



# The Effect of Onomatopoeia to Enhancing User Experience in Virtual Reality

Jiwon Oh and Gerard J. Kim<sup>(✉)</sup>

Digital Experience Laboratory, Korea University, Seoul, Korea  
{hanajitsu94, gjkim}@korea.ac.kr

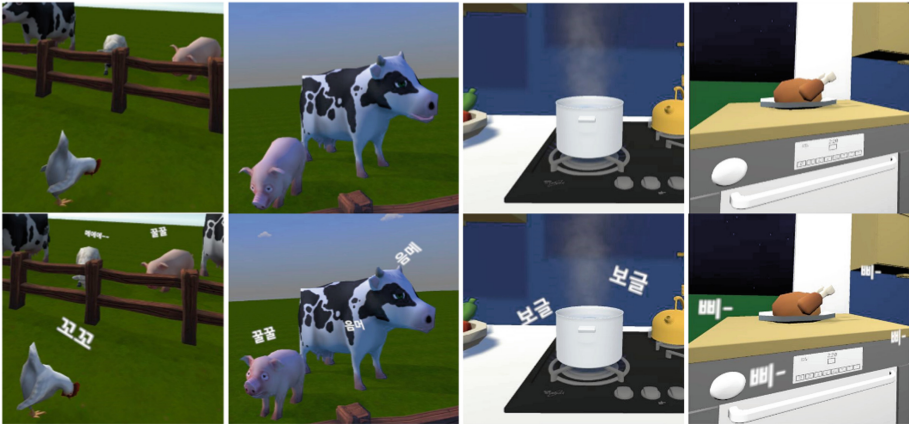
**Abstract.** Onomatopoeia refers to a word that phonetically imitates, resembles, or suggests the sound (or motion that accompanies sound) that it describes. It is often used in comics, games, and videos, along with the actual sound in caption. This usage is a way to emphasize, exaggerate, dramatize and draw attention the situation. In this paper we explore if the use of onomatopoeia, associated with sound feedback, could also bring about similar effects and improve the user experience in immersive virtual reality. We present an experiment comparing the user's subjective experiences and attentive performance in two virtual worlds, each configured in two test conditions: (1) sound feedback with no onomatopoeia and (2) sound feedback with it. Our experiment has found that the moderate and strategic use of onomatopoeia can indeed help direct user attention, offer object affordance and thereby enhance user experience and even the sense of presence and immersion.

**Keywords:** Onomatopoeia · Virtual reality · Sound visualization · User experience

## 1 Introduction

Onomatopoeia refers to a word that phonetically imitates, resembles, or suggests the sound (or motion that accompanies sound) that it describes. Common occurrences of onomatopoeia include animal noises such as “oink”, “meow”, “roar” and “chirp” [16]. Onomatopoeia is much used in paper comics that have no aural feedback, to enrich the static image cuts [4]. However, it is used less in a spoken fashion especially in the Indo-European languages. The extent of use of onomatopoeia can differ between languages and cultures, and Koreans, in particular, make use of onomatopoeia in both written (not necessarily for comics) and spoken ways very frequently. It is often used even along with the actual sound in caption for entertainment shows (in Korea [6]) and games [8]) as a way to dramatize, emphasize, exaggerate and draw attention the situation [19].

On the other hand, one of the important goals to achieve in virtual reality (VR) is to provide a rich content experience by giving users a high level of immersion and presence (feeling as if situated in the [5, 13]). There are variety elements that can affect and improve the sense of presence, including the first person viewpoint, visual realism, interaction, shielding from the operating environment, and physical immersion, just to name a few.



**Fig. 1.** The two test virtual environments: “Animal Farm” (left) and the “Busy Kitchen” (right), “Sound-only” (above) and With-words” (below).

Based on such observations and research results, we explore if the use of onomatopoeia, associated with sound feedback, could also bring about the similar effect, i.e. enhancement in presence and immersion for VR. Note that showing of onomatopoeia is not physically possible in the real world, however, there have been many instance of purely virtual worlds or proper mixing of physically unrealistic effects (e.g. showing motion profiles of a flying ball in a tennis game [1]) such that it can still induce the suspension of disbelief and not come in the way of, or even enhance the level of immersion and presence [5, 13]. On the other hand, we can also expect overusing onomatopoeia can be distracting, too unrealistic and break the sense of immersion and illusion [12].

