, 2019). An important difference when implementing a game design element is whether the element is designed cooperatively or competitively. In competitive games, such as car racing, the objective is to compete with one or more opponents and win against them. It is about being faster, smarter, or simply more skilled than the opponents (Kapp, 2012). Cooperative design is about working together to achieve a common goal or result. This concept is often found in app games for smart devices, such as Candy Crush or FarmVille. By cooperating with friends and exchanging or giving virtual goods, the progress of the individual players can be accelerated significantly. They are encouraged not to work alone towards a specific goal but to team up with other players (Kapp, 2012).

How competition affects performance in a gamified setting was, e.g., experimentally investigated by Landers, Collmus, and Williams (2019) in the context of brainstorming. In comparison to a control group, it was shown that both creativity and the quality of the brainstorming contributions could be improved by competition. The authors assume that the motivational effect of the competition must have come about either through unconscious influence or through the implicit creation of extrinsic rewards for increased effort (Landers et al., 2019).

Morschheuser, Hamari, and Maedche (2019) conducted a study to discover the differences between cooperative, competitive, and inter-team competitive gamification. In a field experiment in a crowdsourcing context, the authors demonstrated that inter-team competition had the most significant positive effect on the participation and enjoyment of participants. However, a different study that examined the effect of cooperative and competitive gamification on the flow perception of students did not find differences between the different design approaches (Marinho et al., 2019). Nevertheless, the study results are limited due to the data of only 18 individuals.

Dindar et al. (2021) investigated the effects of cooperative and competitive gamification with the help of a gamified app for learning vocabulary. The comparison between two groups, one cooperative and one competitive, did not reveal any differences in the subjects' learning success, performance, or motivation. Unfortunately, no control group without gamification was used in their study. Thus, it cannot be ruled out that the overall gamification design may have achieved the desired effect. The authors claim that "The current findings emphasize that the positive influence of gamified cooperation on creating meaningful connections amongst learners should not be ignored, even though it facilitates similar learning and motivational outcomes as gamified competition." (Dindar et al., 2021, p. 142). This is contrary to the claims of, e.g., Morschheuser et al. (2019) and Landers et al. (2019).

Although the research gap of cooperative and competitive gamification has been identified and initial studies on this topic exist, the results are not always conclusive, and their interpretation is not always clear. Furthermore, there is a lack of studies that compare the effect of cooperative and competitive gamification in a controlled setting. Therefore, to close this research gap and test whether the advantages of cooperative team competition can be transferred to the context of audio-gamification, RQ3 and RQ4 were formed.

RQ3. How do cooperation and competition influence motivation when using audio-gamification?

RQ4. How do cooperation and competition influence performance when using audio-gamification?

2.2. Hypotheses

So far, no studies have explicitly addressed audio-gamification and its effects on motivation and task performance. The research goal of this pioneering study is to investigate how audio-only implemented game design elements can be deployed successfully and how they can enhance motivation and performance. Six hypotheses were established to systematically investigate this research gap and to answer all four research questions presented in this chapter.

Hypotheses H1a and H1b address RQ2, the relationship between audio-gamification and performance. We argue that, based on the results of Montalvo et al. (2021), task performance when using an IVA is not negatively affected by the absence of a screen. At the same time, we assume that the positive effect on performance induced by primarily exclusively visual gamification can also be achieved by audio-gamification, similar to audio games.

H1a. Subjects in the experimental condition with *competitive* audio-gamification perform tasks significantly faster than the control group.

H1b. Subjects in the experimental condition with *cooperative* audio-gamification perform tasks significantly faster than the control group.

Hypotheses H2a and H2b address RQ1, namely the relationship between audio-gamification and motivation. According to our assumptions when developing the hypotheses about performance, we believe that the numerous beneficial results obtained via gamification on motivation can also be obtained through audio-gamification.

H2a. Subjects in the experimental condition with *competitive* audio-gamification feel significantly more motivated than the control group.

H2b. Subjects in the experimental condition with *cooperative* audio-gamification feel significantly more motivated than the control group.

The last two hypotheses address the difference between cooperatively and competitively implemented audio-gamification. Hypothesis H3 addresses RQ3, and RQ4 is handled by hypothesis H4.

H3. Subjects in the experimental condition with *cooperative* audio-gamification perform tasks significantly faster than the subjects in the experimental condition with *competitive* audio-gamification.

H4. Subjects in the experimental condition with *cooperative* audio-gamification feel more motivated than the subjects in the experimental condition with *competitive* audio-gamification.

3. Method and data

To investigate the effects that audio-gamification can have on motivation and performance, a lab experiment was conducted. The following section describes the experimental design, the operationalization of the dependent and independent variables, and descriptive data about the subjects of the study in detail.

3.1. Experimental design

We created for our experiment a skill for the IVA Amazon Alexa to test in a lab experiment. A skill is similar to an app for a smartphone (Zhang et al., 2019). The skill extends the range of actions that can be performed by the IVA, using voice control provided by Amazon via a corresponding device, e.g., Amazon Echo.

The study uses a 3 × 2 between-subject design. The subjects are randomly assigned to one of three groups. One group serves as a control group without any game design elements. In contrast, the other two experimental groups (competition group and cooperation group) use audio-implemented game design elements, both with five different game design elements: points, time pressure, level, narrative, and ranking. The groups are compared after the experiment to identify possible differences in the performance and motivation of the subjects to find support for our hypotheses, including data from a subsequent questionnaire administered immediately after the lab experiment.

Similar to the study by Silva-Coira et al. (2016), the IVA supports the user in performing tasks. The tasks and the general course of the experiment do not differ between the groups. Only the subjects in the experimental groups are given additional audio game design elements built into the skill. In the following, we will first describe the general design of the experiment as it was implemented for the control group. Then we will describe the different game design elements and how they were implemented in the two experimental conditions.

3.1.1. Control group

Each subject was greeted and briefed that the purpose of the experiment was to evaluate a new skill. The subjects are not aware that there is more than one test condition. First, a tutorial familiarized the subjects with the use of Alexa and the skill. They learned which commands they had to use, for example, to start new tasks or how to get information about the current task. Each subject was then taken to a separate room where there was an Alexa Echo Dot speaker, and the subjects did start with the experiment. During the whole experiment, a supervisor was present in the same room. The supervisor was available to help in the unlikely event that a subject has problems interacting with the IVA or to help them with any technical or other issues.

The IVA explains six different tasks to the subjects during the experiment, one by one. The tasks should be reminiscent of daily household chores, which are usually done rather reluctantly. The tasks include activities such as cleaning the room or sorting books into a shelf. To receive new tasks or to confirm when a task has been completed, the subject interacts with the Alexa Skill. By saying, "I'm done!" the subject confirms that he or she has successfully completed the last task. Fig. 1 shows an example dialogue of a subject in the control group.

Because in a real-world setting usually, it is not possible to check whether the tasks have actually been finished completely and correctly, the supervisor also did not intervene in our laboratory setting. We expected that the mere presence of a supervisor will not lead to cheating. If a subject deviates from the task, the supervisor of the experiment will note this silently to ensure a high internal validity of the experiment.

After the completion of all tasks, a questionnaire had to be filled out to obtain information about the motivation during the activity as well as feedback on the audio implementation. The subjects in the experimental groups are also asked whether and how they perceived the individual game design elements. Since other studies have already found differences in the effect of gamification on gender (Codish & Ravid, 2017; Koivisto & Hamari, 2014) and age (Jent & Janneck, 2018), this information is also collected. We provide additional information in section 3.2 about the questionnaire.

