

Factor analysis of the sense of presence in daily scenes

Masayuki Iizuka (1), Kenji Ozawa (2), Yuichiro Kinoshita (2) and Kazutomo Fukue (2)

(1) Faculty of Engineering, University of Yamanashi, 4-3-11 Takeda, Kofu 400-8511, Japan

(2) Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4-3-11 Takeda, Kofu 400-8511, Japan

PACS: 43.66.Lj, 43.66.Pn

ABSTRACT

Although the sense of presence is crucial for evaluating AV (audio-visual) equipment, the meaning of the sense remains vague. To clarify the structure of the sense, we conducted an experiment on AV contents. Initially 374 adjectives, which can express the sense of presence, were collected by interviews, a dictionary, and magazines. Then this number was reduced to 29 pairs of adjectives based on the KJ (Kawakita Jiro) method. Thirty-three daily scenes (e.g. a passing train) were recorded with a high-definition video camera while the sounds were recorded using a dummy head. Each AV content was reproduced with a 65-inch display and headphones, and evaluated by the SD (Semantic Differential) method using a five-point category scale for the 29 pairs of adjectives. Sixteen subjects in their twenties participated in the experiment. The experimental data was analyzed by the factor analysis method, and seven factors were extracted where the accumulated contribution ratio was 82.5%. The first factor, F1, represents *activity* because the pairs such as *quiet-noisy* and *persisting-nonpersisting* have the largest loadings on this factor. Similarly F2 through F7 represent *naturalness*, *dailiness*, *psychological loading*, *potency*, *entertainment*, and *decorativeness*, respectively.

INTRODUCTION

Although the sense of presence is a key to evaluate the performance of current AV (Audio-Visual) equipment, the characteristics of the sense have yet to be sufficiently clarified. The lateral meaning of presence includes the subjective experience of being in one place or environment when one is physically situated in another [1]. According to this meaning, the sense is determined by how accurately an AV system reproduces the physical characteristics of the original field. However, our preliminary studies on auditory presence, which is the presence evoked by auditory stimuli, revealed that the perceived presence depends on the reproduced content even if one recording and reproduction system was used [2, 3]. Thus, we divided sense into two parts: *system presence*, which is determined by the characteristics of the AV system used, and *content presence*, which depends on the characteristics of the contents reproduced by a system [3]. This article focuses on *content presence*.

Previous investigations have suggested a multidimensional structure of the sense of presence [1, 4-8]. Kim and Biocca [4] have shown two factors, *arrival* (a sense of being in the mediated environment) and *departure* (a sense of not being in the physical environment) by factor analysis. Schubert *et al.* [5] have revealed three components related to presence: *Spatial presence*, *Involvement*, and *Realness*. Lessiter *et al.* [6] have indicated four factors, *physical space*, *engagement*, *naturalness*, and *negative effects*, whereas Witmer *et al.* [7] have suggested four factors, *involvement*, *adaptation/immersion*, *sensory fidelity*, and *interface quality*, based on the results of principal-components analyses of presence questionnaire data from participants following exposure to immersive virtual environments. All the above factors are related to *system presence*.

In contrast, Teramoto *et al.* [8] have shown four factors: *evaluation*, *potency*, *activity*, and *mechanicalness*. They employed a factor analysis of questionnaire data based on memories or mental images of the subjects. Because the participants of their questionnaire did not have experiences of any special virtual environment, the presence factors seem to be closely tied to *content presence*. However, the multidimensional construct of *content presence* remains vague because they did not use actual AV content.

Thus, to clarify the multidimensional structure of *content presence*, AV contents reproduced by a unique system were evaluated. The evaluation experiment was conducted by the SD (semantic differential) method [9], and the multidimensional structure was revealed by factor analysis.

COLLECTION OF AV CONTENTS

Selection of AV contents

In this study, daily scenes were addressed as AV contents. Our previous study suggests that the movements of sound and visual images are important in evaluating presence [2, 3]. Thus, we selected daily scenes with an emphasis on the movement of sound and visual images. Table 1 shows the 33 AV contents selected. Here, the meaning of *movement* is not moving in the picture, but changing in position relative to an observer. For example, the visual image of a *river's current* is part of a *stationary visual image*.