In the following, we present a pilot experiment (and its results) comparing the user’s subjective user experience, presence and immersion in two virtual worlds, each configured in test conditions: (1) sound feedback with no onomatopoeia and (2) sound feedback with it. First, we shortly review related research in the next section.

## 2 Related Work

There have been several works that has artificially added “text (or glyphs)” and localized “sound word animation” to images [9] and videos [15] with the intent to add the effects of, e.g. dynamics and excitement. As for the use of onomatopoeia, perhaps due to the cultural orientation, there have been just few researches mainly from Japan, as applied to images or computer graphic contents, but not for immersive virtual reality. For example, Wang et al. developed a method for automatically transforming non-verbal video sounds to animated sound words and positioning them near the sound source objects in the video for visualization. Furthermore, they conducted a user study to show that animated sound words helped clarify the sound source and made the video watching more enjoyable [15]. Yamamoto et al. developed a method to transform the

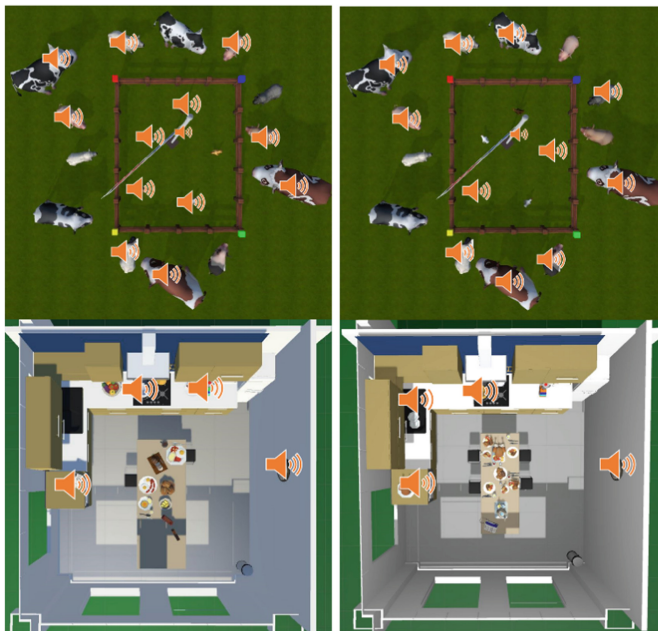
environmental sound into the corresponding onomatopoeia and visualized them in various ways (e.g. varying fonts and sizes to reflect sound qualities and loudness) [18]. In a similar vein, Fukusato and Morishima considered the estimation and depiction of onomatopoeia in computer-generated animation based on physical parameters [3]. More recently, Shimoda and Yani have applied the deep neural network in labeling an image with the appropriate onomatopoeia [12].

Among many presence enhancing elements, we take a note of the role of attention and affordance [10, 11, 17]. The use of onomatopoeia, to direct user's attention, can be regarded as a similar approach. Illusion effects and multimodal effects for enriching VR contents are also related approaches [2, 7, 20].

### 3 Experiment

#### 3.1 Experiment Design

A usability experiment was carried out to assess the projected merits of the use of onomatopoeia, as an additive to the regular sound effect, towards directing user attention, affordance, and thereby to enhancing immersion and presence. The assessment was made over two different virtual reality scenes. Thus, the main factors were the use of onomatopoeia and scene type, making the experiment designed as a 2-factors x 2-levels within subject repeated measure.