Next, we provide a short overview of the different game design elements and their implementation in the experiment. The basic structure of both experimental groups corresponds to the structure of the control group and differs only in these game design elements.

3.1.2. Points

Points are the basis of most gamified applications (Zichermann & Cunningham, 2011) and offer a possibility to show the progress of the users in numerical form (Seaborn & Fels, 2015). Subjects in the gamified experimental groups get for each successfully accomplished task 50 points. In addition, extra points can be earned depending on the speed performance. A maximum of 25 bonus points is awarded if the subject completes a task within the minimum duration, which was previously determined based on a pre-test. Therefore, a maximum of 75 points per task can be obtained, which makes a maximum of 450 points for the complete experiment (six tasks).

The subject will be informed about how to earn points and bonus points when the skill is started. This description and the number of points achieved are always communicated to the subject exclusively by audio through the Alexa Speaker. An example of a dialogue with a gamified version of the skill is given in Fig. 2.

3.1.3. Time pressure

By awarding the extra points and announcing them at the end of a

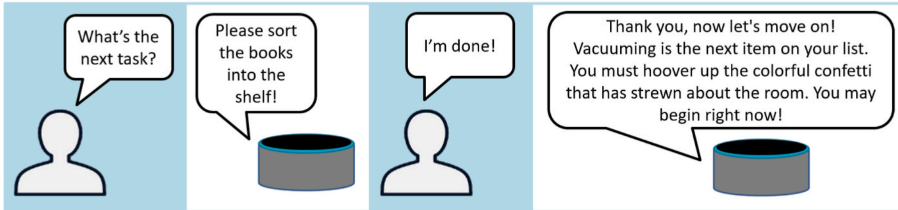

Fig. 1. Example of a dialogue of a subject in the control group.

Fig. 2. Example of a dialogue between a subject and a gamified version of the Alexa skills.

task, time pressure is created in this study. In many games, this game design element is used, often in a visual form, e.g., by a clearing hourglass or displaying a countdown. By communicating time pressure, the stress level of the players is increased, and the motivation to act is stimulated (Kapp, 2012).

Time pressure can be used to create a form of challenge. However, it is not always helpful, especially when tasks need to be completed precisely (Thiebes et al., 2014). For this reason, time is only measured here for the awarding of extra points.

3.1.4. Level

Levels can have a similar effect as given goals and motivate a subject to achieve them (Koivisto & Hamari, 2019). In games, they can display ranks that the player can reach or refer to stages or areas in a game world (Seaborn & Fels, 2015). Often, progress bars are used to show how far a player has progressed (Zichermann & Cunningham, 2011).

If a certain number of tasks are completed in our studies, the subject is informed via sound output that a new level has been reached. The IVA says: "Yippee, you've reached level two!". A total of three levels were provided. The levels are firmly linked to the completion of the tasks. The subject moves up a level for the first time by completing the first three tasks. The second level is completed as soon as the fifth task is completed. The third and last level ends with the last task. As a result, we used levels to inform the subjects about their progress.

3.1.5. Narrative

In addition to points, time pressure, and levels, the game design element *narrative* is integrated into the skill. If a user earns points, the story unfolds gradually. This form of gamification is reminiscent of one of the first forms of computer games, so-called text-adventures (Atkinson, Baier, Copplestone, Devlin, & Swan, 2019). In this type of game, the player goes through a story that is told exclusively through text output. Also, any interaction of the player takes place via text input. A similar principle was applied in this study to design the narrative element. By interacting with the IVA via speech, the subject progresses through the story. Since Ji et al. (2017) showed that users prefer natural conversations to menu-based navigation, it was expected that the integration of the narrative would have a positive impact on the perception of the skill.

To create the most immersive experience possible with the IVA, an attempt was made to choose the narrative to match the context as much as possible (Keusch & Zhang, 2017). The narrative that was developed for this skill is loosely based on the video game "Portal" ("Portal," 2019). The subjects interact with an artificial intelligence called "TILA." TILA is rather grumpy in the beginning but will become nicer the more tasks are fulfilled. While the user is working on the tasks, TILA tells anecdotes and facts that are vaguely related to the content of the current task. When cleaning up, for example, she curses about the pollution of the environment by humans.

Throughout the story, the subject notices that TILA gives the subject tasks to help her get a password to escape from the speaker. The intermediate dialogs not only help develop the narrative but also assist from a technical perspective. The Alexa Skills Kit limits the maximum break time a skill remains active when there is no user input to 10 s (Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit, n.d.). After this period, the speaker shuts down and must be restarted by the wake word. The intermediate dialogs thus help to extend the time limit of 10 s.

3.1.6. Ranking

A social component can be created by comparing the results with others using rankings or leaderboards (Kapp, 2012). Rankings are therefore also a popular element in gamification research for generating cooperation or competition (Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818). In this study, the game design element ranking was implemented in two different conditions, cooperative and competitive. An overview of how the ranking was implemented in both gamified experimental conditions is shown in Fig. 3.

In the cooperative group, the subjects are told at the beginning of the story that they are part of Team Bravo and that this team competes against the fictitious Team Alpha. When the intermediate ranking is announced, subjects in the cooperative group receive information on how their Team Bravo compares with the fictitious Team Alpha. The information on how their team compares with the Alpha Team is given four times: after completing the second, third, fifth, and sixth task. It was decided not to do this after each task since it would be too monotonous in combination with the regular announcement of the points, and TILA would have talked too much, maybe leading to distraction.

Subjects in the competitive group do not receive any information at the beginning that a comparison with other people is taking place. Just after the second task has been completed, the subjects are informed whether they performed better, equally well, or worse than the previous subjects. It was decided that the design of the ranking should be based on a general statement (see Fig. 3) and not on a list of best scores with exact rankings. The reason for this decision was that it is much easier to understand instead of a leaderboard where several users and scores have to be listed. The competitive ranking is announced at the same four stages as in the cooperative group.

Whether a subject performs well, moderately, or poorly in the

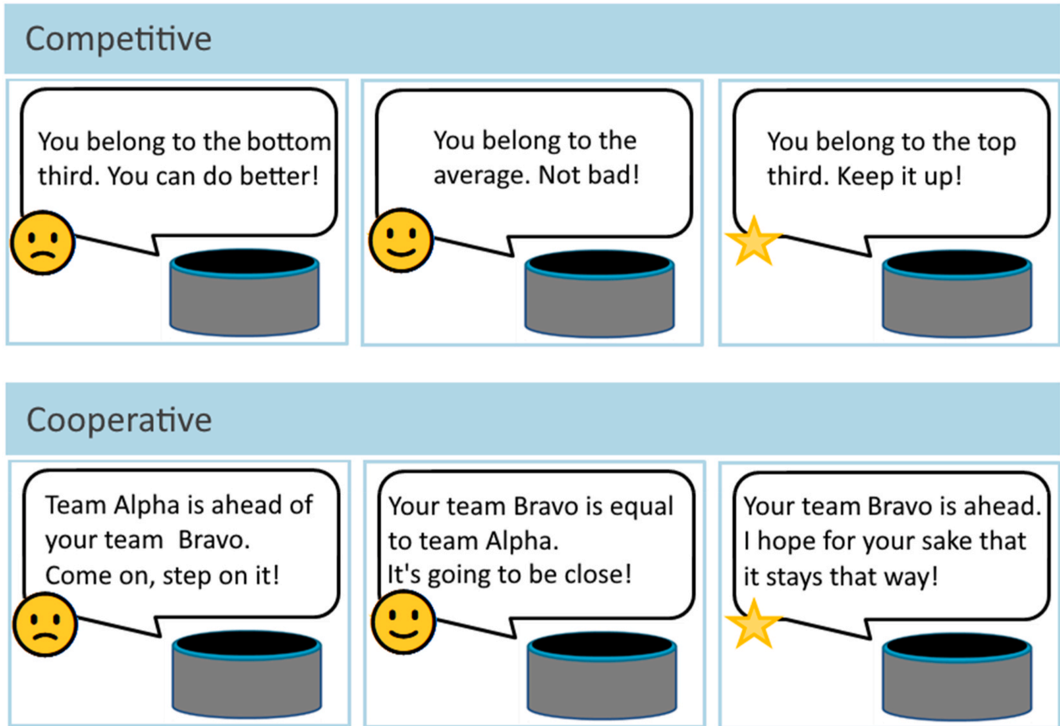

Fig. 3. Description of the cooperative and competitive ranking in the three gradations negative, neutral, and positive.