Furthermore, the variability in presence was also taken into account when selecting the AV contents. Our previous study on auditory presence showed that contents such as a *passing train* and a *river* were evaluated to have a higher presence [2].

On the other hand, contents such as *sounds of winds* and *insects singing* were assessed as having a lower presence. These contents were included in the selected AV contents in this study.

Recording and reproduction of AV contents

The aforementioned AV contents were recorded using a HD (high-definition) video camera (Sony, XDCAM EX PMW-Ex1 or PMW-EX1R). Moving pictures were recorded in the full-HD format (1920 x 1080/60i).

To accurately record the original sound field, the binaural technique [10] with a dummy head (Koken, SAMRAI) was used. Two microphones were mounted in the ears of the dummy head, and the sounds were recorded in the LPCM (Linear Pulse Code Modulation) format with an accuracy of 16 bits and a sampling frequency of 48 kHz.

Each AV content, which had a duration of approximately 30 s, was reproduced in an office room as a stimulus in an interview and in an experiment described below. Video stimuli were presented with a 65-inch full-HD display (Sharp, AQUOS LC-65RX1W), while auditory stimuli were reproduced through headphones (Sennheiser, HD600). The subject sat on a sofa while receiving the AV stimuli. The distance between the display and a subject was 2.4 m, which corresponds to 3H (Height of a display) in the ITU standard [11].

COLLECTION OF EXPRESSIONS

Interviews to extract expressions

The first step to collect the appropriate expressions for evaluating the sense of presence was to conduct interviews. Sixteen undergraduate or graduate students (ten males and six females) with normal vision and hearing acuity participated in the interviews. The participants were divided into four groups of four people. Each group included at least one female subject to avoid a gender difference in the expression of the presence evaluation.

Five different AV contents were assigned to each group as stimuli. Subjects were exposed to a stimulus with the full-HD display and headphones, and then he or she was asked the following six questions by an experimenter sitting beside him or her.

- Q1: How did you feel the sense of the presence of this AV content?
- Q2: What were the reasons for your answer to Q1?
- Q3: What was your impression of the sound(s) in this content?
- Q4: What was your impression of the moving picture in this content?
- Q5: Did you have something specifically impressive or interesting with respect to this content?
- Q6: Can you identify the location of this content?

Among these questions, Q2 was critical because expressions existing in the answer of Q2 were expected to be candidates of expressions for the presence evaluation. Similarly, we expected to extract expressions for auditory and visual presence from the answers of Q3 and Q4, respectively. Question 5 was intended to extract further impressions of the subject. The answers to Q1 and Q6 were recorded, but not used in analysis.

This sequence was repeated for the five contents assigned to each group. From these interview sessions, 167 expressions were extracted. Although most of the expressions were adjectives, some were sentences (e.g. I tried to find the real thing involuntarily.). At this point, every expression was preserved to prevent losing possible candidates.

Extraction of expressions from a dictionary and magazines

The first author extracted potential expressions for evaluating *content presence* from an adjective dictionary [12]. This resulted in 104 expressions. Moreover, two of the authors extracted 119 expressions from magazines relating music and cinema [13, 14].

Table 1. Collected AV contents.