**Fig. 2.** The location of sound making object. (above – Animal Farm, below – Busy Kitchen)

### 3.2 Experiment Task

The experimental task involved the subject to experience two different scenes, “Animal Farm” and “Busy Kitchen” (see Fig. 1) and assess them guided by a subjective survey. Each scene type was configured in the two test conditions, i.e. (A) one with ordinary sound effects (denoted “*Sound-only*”), e.g. from the animals (cow, sheep, chicken, pig), cooking objects (kettle, boiling pot, operating mixer, baking oven), and other miscellaneous ones (flag on a pole, clock on the wall, running water from a faucet, ringing phone), (B) the other with the added onomatopoeia (denoted “*With-words*”). Therefore, as a within-subject experiment, there were four virtual environments for the subject to experience and assess, presented in a balanced order: (1) Farm/Sound-only, (2) Farm/With-words, (3) Kitchen/Sound-only, (4) Kitchen/With-words. To minimize any learning effect, the respective scenes (Farm and Kitchen) was slightly reconfigured between experiencing (1) and (2), and (3) and (4): as for the Farm, the number and kinds of animals making sounds were made and distributed differently (the total staying the same), and as for the Kitchen, the types of food prepared, several types of cooking apparatus and time of day (one for breakfast and other supper preparation) were changed. The total number of objects making sounds in the Farm was about 12 and 15 for the Kitchen, distributed evenly throughout the scene (see Fig. 2). For the condition (B), the onomatopoeias were positioned right next or above to the pertaining objects with sufficient contrast for visibility. The actual sound words (onomatopoeia) used in the experiment all common and registered in the standard Koreans dictionary.

### 3.3 Experiment Procedure

Thirty six people (18 females and males) aged between 19 and 38, (mean = 23.4, SD = 3.3) participated in the study. Each subject experienced and freely navigated around the four test VR scenes (for 5 min each) in an order balanced around the content theme (Farm/Kitchen) and the experimental factor (with or without the onomatopoeia). The subject was told to count the total number of notable objects (e.g. animals, kitchen objects) and sound making objects, and also identify which objects or how many of them were making sounds. Such measures were collected to evaluate quantitatively the effect of the onomatopoeia with respect to user focus/attention and awareness.

After experiencing the each scene (2-conditions x 2-content types), the subject answered a survey which asked of the general usability (fatigue, sickness, naturalness), content perception (liveliness, realism, affordance and attention), presence/immersion, and enjoyment/preference level, all in 7 level Likert scale (see Table 1). Finally, the subjects were asked of various situations in the respective scenes (see Table 2) and to describe the experience virtual space in words, which was linguistically analyzed for its richness in expression similar to the approach in [14].

The testing platform was implemented with Unity3D and run on a desktop PC with the HTC VIVE head-set and interaction controller (for navigation). Further experiment procedural details are omitted due to space restriction. Our basic expected outcomes

**Table 1.** The subjective survey assessing a variety of aspects in user experience (all answered in 7 level Likert scale).

<b>Fatigue</b>
F1 I am mentally tired looking around the environment
<b>Simulation sickness</b>
S1 I feel sick and nauseated
S2 I have a headache
S3 I feel dizzy
S4 I have a problem in concentrating
<b>Naturalness</b>
N1 The visual scene looks natural
<b>Perceived Realism</b>
R1 The visual realism is high
R2 The overall scene looks realistic
<b>Perceived Dynamism/Liveliness</b>
D1 The content feels to be lively and dynamic (e.g. lots of activities, things moving, sounds, etc.)
<b>Affordance/Attention</b>
A1 My attention is naturally drawn to various objects
A2 I am able to attend to objects making sounds
<b>Immersion</b>
I1 I feel immersed in the virtual space (spatially)
I2 I am immersed into the task of viewing the virtual space (mentally)
<b>Presence</b>
P1 I felt as if I was at the Farm/Kitchen
P2 I lost the track of time while experiencing the virtual space
P3 I felt and was drawn to act upon and interact with the objects
<b>Enjoyment/Preference/Satisfaction</b>
E1 I would like to experience the virtual space again if given the chance
E2 I enjoyed the virtual space