ranking is determined for each task based on the time it takes to complete the task. Therefore, the data for assessment on which the ranking is based is the same for all subjects and, like the points, is based on empirical data that has been collected by testing the tasks several times in a pre-test.

3.1.7. Audio feedback

In addition to the five game design elements mentioned, audio feedback is used to make the application more appealing. According to Thiebes et al. (2014), audio feedback includes both sound effects and music. We used short positive and negative mood audio files to provide positive or negative feedback to subjects. Furthermore, we used sound effects like a drum roll to create excitement. A combination of sound effects and music can also be found in audiobooks, for example, where they are used to create mood and atmosphere (Steinhaeusser, Schaper, & Lugrin, 2021). The audio files of the study supported the narrative and helped to give feedback in a simpler and more concise way than it would have been possible through detailed verbal responses (Röber & Masuch, 2005).

3.2. Measurements

In the study, the subjects' task completion time and motivation were analyzed. To evaluate the task completion time, the time required by the subjects to perform the individual tasks was measured. To be able to compare the times of the different test conditions, the exact completion time was collected. For the analysis, this completion time was adjusted by the time TILA was talking. For the final analysis, the total adjusted time on task for each individual was used.

An additional survey was immediately conducted after the laboratory experiment. The subjects had to answer a short questionnaire via a Google Forms sheet. In addition to demographic data (age and sex), the subjects were also asked about their previous usage behavior and experience with IVAs. The subjects could choose between six possible options ("never," "occasionally," "several times a month," "several times a week," "daily," and "several times a day"). Additionally, the subjects were asked to indicate how motivated they were during the experiment. For this purpose, the following three questions were asked, which were evaluated based on a 5-point Likert scale (1 = not at all; 5 = very much):

- Q1 - Were you motivated by the virtual assistant (TILA) to solve the tasks?
- Q2 - Were you motivated by the virtual assistant (TILA) to solve the individual tasks faster?
- Q3 - Did you have fun with the virtual assistant (TILA)?

To perform a manipulation check, we asked the subjects in the questionnaire whether and if so which game design elements they perceived during the study. The subjects could first choose whether they thought they had perceived game design elements and then tick which of the five game design elements (points, time pressure, level, narrative, and ranking) they had perceived.

3.3. Participants

Data were collected from 91 participants at a German university. Of these, ten had to be excluded from evaluation because of incomplete data or technical or other issues. For example, contradictory results were obtained due to Internet failures or because subjects skipped entire tasks. Of these ten excluded subjects, four were in the control group, three in the group with cooperative gamification, and three in the group with competitive gamification. In total, a final sample size of N = 81 is obtained for further analysis.

The average age of the sample was 23.56 years (SD 4.42), with the youngest subject 17 years old and the oldest 41 years. 53 subjects were male and 28 female. The high proportion of students (85%) is given that the acquisition of the subjects was carried out at the university. The other subjects were employees.

Concerning the adoption and frequency of use of IVAs, 57 of the subjects (70.4%) stated that they had never used IVAs. Compared to the average use of IVAs, 26.2% in the U.S. and 25.8% in Germany (Kinsella, 2021), the proportion of people who at least occasionally use an IVA is

29.6% in our sample, which is comparable. An overview of the exact distribution in our study is given in Table 1. In addition, comparing the usage with a pre-COVID-19 survey with more than 1660 adults in the U. S.A., similarities can be found for weekly and monthly usage behavior (NPR and Edison Research, 2020).

4. Results

In this section, descriptive statistics are first briefly presented. Subsequently, the hypotheses are evaluated in three stages: first, the effect of audio-gamification on performance is examined, followed by the effect on motivation and the comparison between cooperative and competitive implementation. Finally, we consider whether there is an influence of variables such as age, gender, or previous experience with IVAs on the variables investigated.

Table 2 gives an overview of the number of subjects in the three test conditions, as well as the mean and median value and standard deviation of the time in seconds required by the subjects to complete all tasks. The unequal number of subjects per test condition is due to the random distribution of the subjects. Overall, the mean value for the adjusted required time is 544.78 s with a standard deviation of 115.77 s, and the median is 547.34 s. The fastest subject is from the cooperative condition with 331.58 s, and the slowest is also a subject from the same group with 838.32 s.

4.1. Effect on performance

To evaluate the hypotheses, it is first checked whether the collected data are normally distributed. For the hypotheses, H1a and H1b (subjects in experimental conditions are faster than the control group) as well as H3 (the cooperative group is faster than the competitive group), the normal distribution of the adjusted required time is tested first. According to the Shapiro-Wilk test (Field, 2017), no statistically significant result is provided, $W(81) = 0.98$, $p = .308$, and therefore a normal distribution of the data can be assumed.

Then the variance homogeneity of the sample is checked. The Levene test (Field, 2017) does not provide a statistically significant result, $F(2, 78) = 0.07$, $p = .936$. Thus, it can be assumed that the variance of the mean values for all three groups is not different, and a t-test can be performed to evaluate the hypotheses H1a, H1b, and H3.

Already from the descriptive results in Table 2, it can be seen that the subjects from the competitive group completed the tasks faster than those in the control group. Furthermore, the statistical comparison of the mean values between the competitive group ($M = 490.37$, $SD = 99.31$) and the control group ($M = 635.57$, $SD = 94.00$) also provides a statistically significant result with a large effect, $t(46) = 5.17$, $p = .000$, $d = 1.50$. Thus, H1a can be supported, and it is assumed that integrating audio-gamification with competitive design can positively influence performance.

The cooperative group's descriptive results from Table 2 also show that the tasks were completed faster than in the control group. Furthermore, the mean value comparison between the cooperative group ($M = 527.11$, $SD = 107.73$) and the control group ($M = 635.57$, $SD = 94.00$) is also statistically significant with a large effect, $t(53) = 3.84$, $p = .000$, $d = 1.07$. Thus, H1b can also be supported, and it can be assumed that the inclusion of both competitive and cooperative audio-gamification will positively impact performance.

Table 1
Overview of the degree of previous experience with IVAs of the subjects.

Frequency of use	*N*	Percentage of active users
Never	57	–
Occasionally	14	58.33%
Several times a month	4	16.67%
Several times a week	2	8.33%
Daily	4	16.67%
Serval times a day	0	0%

Table 2
Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group.