	Visual images are moving.	Visual images are stationary.
Sound images are moving.	[N = 7] Scene of passing vehicles, Scene of a passing train (Short-distance view), Scene of a passing train (Long-distance view), Scene with pedestrians and buses in an open space of a railway station, Scene of a passing boat, Scene of a small stone skipping on the surface of a lake, Scene of playing catch	[N = 3] Scene of playing Japanese-archery, View from a person sitting in a office, Loudspeakers reproducing a piece of music
Sound images are stationary.	[N = 6] View from a pedestrian in a corridor of a building (Fast walking. Footsteps are stationary for the walker.), View from a pedestrian in a corridor of a building (Slow walking), View from a pedestrian on a gravel road, View from a person sliding on a playground slide with rolling bars (Sounds from the playground slide are stationary for the person.), View from a person in a ropeway gondola (Sounds in the gondola are recorded.), View from a person in a moving car (Sounds of its engine are recorded.)	[N = 17] Night scene where insects are singing, View from a person listening to sounds of a <i>suikinkutsu</i> (Japanese traditional large bottle that is embedded in a ground. A listener can enjoy the sounds of water dropping into it.), Waterfall, Scene of a school at dusk (Sound of chimes and bird singing are heard). Current of river with murmuring, Long-distance view of Mt. Fuji with the hum of voices from sightseers. Trees murmuring in the wind, Water fountain at night, <i>Shishiodoshi</i> (Japanese traditional equipment that makes an impulsive sound by hitting a stone with a rod of bamboo, which is driven by dropping water). Scene of cooking, Steaming kettle, Reverberant voice in pedestrian subway, Bustle in a park, Nightscape, Lakeside, Ringing cellphone, Scene of typing



(a) Passing vehicles



(b) Passing train (Short-distance view)



(c) Waterfall



(d) Water fountain at night

Figure 1. Examples of collected AV contents.

Grouping of collected expressions

As described above, a total of 390 expressions were collected. However, some had the same meaning (e.g. *Araarashii* and *Arai* in Japanese). Thus, redundant expressions were excluded, which reduced this number to 374.

These 374 expressions were grouped using the KJ (Kawakita Jiro) method [15]. Six male subjects, who were not part of the interviews, participated in this task. Each of the expressions was written on a small card and placed on a table. The subjects tried to group the expressions by meaning and then selected a representative expression for each group. If it was difficult to choose a representative for the group, then a new expression was written on another card and used as the representative. Consequently, 63 groups were obtained.

Forming adjective pairs for the SD experiment

As the final step to collect the expressions, the authors formed pairs of expressions, which were subsequently used in the following experiment by the SD method. Expressions representing *system presence* were excluded. All of these were sentences (e.g. Sound localization was well.). Then words that did not have an antonym were omitted. Finally adjective pairs with opposite meanings were formed. Consequently, 29 pairs were obtained.

IMPRESSION EVALUATION EXPERIMENT BY THE SD METHOD

Method

Sixteen subjects (12 males and 4 females), where fourteen of them participated in the interview, were presented the 33 AV contents in random order. The subjects rated their impres-

sions of each stimulus using a five-point category scale for the 29 pairs of adjectives.

Factor analysis

Prior to the factor analysis, the raw data with respect to each adjective pair was examined. If more than half of the subjects (i.e. more than eight) evaluated an adjective pair as *neutral*, which corresponded to the center of the scale for more than half of the stimuli (i.e. more than 16), the adjective pair was deemed to be inappropriate for evaluating the *content presence* and such pairs were excluded. Consequently, two pairs, *new-old* and *stylish-unstylish*, were excluded. The average evaluation scores over all subjects for the other 27 pair of adjectives were analyzed by factor analysis (Principal factor solution, varimax rotation).

Based on the criterion that eigenvalues must be greater than one, we extracted seven factors. Table 2 shows factor loadings for every pair of adjectives, contribution ratios, and cumulative contribution ratios in terms of the seven factors. As shown in the table, the cumulative contribution ratio up to the 7th factor was 82.5%, indicating the seven factors well explain the variance in the data.

In Table 2, the pairs are sorted by the largest factor loading, which is underlined. All items loaded at least 0.4 on one or factors. The first factor, F1, represented *activity* because pairs such as *quiet-noisy* and *persistent-plain* had the largest loadings on this factor. The second factor, F2, indicated *naturalness* because pairs such as *artificial-natural* and *intuitive-logical* had the largest loadings on this factor, whereas the third factor, F3, was concerned with *dailiness* because this factor was composed of *realistic-fantastic* and *unique-ordinary*. The fourth factor, F4, was regarded as *psychological loading* because pairs such as *tense-relaxing* and *scary-reassuring* had the largest loadings. The fifth factor, F5, re-

flected *potency* because this factor was composed of *dignified-frivolous* and *heavy-light*. The sixth factor, F6, was related to *entertainment* because pairs such as *delightful-boring* and *enthusiastic-calm* had the largest loadings. Finally, the seventh factor, F7, was associated with *decorativeness* because this factor was composed of *gorgeous-plainly* and *simple-complex*. Figure 2 summarizes these results.