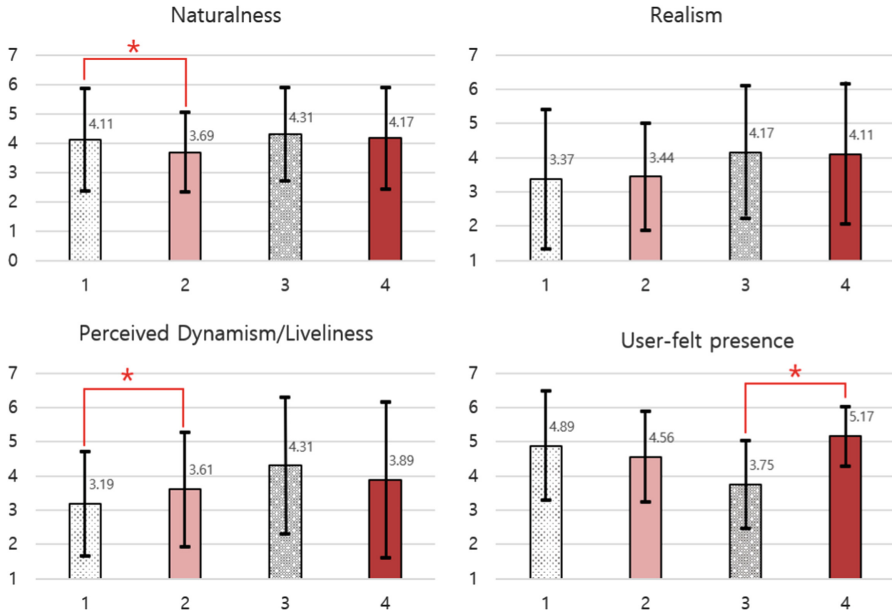
**Table 2.** Object attention questions (for Busy Kitchen - Breakfast scene).

What was being cooked in the pot?
What was being grinded in the mixer?
Where was the telephone?
How many different breads were there on the table?
What color was the trash can?

were that both quantitative (e.g. more accurate assessment of the objects in the scene) and subjective (e.g. preferred, helpful in perceiving the content to be rich, and leading to a higher level of awareness, presence, and immersion) performance would be better with the additive use of the onomatopoeia to sound.

## 4 Results and Discussion

See Fig. 3.



**Fig. 3.** The levels of various aspects of user experience for the four conditions: (1) Farm – Sound only, (2) Farm – With words, (3) Kitchen – Sound only, (4) Kitchen – With words.

### 4.1 General Usability (Fatigue, Naturalness, Sickness)

ANOVA/t-test was applied for the analysis of these dependent variables (corresponding responses to the survey in Table 1). A statistically significant difference was found in the fatigue level (F1) when onomatopoeias were used (vs. not using) for the Farm scene ( $t(35) = -2.935, p = 0.006^*$ ), but not for the Kitchen. The Farm containing relatively more onomatopoeias per unit area (and thus also for the momentary visual span) seems to have caused this difference.

Subjects felt the scene to be, expectedly, relatively unnatural (N1) when the onomatopoeias were used for the Farm scene ( $t(35) = 1.898, p = 0.066$ ), but again not for the Kitchen. Considering that the overall realism did not suffer despite the use of the sound words (see Sect. 4.2), it seems this result emanates from the scene/object specific characteristics (e.g. animals making occasional sounds/text vs. cookerries making constant sounds).

Sickness levels were low (below 2 out of 7) as the virtual environment incurred no substantial navigation (just slow-paced local exploration) and no differences were found between Sound-only and With-words.

## 4.2 Content Perception (Realism, Liveliness)

It was actually projected that the additive use of onomatopoeia possessed a danger of lowering the realism (R1, R2 - since obviously, such a thing does not exist in the real world). However, no significant effect was found on this variable. Aside from the usual suspension of disbelief, we attribute this partially to the fact that most people are already accustomed to using and seeing onomatopoeia in everyday living, comics and other media form (at least in Korea).

ANOVA/t-test found that users generally felt the scenes to be more dynamic and livelier (D1) when the onomatopoeias were used in the Animal Farm ( $t(35) = -2.076$ ,  $p = .045$ ), but not in the Busy Kitchen. Animals compared to kitchen objects moved and were expected to do so freely and onomatopoeia helped the users feel improved dynamics and liveliness. In contrast, the kitchen objects were mostly static to begin with, and the added sound words would be ineffective, and rightly so. We believe that the dynamism and liveliness can further increase for moving objects, if we also animate the words in tune with the object motion. The perceived liveliness might improve even for the non-moving objects with onomatopoeia animated in synch with the sounds made.