Condition	*N*	*M*	*Mdn*	*SD*
Control	22	635.57	607.79	94.00
Competitive	26	490.37	504.80	99.31
Cooperative	33	527.11	526.48	107.73

4.2. Effect on motivation

The normal distribution of the questionnaire items *Q1, Q2, and Q3* is checked first to evaluate the hypotheses H2a and H2b (subjects in the experimental conditions are more motivated than the control group) and H4 (the cooperative group is more motivated than the competitive group). The Shapiro-Wilk test provides a significant result for all three items: Q1 - *General Motivation* $W(81) = 0.89$, $p = .000$; Q2 - *Motivation to work fast* $W(81) = 0.90$, $p = .000$; Q3 - *Feeling of fun* $W(81) = 0.87$, $p = .000$. The data of all three items are therefore not normally distributed, and a non-parametric method (Mann-Whitney-U-Test) is used for further evaluation.

The results of the mean value comparisons for the evaluation of H2a are summarized in Table 4, the corresponding median values can be found in Table 3. For all three items, the Mann-Whitney U Test provides a statistically significant result. Furthermore, a medium effect was determined for the items Q1 - *General Motivation* ($r = 0.33$) and Q2 - *Motivation to work fast* ($r = 0.44$), the item Q3 - *Feeling of fun* achieves only a weak effect ($r = 0.24$). Thus, hypothesis H2a can be supported, and it is assumed that the perceived motivation of the subjects can be increased by competitive audio-gamification..

Table 5 presents the results of the mean value comparisons for the evaluation of H2b. The Mann-Whitney-U-Test provides statistically significant results for comparing the cooperative group and control group for all three items. In addition, the effect size for the two items Q1 – *General Motivation* ($r = 0.35$) and Q3 – *Feeling of fun* ($r = 0.35$) reach a medium value, and for Q2 – *Motivation to work fast,* a high effect size value ($r = 0.51$). Thus, H2b is also supported, and it can be assumed that cooperative gamification has a positive effect on the perceived motivation.

4.3. Comparison of cooperative vs. competitive

In the descriptive results in Table 2, the competitive group completed the tasks on average faster than the cooperative experimental group. Thus, our hypothesis H3 (subjects in the cooperative group are faster than in the competitive group) cannot be supported as it was set up initially with an opposite expected result. Therefore, we accept the null hypothesis.

Finally, we check our hypothesis H4 (subjects in the cooperative group are more motivated than in the competitive group). The descriptive data for the three items in Table 3 show that the cooperative group achieves a higher mean value for all questionnaire items than the competitive group. The results of the statistical analysis are summarized in Table 6.

Comparing the mean values leads to no statistically significant results for all three items. Thus, H4 cannot be supported, and it must be assumed that it has no statistically significant influence on the motivation whether a cooperative or competitive design is applied.

4.4. Additional statistical analyses

In addition to the hypotheses, it was also examined whether an influence of the factors age, gender, or previous knowledge in dealing with

Table 3
Mean and standard deviation for all three motivational items (General motivation/Motivation to work faster/Feeling of fun) per experimental group.

Condition	General Motivation			Motivation to work faster			Feeling of fun		
	M	Mdn	SD	M	Mdn	SD	M	Mdn	SD
Control	2.64	3.00	1.22	2.23	2.00	1.02	3.00	3.00	1.20
Competitive	3.38	4.00	0.98	3.38	4.00	1.30	3.58	4.00	0.95
Cooperative	3.52	4.00	1.06	3.58	4.00	1.23	3.76	4.00	1.20

Table 4
Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the competitive group regarding the three items for measuring motivation.

Item	*U*	*z*	*p* (one-tailed)	*r*
Q1 General Motivation	180.50	−2.29	.011	.33
Q2 Motivation to work fast	142.00	−3.06	.001	.44
Q3 Feeling of fun	210.50	−1.66	.049	.24

Table 5
Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the cooperative group concerning the three items for measuring motivation.

Item	*U*	*z*	*p* (one-tailed)	*r*
Q1 General Motivation	217.50	−2.59	.005	.35
Q2 Motivation to work fast	149.50	−3.76	.000	.51
Q3 Feeling of fun	231.50	−2.36	.010	.35

Table 6
Results of the Mann-Whitney-U-Test for the comparison between the cooperative and the competitive group regarding the three items for measuring motivation.

Item	*U*	*Z*	*p* (one-tailed)
Q1 General Motivation	409.0	−0.32	.373
Q2 Motivation to work fast	395.5	−0.53	.298
Q3 Feeling of fun	363.5	−1.06	.146

IVAs on the three motivation items and the time required could be determined. The reason is to control for possible influences of extraneous variables. The correlations were calculated for the entire data set. Since the three items used to measure motivation are not normally distributed, as shown in section 4.2., a Spearman correlation is calculated. The analysis shows no significant correlations between age, gender, or the experience with IVAs and one of the other variables. Interestingly, the three questions concerning motivation correlate, suggesting good construct validity. The results are summarized in Table 7. Negative correlation values for gender are in favor of male subjects, and positive correlation values are in favor of females.

The variables age and gender appear interesting in the context of the study. Reminder, 53 subjects are male, and 28 are female. When conducting the study, males were on average 24.19 years old ($SD = 4.57$) and females 22.36 years old ($SD = 3.91$). The age distribution between the genders within the three experimental conditions is unremarkable. If we take the results of Table 2 and further analyze the results according to gender, we get the following results, as displayed in Table 8.

Although the results do not become statistically significant with the *t*-test, strong trends are apparent. In the control and the cooperative conditions, the male subjects took more time to accomplish the tasks than the female subjects. In contrast, female subjects were slower in the competitive condition. A more balanced gender distribution per condition might have led to more conclusive results in this case. Still, this is an exciting result, which shows that gender effects can play an essential role in performance, which also happened to be the case on other occasions (Bräuer & Mazarakis, 2019). Age- and gender-related differences in the use of voice-activated artificially intelligent (voice-AI) devices have not gone unnoticed and are already being addressed in studies (Zellou, Cohn, & Ferenc Segedin, 2021). The need for future research in this area should already be noted at this point.

For the motivational aspects (general motivation and motivation to work fast), it is already evident from the correlation table that the female subjects were more motivated than the males. The questionnaire items for general motivation were tended to be rated lower by males ($M = 3.08$, $SD = 1.14$, $Mdn = 3.00$) than for females ($M = 3.54$, $SD = 1.08$, $Mdn = 4.00$), $U = 564.50$, $z = -1.84$, $p = .065$. In addition, the questionnaire items for the motivation to work fast were also tended to be

Table 8
Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group and gender.

Condition	*N*	*M*	*Mdn*	*SD*	*p* (two-tailed)
Control Male	16	654.65	646.27	100.40	.122
Control Female	6	584.70	587.53	51.03	
Competitive Male	16	473.04	466.49	104.10	.269
Competitive Female	10	518.10	514.82	89.12	
Cooperative Male	21	552.77	549.62	113.86	.069
Cooperative Female	12	482.20	472.42	82.08	

Table 7
Correlation matrix of the additional statistical analyses. *P*-values are two-tailed.

		Age	Gender	Experience with IVAs	General Motivation	Motivation to work fast	Feeling of fun	Adjusted required time
Age	*r*	1.00	−.196	.034	−.130	−.017	−.010	−.211
	p		.079	.761	.248	.882	.927	.059
Gender	*r*		1.00	−.190	.206	.211	−.022	−.163
	p			.089	.065	.059	.844	.145
Experience with IVAs	*r*			1.00	−.109	−.011	.129	−.027
	p				.331	.923	.253	.810
General Motivation	*r*				1.00	.387**	.527**	−.100
	p					.000	.000	.373
Motivation to work fast	*r*					1.00	.405**	−.198
	p						.000	.077
Feeling of fun	*r*						1.00	−.108
	p							.337
Adjusted required time	*r*							1.00
	p							

rated lower by males ($M = 2.94$, $SD = 1.34$, $Mdn = 3.00$) than for females ($M = 3.54$, $SD = 1.20$, $Mdn = 4.00$), $U = 557.00$, $z = -1.89$, $p = .060$. Again, a more balanced gender distribution may have resulted in more definitive answers in this context. However, between the three experimental conditions, the results are unexceptional.