Discussions

As described in the Introduction, Teramoto *et al.* [8] have demonstrated that the sense of presence has a multidimensional construct with *evaluation*, *potency*, *activity*, and *mechanicalness*. Comparing our study to their results indicates that two factors, *activity* and *potency*, correspond. Our F7, *decorativeness*, seems similar to *evaluation* in their study because *gorgeous-plainly* is occasionally used as an adjective pair of *evaluation*. Furthermore, F2, *naturalness*, has the

opposite meaning of their *mechanicalness*. Thus, their four factors can be found in the seven factors of this study. In other words, this study revealed three new factors: *dailiness*, *psychological loading*, and *entertainment*. The difference between the studies is due to the varying methods; their study was conducted using a questionnaire based on memories or mental images of the subjects, whereas our results were based on actual evaluations of AV contents.

CONCLUSIONS

To clarify the multidimensional construct of *content presence*, 33 AV contents from daily scenes were evaluated by the SD method using 29 pairs of expression words, which were selected as candidates to describe the sense of presence. Based on the factor analysis results, seven factors were extracted: *activity*, *naturalness*, *dailiness*, *psychological loading*, *potency*, *entertainment*, and *decorativeness*. A further study is

Table 2. Factor loadings for 27 pairs of expressions after a varimax rotation. Largest factor loading for each pair is underlined.

	F1	F2	F3	F4	F5	F6	F7
quiet – noisy	<u>0.86</u>	-0.09	0.20	-0.12	-0.09	-0.19	-0.37
persisting – nonpersisting	<u>-0.81</u>	-0.22	-0.09	0.03	0.15	0.07	0.07
comfortable – uncomfortable	<u>0.66</u>	0.52	0.25	-0.30	-0.05	0.04	0.25
lovable – hateable	<u>0.64</u>	0.46	0.39	-0.05	0.09	0.26	0.29
busy – unhasty	<u>-0.63</u>	-0.26	-0.44	0.42	0.04	0.14	0.22
congenial – uncongenial	<u>0.59</u>	0.58	0.37	-0.17	-0.02	0.19	0.23
artificial – natural	-0.18	<u>-0.70</u>	-0.25	0.42	-0.15	0.16	0.22
open – closed	0.34	<u>0.69</u>	0.06	-0.24	0.03	0.08	0.13
intuitive – logical	-0.01	<u>0.66</u>	0.15	0.04	-0.03	0.24	-0.04
clear – murky	0.43	<u>0.62</u>	0.52	-0.05	0.09	-0.29	0.07
beautiful – ugly	0.46	<u>0.55</u>	0.52	-0.14	0.20	-0.09	0.27
large-scale – small-scale	0.18	<u>0.55</u>	0.24	-0.16	0.52	0.08	0.33
bright – dark	-0.10	<u>0.50</u>	-0.02	-0.44	-0.10	0.44	0.20
realistic – fantastic	-0.22	-0.20	<u>-0.88</u>	0.07	-0.17	0.08	-0.06
unique – ordinary	0.19	0.18	<u>0.86</u>	0.09	0.23	0.09	-0.05
sharp – dull	-0.09	-0.12	0.03	<u>0.91</u>	0.07	-0.09	0.11
tense – relaxing	-0.24	-0.40	-0.18	<u>0.68</u>	0.28	0.07	-0.34
strong – weak	-0.26	0.13	-0.01	<u>0.66</u>	0.61	0.09	0.03
scary – reassuring	-0.32	-0.45	-0.11	<u>0.57</u>	0.40	0.11	-0.32
dignified – frivolous	0.01	0.12	0.22	0.09	<u>0.91</u>	-0.02	0.10
heavy – light	-0.15	-0.09	0.11	0.23	<u>0.89</u>	-0.01	0.04
delightful – boring	0.17	0.27	0.28	-0.06	0.01	<u>0.78</u>	0.31
enthusiastic – calm	-0.54	0.03	-0.11	0.07	0.00	<u>0.60</u>	0.20
dynamic – static	-0.28	-0.20	-0.36	0.08	0.25	<u>0.52</u>	0.02
warm – cold	-0.04	0.27	-0.25	-0.44	-0.32	<u>0.47</u>	0.12
gorgeous – plainly	0.00	0.24	0.16	0.17	0.34	0.22	<u>0.71</u>
simple – complex	0.18	0.05	0.07	0.11	0.01	-0.17	<u>-0.74</u>
Contribution ratio (%)	16.1	15.9	11.9	11.7	11.1	7.9	7.9
Cumulative contribution ratio (%)	16.1	32.0	43.9	55.6	66.8	74.7	82.5