Even though the scenes were not interactive, users reported they were led to interact with the objects in both contents, implicating a strong sense of affordance when the onomatopoeia were used. Unfortunately, to the two relevant questions of object attention and affordance (A1 and A2), statistically significant effects were not found. However, other measures of object attention were found to be different with a statistically significant difference (see Sect. 4.4).

## 4.3 Immersion and Presence

ANOVA/t-test found a significant difference for the user felt immersion and presence (P1, P3 collectively) for the Kitchen scenes (Breakfast -  $p = .002$ ; Supper -  $p = .027$ ), but not for the Farm scenes. In the Busy Kitchen scenes, user felt immersion and presence were higher when onomatopoeias are used. Note that it was observed and found that subjects felt a substantial level of affordance and interest (for interaction possibility) through the onomatopoeia (see Sect. 4.2). This seems to have led the subjects to perceive higher immersion and presence. On the other hand, in the Farm scene, subjects felt that the animals were arranged too crowdedly and there were excessively many additive texts (in unit area), raising the level of fatigue, and eventually distracting them and breaking the sense of immersion/presence. The onomatopoeia did not help the scene to be felt interactive because of the very nature of the content type (nothing much to do with animals vs. kitchen objects having clear functional purposes). On the other hand, subjects reported that it was much easier to focus, attend and localize the objects in the less crowded Kitchen scene with the onomatopoeia (see Sect. 4.2). In fact, there are research works indicating a strong correlation between presence and affordance/attention [17] and vice versa with distraction [17].

#### 4.4 Situation Awareness

The situation awareness was evaluated in two ways: (1) accuracy in the counting and naming objects (total or sound making) (2) linguistically analyzing the subject's scene descriptions (evaluated based on the richness of the vocabularies or expressions). In both the Animal Farm and Busy Kitchen scenarios, subjects were generally more correct in the counting and identifying objects and grasping features of objects when there was onomatopoeia either for the total or the sound making ones (we omit the exact statistical figures). In fact, when there was no onomatopoeia, the counts were significantly inaccurately less. Despite the no effect not being found for questions A1 and A2 (see Sect. 4.2), based on this result, we believe the subjects had an object-wise understanding.

Subjects were also asked to freely write out and describe the scenes they experienced. We applied the neuro-linguistic programming to code the subjective experience [14]. In general, the subject's descriptions were richer when the onomatopoeia accompanied the sound effect. For example, the descriptions were longer and changed from simple depiction to a more refined story (see Table 3). Also, the descriptions were partly imagined out with colorful vocabularies with the perceptual position changed to the first person from the third (see Table 4).

**Table 3.** Examples of Farm scene description from a subject (Sound-only and With-words conditions) – originally written in Korean and translated to English.

<b>Sound-only</b>	“Cows, pigs, chickens and sheep cry outside the fence and are observing what is going on in the fence.”
<b>With-words</b>	“The pigs, sheep and cows have gone outside because the fence was low, but I and the chickens cannot get over it and feeling restless while, the other animals are urging us to come over.”

**Table 4.** Examples of Kitchen scene description from a subject (Sound-only and With-words conditions) – originally written in Korean and translated to English.

<b>Sound-only</b>	“Based on the food on the table, it seems like it is morning time, and from the sounds, it feels like a hectic situation.”
<b>With-words</b>	“I think I was invited for a meal and looking at the kitchen, it seems the host briefly stepped out.”

#### 4.5 Enjoyment/Preference

ANOVA/t-test found mixing results for the user enjoyment/preference as well. For example, subjects showed a significantly more willingness/preference for the content with the onomatopoeia for the Busy Kitchen, but not for the Animal Farm. As posited, this seems to relate to the difference in the object affordance and attention to them, leading the subjects to raise their interest and the desire to interact.



## 5 Conclusion and Future Work

In this paper we verified that using onomatopoeia as added to the sound feedback could improve the user experience in virtual reality. Our experiment result has found that the judicious use of onomatopoeia can indeed help direct user attention, offer object affordance and thereby enhance user experience and even the sense of presence and immersion, without degrading the perceived realism. Summarizing the specifics, important factors that would make an effective application of onomatopoeia to sounds were: (1) having a scene (and associated objects) that has an interactive theme and highly functional, (2) not crowding the scene with too much additive texts.