5. Discussion

This pioneering study was designed to investigate a new innovative field of application of gamification and IVAs for human behavior, which we call audio-gamification. The goal was to systematically determine whether game design elements also have a motivating and performance-enhancing effect in purely audio-based applications for IVAs. Following our expectations, a statistically significant effect on the motivation of the subjects as well as on the speed at which they complete given tasks could be determined. Thus the results are in accordance with previous gamification research in other fields (Johnson et al., 2016; Majuri et al., 2018; Seaborn & Fels, 2015; Warmelink et al., 2018).

As Murad, Munteanu, Clark, and Cowan (2018) point out, feedback is of central importance in the design of IVAs. Audio-gamification can help to provide the user with the necessary feedback in a familiar way. At the same time, care must be taken not to overwhelm the user with additional elements which could distract or bore them. Our study is compliant with design recommendations for non-complex interaction models (Murad et al., 2018), and our results show that audio-gamification can improve the experience with IVAs. Our results provide implications for improving user motivation of IVAs, which might be helpful in many application domains. Possible areas of application could be, for example, language learning, which is currently a popular field of research on the use of IVAs (Dizon, 2021). Since it is known that gamification can improve both motivation for learning and learning success (Bai et al., 2020; Majuri et al., 2018; Sailer & Homner, 2020), audio-gamification could possibly also produce such an effect in learning with IVAs. Various applications for audio-gamification are also possible in the realm of sustainability research. Drivers could be incentivized to adopt ecologically friendly driving behavior, for example, by incorporating game design features in cars, similar to the Landesberger et al. (2020)study. Alternatively, in the application in the smart home, users could be motivated to use the devices in the household in a more energy-efficient way (Mendez et al., 2020).

Research into the effects of audio-gamification opens new possibilities for the further development of IVAs. Another field of application in which gamification can be implemented without visual support is applications for visually impaired individuals (Mazarakis, 2021; Sciarretta & Alimenti, 2021). Through an audio-only implementation, obstacles such as the use of color to convey instruction, to indicate an action, or to prompt a response, which often occur during the implementation of gamification, can be bypassed (Smith & Abrams, 2019). In this respect, our results help to go beyond the work of Tejedor-Garcia et al. (2020) to show that for the implementation of gamification in the context of IVAs, a smartphone or other device with a display is not necessarily required. To enable the most inclusive design of IVAs, we create new possibilities to make applications more appealing to people with limited vision by using only sound and speech to represent the game design elements. Our study provides valuable insight that offers a new perspective on the potential of gamification in a wide range of applications.

The positive results of our study can also be attributed to the successful embedding of the tasks in the narrative around the AI TILA. With the help of the narrative, the user was guided through the menu of the application. As a result, receiving and confirming the tasks was more like a conversation with TILA. This also is in line with the results of Ji et al. (2017), according to which users prefer natural conversations over menu-based navigation. This leads to the assumption that by including a narrative element, an enormous potential can be unlocked regarding the improvement of the acceptance of IVAs.

5.1. Limitations

In addition to the effect of audio-gamification on motivation and performance, this study examined whether differences between cooperative and competitive game design elements could be demonstrated. Contrary to the results of, e.g., Morschheuser et al. (2019) and our expectations, no difference in motivation or performance could be detected between the groups. Other studies also struggled with the differences between cooperative and competitive game design elements, e.g., they were not always clear either (Dindar et al., 2021; Marinho et al., 2019). In the case of our study, there may be too marginal differences in the design of the two groups. The same five game design elements (points, time pressure, level, ranking, and narrative) were used in both groups. The only difference in terms of game design elements and besides the different amount of information provided between the cooperative and the competitive group was the implementation of the ranking. To achieve a clear impact, the differences between cooperative and competitive gamification should be highlighted more clearly. For example, the narrative could be adapted more or challenges designed cooperatively or competitively could be used.

A possible general limitation should not remain unmentioned: a cooperative setting may not always be beneficial. Despite the efforts of, e.g., Morschheuser (2019), others provide different findings. A highly competitive game design element like a leaderboard or ranking can have a massive impact on performance (Bräuer & Mazarakis, 2019). In addition, Dindar et al. (2021) show that subjects in their study completed the tasks in the competitive condition statistically significant faster than in the cooperative condition. Therefore, the context in which gamification is to be used must always be considered (Bräuer & Mazarakis, 2019; Finckenhagen, 2017; Koivisto & Hamari, 2014; Mazarakis, 2021; Richards, Thompson, & Graham, 2014).

In our study, the motivation of the subjects was determined based on a questionnaire with three self-developed questions. The procedure for measuring motivation could have been based on established questionnaire tools, such as the Intrinsic Motivation Inventory (Intrinsic Motivation Inventory (IMI) – Selfdeterminationtheory.Org, n.d.). The advantage of using such a tool would be, besides the proven validity of the questionnaire, the possibility to collect information about the satisfaction of different basic needs (autonomy, relatedness, and competence). This would allow more precise conclusions to be drawn about the effect of audio-gamification on motivation.

Also, the effect that the individual game design elements might have on each other was not considered in the study. As Mekler et al. (2017), and Mazarakis and Bräuer (2018, 2022) were able to show, different game design elements have quite different effects when applied individually. Since the focus of our study was to test whether an effect can be achieved at all, several game design elements were used in combination so that no conclusions could be drawn about the mode of action of the individual game design elements. Especially about the audio feedback, information on the individual effects could be extremely helpful for the design of skills. In addition, it may be necessary to take individual needs into account when designing audio game design elements and to offer the user the possibility to adjust or switch off the elements.

As with most studies in the field of gamification, there is a lack of evaluation of the long-term impact of gamification (Mazarakis, 2021; Rapp et al., 2019; Seaborn & Fels, 2015). The effect investigated was demonstrated in a laboratory setting in which the subjects interacted with the system once and for no longer than half an hour. Therefore, it remains questionable how audio-gamification would affect motivation if a user used a similar incentive system several times over a longer period of time. To answer this question, a follow-up study should be conducted. In such a study, however, the setting would have to be adapted since the skill developed for the laboratory experiment with the predefined tasks would not be reusable in a long-term field study. Alternatively, the audio-gamification concept might be transferred to a foreign language learning application, for example. Another imaginable scenario to implement the research design in a less restrictive laboratory setting would be to emphasize health exercises, specifically relaxation exercises, emphasizing breathing exercises.

A final limiting aspect in conducting the study is the relatively low proportion of daily users of IVAs (16.67%) in the sample. Although the ratio of subjects without experience with IVAs compared to subjects with experience in our study corresponds to the prevalence of experience with IVAs in Germany (Kinsella, 2021), it is still possible that the usage behavior could be influenced by this aspect. In addition, while no major differences in interaction with the skill between experienced and inexperienced users were found (see Table 7), future studies should include a larger group of users with more prior experience, e.g., daily use. This could lead to group-specific insight, e.g., different results for power users.