In the future, we would like to investigate in how to present the onomatopoeias in various ways, such as in their position, pose, size, animation, timings, and how they might also affect the user experience in virtual and mixed reality. Using mimetic words for emphasizing the object movement is also another consideration along this line. The ultimate goal is to derive a specific guideline for utilizing onomatopoeia in VR/MR. Finally, how onomatopoeia can help give experience to the hearing-impaired or how it can be associated with other modalities such as the haptic and visual (e.g. object animation) are another interesting future work topic.

**Acknowledgements.** This work was partially supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government (MEST) (NRF-2015M3A6A3076490), and by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2017-0-00179, HD Haptic Technology for Hyper Reality Contents).

## References

1. VR Tennis Online. <https://colopl.co.jp/en/products/vr/vrtennis.php>. Accessed 15 Aug 2018
2. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: a survey of principles, models and frameworks. In: Lalanne, D., Kohlas, J. (eds.) *Human Machine Interaction*. LNCS, vol. 5440, pp. 3–26. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-00437-7\\_1](https://doi.org/10.1007/978-3-642-00437-7_1)
3. Fukusato, T., Morishima, S.: Automatic depiction of onomatopoeia in animation considering physical phenomena. In: *Proceedings of the Seventh International Conference on Motion in Games*, pp. 161–169. ACM (2014)
4. Guynes, S.A.: Four-color sound: a Peircean semiotics of comic book onomatopoeia. *Public J. Semiot.* **6**(1), 58–72 (2014)
5. Heeter, C.: Being there: the subjective experience of presence. *Presence Teleoperators Virtual Environ.* **1**(2), 262–271 (1992)
6. Jung, S.-Y.: A Study on the feature and the function of TV visual character: focusing on the real variety shows. *Korean J. Journal. Commun. Stud.* **53**(6), 153–176 (2009)
7. Lee, J., Kim, G.J.: Visual information for inducing interactivity and presence. In: *Proceedings of the Korean Society for Emotion and Sensibility Conference*. The Korean Society for Emotion and Sensibility
8. Masuch, M., Röber, N.: Game graphics beyond realism: then, now and tomorrow. Level UP: Digital Games Research Conference. DIGRA, Faculty of Arts, University of Utrecht (2004)

9. Nienhaus, M., Dollner, J.: Depicting dynamics using principles of visual art and narrations. *IEEE Comput. Graph. Appl.* **25**(3), 40–51 (2005)
10. Norman, D.: *The design of everyday things: Revised and expanded edition*. Constellation (2013)
11. Norman, D.A.: *The psychology of everyday things (The design of everyday things)* (1988)
12. Shimoda, W., Yanai, K.: A visual analysis on recognizability and discriminability of onomatopoeia words with DCNN features. In: *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE
13. Slater, M., Usoh, M.: Presence in immersive virtual environments. In: *IEEE Virtual Reality Annual International Symposium*, pp. 90–96. IEEE (1993)
14. Slater, M., Usoh, M.: Representations systems, perceptual position, and presence in immersive virtual environments. *Presence Teleoperators Virtual Environ.* **2**(3), 221–233 (1993)
15. Wang, F., Nagano, H., Kashino, K., Igarashi, T.: Visualizing video sounds with sound word animation to enrich user experience. *IEEE Trans. Multimed.* **19**(2), 418–429 (2017)
16. Wikipedia. Onomatopoeia — Wikipedia, The Free Encyclopedia (2018). <http://en.wikipedia.org/w/index.php?title=Onomatopoeia&oldid=854819117>. Accessed 15 Aug 2018
17. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. *Presence* **7**(3), 225–240 (1998)
18. Yamamoto, T., Matsubara, M., Saito, H.: Visualization of environmental sounds using onomatopoeia and effective fonts. *J. Soc. Art. Sci.* **11**(1), 1–11 (2012)
19. Jick, J.Y., Seok, L.H.: The expression techniques of cartoon style in the subtitles of TV entertainment programs. *J. Digit. Des.* **15**(3), 779–788 (2015)
20. Yuan, Y., Steed, A.: Is the rubber hand illusion induced by immersive virtual reality? In: *2010 IEEE Virtual Reality Conference (VR)*, pp. 95–102. IEEE (2010)