5.2. Future research

Our study provides first results in the field of audio-gamification and shows the potential of integrating game design elements without the use of a screen. At the same time, this study raises several further questions that should be addressed by future research to provide a valid basis for the design of IVAs based on our results. A first starting point is provided by the additionally considered differences between the genders. In both fields of gamification and research on IVAs, initial studies address the effects of age and gender (Codish & Ravid, 2017; Jent & Janneck, 2018; Zellou et al., 2021). Our data also suggest that there might be differences in the motivational effect of audio-gamification due to these influencing factors. Similarly, users' prior experience with IVAs could also be an essential factor influencing the impact of audio gamification. Future research could address this potentially limiting factor.

Many users of Amazon Alexa are annoyed after some time because the IVA always gives them the same long answers. Amazon has therefore now built in the option to receive only short answers such as "yes" or a tone instead of a spoken answer (Alexa Brief Mode Explained, 2019). Since the addition of game design elements automatically extends the responses of the IVA, it is possible that a similar fatigue effect can result. By conducting an online study over a longer period of time, conclusions can be drawn as to whether this effect also occurs with audio game design elements and, if so, what design options there are to counteract this phenomenon.

An important point in designing applications for IVAs is to make the menu navigation as efficient and straightforward as possible (Murad et al., 2018). Unfortunately, the integration of game design elements counteracts this principle and could therefore harm usability. However, since the results on motivation and performance improvement of our study were positive, it should be determined whether the integration of game design elements really has a negative impact on usability. To find out whether such an effect can be verified, this should be investigated using suitable evaluation procedures, such as appropriate questionnaires like the system usability scale (Ghosh, Foong, Zhang, & Zhao, 2018).

A third important point that should be considered for future research is the direct comparison between visual and audio implementation of game design elements. To be able to conclude how important a visual component is in the implementation of game design elements, more studies with different devices and settings are needed. For example, some speakers with IVAs like the Alexa Echo Show or the Google Nest Hub already have integrated a screen. A comparison of the usage behavior between two skills that differ only in the presentation of the game design elements (visual vs. audio) could provide insights into how much users are influenced by a visual component (Montalvo et al., 2021).

One factor that could decisively influence the effect of audio-gamification is the presence of other people in the room. In contrast to working on screens, where users usually use one device alone, several users can simultaneously interact with an IVA (Trajkova & Martin-Hammond, 2020). The assistants can be trained on different voices and thus identify various users (Amazon.Com Help: About Alexa Voice Profiles, 2020). It has already been shown that working together in a group with IVAs has a positive impact on cooperation within the group and improves the output of group work (Winkler, Söllner, Neuweiler, Conti Rossini, & Leimeister, 2019). On the other hand, the presence of other people can also act as a deterrent and discourage the use of the IVA (Trajkova & Martin-Hammond, 2020). Depending on the implementation of gamification and the personality traits of the user, the presence of other people could lead to amplification or attenuation of the motivating effect.

6. Conclusion

Our pioneering study with 81 subjects empirically investigated the effect of audio-only game design elements on motivation and performance when using an IVA. We answered our research questions regarding how audio-gamification affects motivation and performance. The positive results are consistent with previous studies on the general effect of gamification (Johnson et al., 2016; Majuri et al., 2018; Seaborn & Fels, 2015; Warmelink et al., 2018) and show that a visual component is not necessarily required for the successful implementation of game design elements. We also considered differences in the design of the audio game design elements. However, the comparison between a cooperative and a competitive group did not show any statistically significant difference for our corresponding research questions about motivation and performance. The results thus contradict previous work (Morschheuser et al., 2019) and offer a variety of possibilities to research the contradictions. We believe that our work will contribute to a fundamental paradigm shift for IVAs in developing human-centered skills and creating new, more appealing motivational design approaches, thus possibly promoting the acceptance of IVAs in the long term, especially with gamification.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Credit author statement

Paula Bräuer: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision.; **Athanasios Mazarakis:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

References

Alexa Brief Mode Explained: How to Turn It on and How It Works. (2019, May 6). The Ambient. https://www.the-ambient.com/how-to/how-to-use-alexa-brief-mode-500.

Allam, A., Kostova, Z., Nakamoto, K., & Schulz, P. J. (2015). The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: Randomized controlled trial. *Journal of Medical Internet Research, 17*(1), e14. https://doi.org/10.2196/jmir.3510

Amazon.Com help: About Alexa voice Profiles. (2020). https://www.amazon.com/gp/help/customer/display.html?nodeId=202199440.

Antin, J., & Churchill, E. F. (2011). *Badges in Social Media: A Social Psychological Perspective. CHI 2011 Gamification Workshop Proceedings*, 1–4.

Atkinson, T., Baier, H., Copplestone, T., Devlin, S., & Swan, J. (2019). The text-based adventure AI competition. *IEEE Transactions on Games, 11*(3), 260–266. https://doi.org/10.1109/TG.2019.2896017

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and Synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322. https://doi.org/10.1016/j.edurev.2020.100322

Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Transactions on Learning Technologies, 12*(2), 249–263. https://doi.org/10.1109/TLT.2019.2912167 de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., et al. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications, 147*, 113193. https://doi.org/10.1016/j.eswa.2020.113193

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions, 25*(5), 38–43. https://doi.org/10.1145/3236669

Bräuer, P., & Mazarakis, A. (2019). Badges or a leaderboard? How to gamify an augmented reality warehouse setting. In J. Koivisto, & Hamari (Eds.), *Vol. 2359. Proceedings of the 3rd international GamiFIN conference—GamiFIN 2019* (pp. 229–240) (CEUR-WS.org).

Cechetti, N. P., Bellei, E. A., Biduski, D., Rodriguez, J. P. M., Roman, M. K., & De Marchi, A. C. B. (2019). Developing and implementing a gamification method to improve user engagement: A case study with an M-health application for hypertension monitoring. *Telematics and Informatics, 41*, 126–138. https://doi.org/10.1016/j.tele.2019.04.007

Clark, L. (2003). Speaking pictures: The role of sound and orality in audio presentations of children's picture books. *New Review of Children's Literature and Librarianship, 9* (1), 1–19. https://doi.org/10.1080/1361454032000232077

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., et al. (2019). The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers, 31* (4), 349–371. https://doi.org/10.1093/iwc/iwz016

Codish, D., & Ravid, G. (2017). *Gender Moderation in Gamification: Does One Size Fit All? Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2017.244

Collins, K. (2008). *From Pac-Man to pop music: Interactive audio in games and new media*. Ashgate.

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., et al. (2017). "What can I help you with?": Infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–12). https://doi.org/10.1145/3098279.3098539

Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. Harper Collins.

Czeszumski, A., Ehinger, B. V., Wahn, B., & König, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology, 10*(361). https://doi.org/10.3389/fpsyg.2019.00361

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dellaert, B. G. C., Shu, S. B., Arentze, T. A., Baker, T., Diehl, K., Donkers, B., et al. (2020). Consumer decisions with artificially intelligent voice assistants. *Marketing Letters*. https://doi.org/10.1007/s11002-020-09537-5

Derby, J. L., Rickel, E. A., Harris, K. J., Lovell, J. A., & Chaparro, B. S. (2020). "We didn't catch that!" using voice text input on a mixed reality headset in noisy environments. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 64*(1), 2102–2106. https://doi.org/10.1177/1071181320641509

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From Game Design Elements to Gamefulness: Defining "Gamification." Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Dindar, M., Ren, L., & Järvenoja, H. (2021). An experimental study on the effects of gamified cooperation and competition on English vocabulary learning. *British Journal of Educational Technology, 52*(1), 142–159. https://doi.org/10.1111/bjet.12977

Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal*. https://doi.org/10.1177/00336882211020548, 003368822110205.

Dunn, J. (2016, August 26). Virtual assistants like Siri and Alexa look poised to explode. *Business Insider Deutschland*. https://www.businessinsider.de/virtual-assistants-siri-alexa-growth-chart-2016-8.

Fadhil, A., & Villafiorita, A. (2017). An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, 408–412. https://doi.org/10.1145/3099023.3099112

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Finckenhagen, K. R. (2017). *Context in Gamification—Contextual Factors and Successful Gamification*. https://www.ntnu.edu/documents/139799/1279149990/09+Article+Final_karolirf_fors%C3%B8k_2017-12-08-05-35-05_TPD4505.Karoline.Finckenhagen.pdf/4f269657-0e54-4999-9de8-6ba8e109d386.

Friberg, J., & Gärdenfors, D. (2004). Audio games: New perspectives on game audio. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '04*, 148–154. https://doi.org/10.1145/1067343.1067361

Garcia, F. E., & de Almeida Neris, V. P. (2013). Design guidelines for audio games. In M. Kurosu (Ed.), *8005. Proceedings of the 15th international conference on human-computer interaction* (pp. 229–238). Springer. https://doi.org/10.1007/978-3-642-39262-7_26.

Geelan, B., De Salas, K., Lewis, I., King, C., Edwards, D., & O'Mara, A. (2015). Improving learning experiences through gamification: A case study. *Australian Educational Computing, 30*(1), 1–21.

Ghosh, D., Foong, P. S., Zhang, S., & Zhao, S. (2018). Assessing the utility of the system usability scale for evaluating voice-based user interfaces. In *Proceedings of the Sixth International Symposium of Chinese CHI on - ChineseCHI '18* (pp. 11–15). https://doi.org/10.1145/3202667.3204844

Grobelny, J., Smierzchalska, J., & Czapkowski, K. (2018). Narrative gamification as a method of increasing sales performance: A field experimental study. *International Journal of Academic Research in Business and Social Sciences, 8*(3), 430–447. https://doi.org/10.6007/IJARBSS/v8-i3/3940

Groening, C., & Binnewies, C. (2019). "Achievement unlocked!"—the impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior, 97*, 151–166. https://doi.org/10.1016/j.chb.2019.02.026

Guardia, La, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology, 79*(3), 367–384. https://doi.org/10.1037/0022-3514.79.3.367

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior, 90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034). https://doi.org/10.1109/HICSS.2014.377

Harris, R. A. (2004). *Voice interaction design: Crafting the new conversational speech systems*. Elsevier.

Hofmann, H., Tobisch, V., Ehrlich, U., Berton, A., & Mahr, A. (2014). Comparison of speech-based in-car hmi concepts in a driving simulation study. In *Proceedings of the 19th International Conference on Intelligent User Interfaces* (pp. 215–224). https://doi.org/10.1145/2557500.2557509

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly, 37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Intrinsic Motivation Inventory (IMI) – selfdeterminationtheory.org. (n.d.). Retrieved November 6, 2020, from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/.

Jent, S., & Janneck, M. (2018). Using gamification to enhance user motivation: The influence of gender and age. In L. E. Freund, & W. Cellary (Eds.), *Vol. 601. Advances in the human side of service engineering* (pp. 3–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-60486-2_1.

Ji, J., Wang, Q., Battad, Z., Gou, J., Zhou, J., Divekar, R., et al. (2017). A two-layer dialogue framework for authoring social bots. *Alexa Prize Proceedings*.

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions, 6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.

Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review, 35*(2), 147–166. https://doi.org/10.1177/0894439315608451

Kinsella, B. (2019, March 7). U.S. Smart display user base grew 558% in 2018 and more than doubled in second half of the year, amazon holds two-thirds market share. *Voicebot.Ai*. https://voicebot.ai/2019/03/07/u-s-smart-display-user-base-grew-558-in-2018-and-more-than-doubled-in-second-half-of-the-year-amazon-holds-two-thirds-market-share/.

Kinsella, B. (2021, June 17). Germany smart speaker adoption closely mirrors U.S. Pattern—new report with 30+ charts. *Voicebot.Ai*. https://voicebot.ai/2021/06/17/germany-smart-speaker-adoption-closely-mirrors-u-s-pattern-new-report-with-30-charts/.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179–188. https://doi.org/10.1016/j.chb.2014.03.007

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming, 49*(3), 315–337. https://doi.org/10.1177/1046878118774385

Landers, R. N., Bauer, K. N., & Callan, R. C. (2015). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior, 71*, 508–515. https://doi.org/10.1016/j.chb.2015.08.008

Landers, R. N., Collmus, A. B., & Williams, H. (2019). The greatest battle is within ourselves: An experiment on the effects of competition alone on task performance. *International Journal of Human-Computer Studies, 127*, 51–61. https://doi.org/10.1016/j.ijhcs.2018.09.011

Landesberger, J., Ehrlich, U., & Minker, W. (2020). "What is it?" How to Collect Urgent Utterances using a Gamification Approach. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 19–22). https://doi.org/10.1145/3409251.3411713

Leaning, M. (2015). A study of the use of games and gamification to enhance student engagement, experience and achievement on a theory-based course of an undergraduate media degree. *Journal of Media Practice, 16*(2), 155–170. https://doi.org/10.1080/14682753.2015.1041807

Li, C., & Yanagisawa, H. (2021). Intrinsic motivation in virtual assistant interaction for fostering spontaneous interactions. *PLoS One, 16*(4). https://doi.org/10.1371/journal.pone.0250326. e0250326.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., et al. (2019). Talk to me: Exploring user interactions with the amazon Alexa. *Journal of Librarianship and Information Science, 51*(4), 984–997. https://doi.org/10.1177/0961000618759414

Lopatovska, I., & Williams, H. (2018). Personification of the Amazon Alexa: BFF or a Mindless Companion. In *, 18. Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR* (pp. 265–268). https://doi.org/10.1145/3176349.3176868

Majuri, J., Koivisto, J., & Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *, 2018. Proceedings of the 2nd International GamiFIN Conference, GamiFIN* (pp. 11–19).
Marinho, A., Oliveira, W., Bittencourt, I. I., & Demerval, D. (2019). Does Gamification Improve Flow Experience in Classroom? An Analysis of Gamer Types in Collaborative and Competitive Settings. *Revista Brasileira de Informática Na Educação, 27*(2), 40. https://doi.org/10.5753/rbie.2019.27.02.40
Mazarakis, A. (2015). Using gamification for technology enhanced learning: The case of feedback mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology, 4*(17), 6–9.
Mazarakis, A. (2021). Gamification reloaded: Current and future trends in gamification science. *I-com, 20*(3), 279–294. https://doi.org/10.1515/icom-2021-0025
Mazarakis, A., & Bräuer, P. (2018). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *Proceedings of the Technology, Mind, and Society, 22*(1). https://doi.org/10.1145/3183654.3183667
Mazarakis, A., & Bräuer, P. (2020). Gamification of an open access quiz with badges and progress bars: An experimental study with scientists. In J. Koivisto, M. Bujic, & J. Hamari (Eds.)*, Vol. 2637. Proceedings of the 4th international GamiFIN conference—GamiFIN 2020* (pp. 62–71) (CEUR-WS.org).
Mazarakis, A., & Bräuer, P. (2022). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *International Journal of Human-Computer Interaction*, 1–16. https://doi.org/10.1080/10447318.2022.2041909
McGonigal, J. (2011). *Reality is broken.* Penguin Press.
Mekler, E. D., Bruehlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior, 71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048
Mendez, J. I., Ponce, P., Mata, O., Meier, A., Peffer, T., Molina, A., et al. (2020). Empower saving energy into smart homes using a gamification structure by social products. In *2020 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1–7). https://doi.org/10.1109/ICCE46568.2020.9043174
Mewes, A., Hensen, B., Wacker, F., & Hansen, C. (2017). Touchless interaction with software in interventional radiology and surgery: A systematic literature review. *International Journal of Computer Assisted Radiology and Surgery, 12*(2), 291–305. https://doi.org/10.1007/s11548-016-1480-6
Montalvo, F., Schreck, J., Krause, K., Sasser, J., McConnell, D. S., & Smither, J. A. (2021). Do touchscreens make intelligent personal assistants more useable? An initial usability study. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 65*(1), 899–903. https://doi.org/10.1177/1071181321651283
Morschheuser, B., Hamari, J., & Maedche, A. (2019). Cooperation or competition – when do people contribute more? A field experiment on gamification of crowdsourcing. *International Journal of Human-Computer Studies, 127*, 7–24. https://doi.org/10.1016/j.ijhcs.2018.10.001
Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.)*, Vol. 2. Handbook of research on educational communications and technology* (pp. 745–783). Lawrence Erlbaum Associates.
Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design guidelines for hands-free speech interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18* (pp. 269–276). https://doi.org/10.1145/3236112.3236149
Newman, N. (2018). The Future of Voice and the Implications for News. https://reutersinstitute.politics.ox.ac.uk/our-research/future-voice-and-implications-news.
NPR and Edison Research. (2020). *Smart Audio Report.* https://www.nationalpublicmedia.com/uploads/2020/04/The-Smart-Audio-Report_Spring-2020.pdf.
Nykänen, A., Lopez, M., & Toulson, R. (2016). Safe and sound drive: Design of interactive sounds supporting energy efficient behaviour. *Interactive Audio Systems Symposium, York*, 23 September 2016.
Pe-Than, E. P. P., Goh, D. H.-L., & Lee, C. S. (2014). Making work fun: Investigating antecedents of perceived enjoyment in human computation games for information sharing. *Computers in Human Behavior, 39*, 88–99. https://doi.org/10.1016/j.chb.2014.06.023
Portal. (2019). *In Wikipedia.* https://de.wikipedia.org/w/index.php?title=Portal_(Computerspiel)&oldid=185754590.
Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154–166. https://doi.org/10.1037/a0019440
Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4–13. https://doi.org/10.1002/bs.3830280103
Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. *International Journal of Human-Computer Studies, 127*, 1–6. https://doi.org/10.1016/j.ijhcs.2018.11.007
Richards, C., Thompson, C. W., & Graham, N. (2014). Beyond designing for motivation: The importance of context in gamification. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '14)*, 217–226. https://doi.org/10.1145/2658537.2658683
Röber, N., & Masuch, M. (2005). *Leaving the Screen New Perspectives in Audio-Only Gaming.* https://smartech.gatech.edu/handle/1853/50168.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020
Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford Publications.
Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8
Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2017). Fostering development of work competencies and motivation via gamification. In M. Mulder (Ed.)*, Vol. 23. Competence-based vocational and professional education* (pp. 795–818). Springer. https://doi.org/10.1007/978-3-319-41713-4_37.
Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033
Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w
Schmidmaier, M., Hußmann, H., & Runge, D. M. (2020). Beep beep: Building trust with sound. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8). https://doi.org/10.1145/3334480.3382848
Sciarretta, E., & Alimenti, L. (2021). Smart speakers for inclusion: How can intelligent virtual assistants really assist everybody? In M. Kurosu (Ed.)*, Vol. 12762. Human-computer interaction. Theory, methods and tools* (pp. 77–93). Springer International Publishing. https://doi.org/10.1007/978-3-030-78462-1_6.
Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006
Shang, J., & Wu, J. (2019). Enabling secure voice input on augmented reality headsets using internal body voice. In *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (pp. 1–9). https://doi.org/10.1109/SAHCN.2019.8824980
Shepherd, C., & Liu, N. (2019, october 4). Baidu bets big on smart speakers. *Financial Times, 17*.
Silva-Coira, F., Cortiñas, A., & Pedreira, O. (2016). Intelligent virtual assistant for gamified environments. *PACIS 2016 Proceedings*.
Smith, K., & Abrams, S. S. (2019). Gamification and accessibility. *International Journal of Information and Learning Technology, 36*(2), 104–123. https://doi.org/10.1108/IJILT-06-2018-0061
Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit. (n.d.). Amazon (Alexa). Retrieved July 17, 2021, from https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html.
Stanculescu, L. C., Bozzon, A., Sips, R.-J., & Houben, G.-J. (2016). Work and Play: An Experiment in Enterprise Gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)* (pp. 346–358). https://doi.org/10.1145/2818048.2820061
Steinhaeusser, S. C., Schaper, P., & Lugrin, B. (2021). Comparing a robotic storyteller versus audio book with integration of sound effects and background music. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 328–333). https://doi.org/10.1145/3434074.3447186
Tejedor-Garcia, C., Escudero-Mancebo, D., Cardenoso-Payo, V., & Gonzalez-Ferreras, C. (2020). Using challenges to enhance a learning game for pronunciation training of English as a second language. *IEEE Access, 8*, 74250–74266. https://doi.org/10.1109/ACCESS.2020.2988406
Tejedor-Garcia, C., Escudero-Mancebo, D., Gonzalez-Ferreras, C., Cámara-Arenas, E., & Cardenoso-Payo, V. (2016). Improving L2 production with a gamified computer-assisted pronunciation training tool, tiptoptalk! *Proc. IberSPEECH*, 177–186.
Thiebes, S., Lins, S., & Basten, D. (2014). Gamifying information systems—a Synthesis of gamification mechanics and dynamics. *ECIS 2014 Proceedings*, 1–17.
Trajkova, M., & Martin-Hammond, A. (2020). "Alexa is a toy": Exploring older adults' reasons for using, limiting, and abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3313831.3376760
Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration, 23*(3), 263–280. https://doi.org/10.1037/a0032359
Vtyurina, A., & Fourney, A. (2018). Exploring the role of conversational cues in guided task support with virtual assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). https://doi.org/10.1145/3173574.3173782
Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2018). Gamification of the work floor: A literature review of gamifying production and logistics operations. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 1108–1117).
Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2020). Gamification of production and logistics operations: Status quo and future directions. *Journal of Business Research, 106*, 331–340. https://doi.org/10.1016/j.jbusres.2018.09.011
Weber, P., & Ludwig, T. (2020). (Non-)Interacting with conversational agents: Perceptions and motivations of using chatbots and voice assistants. In *Proceedings of the Conference on Mensch Und Computer* (pp. 321–331). https://doi.org/10.1145/3404983.3405513
Winkler, R., Söllner, M., Neuweiler, M. L., Conti Rossini, F., & Leimeister, J. M. (2019). Alexa, can you help us solve this problem?: How conversations with smart personal assistant tutors increase task group outcomes. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19* (pp. 1–6). https://doi.org/10.1145/3290607.3313090

Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age- and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication, 5*, 600361. https://doi.org/10.3389/fcomm.2020.600361

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., & Qian, F. (2019). Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 1381–1396). https://doi.org/10.1109/SP.2019.00016

Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. O'Reilly Media*.