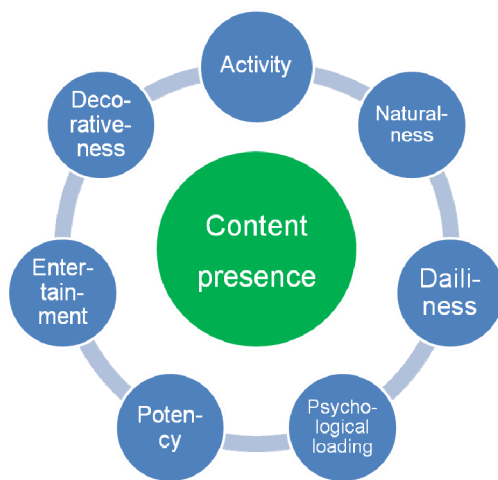


Figure 2. Factors involved in content presence.

necessary to investigate the relationship between these factors and the physical characteristics of moving pictures and sounds in AV contents.

ACKNOWLEDGEMENTS

This study was supported by the National Institute of Information and Communications Technology (NICT) of Japan. The authors thank Mr. Tetsuya Goda, Mr. Takuya Kishita, Ms. Miyu Mashima, Mr. Takumi Nakama, Mr. Kodai Saito, Mr. Tomohiro Sato, Mr. Kento Shirakawa, Mr. Masahiro Takayanagi, and Mr. Satoshi Tsukanaka of the University of Yamanashi for their assistance in collecting the AV contents.

REFERENCES

- 1 B.G. Witmer and M.J. Singer, "Measuring presence in virtual environments: A presence questionnaire" *Presence* **7**, 225–240 (1998).
- 2 K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone, "Contents which yield high auditory-presence in sound reproduction" *Kansei Engineering International*, **3**, 25–30 (2002).
- 3 K. Ozawa and Y. Chujo, "Content presence vs. system presence in audio reproduction systems" *Proc. 2nd International Symposium on Universal Communication (ISUC2008)*, 50–55 (2008).
- 4 T. Kim and F. Biocca, "Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion" *J. Computer Mediated Communication* **3**, on-line (1997).
- 5 T.W. Schubert, F. Friedmann, and H.T. Regenbrecht, "Decomposing the sense of presence: factor analytic insights" *Proc. Presence 1999: 2nd Int. Workshop on Presence*, 1–5 (1999).
- 6 J. Lessiter, J.F. Freeman, E. Keogh, and J. Davidoff, "Development of a new cross-media presence questionnaire: The ITC-Sense of presence inventory" *Proc. Presence 2000: 3rd Int. Workshop on Presence*, 1–7 (2000).
- 7 B.G. Witmer, C.J. Jerome, and M.J. Singer, "The factor structure of the presence questionnaire" *Presence* **14**, 298–312 (2005).
- 8 W. Teramoto, K. Yoshida, N. Asai, S. Hidaka, J. Gyoba, and Y. Suzuki, "What is 'sense of presence'? A Non-researcher's understanding of sense of presence" *Trans. The Virtual Reality Society of Japan* **15**, 7–16 (2010). (In Japanese)
- 9 C.E. Osgood, "The nature and measurement of meaning" *Psychological Bulletin* **49**, 197–237 (1952).
- 10 H. Møller, "Fundamentals of binaural technology" *Applied Acoustics* **36**, 171–218 (1992).

- 11 ITU-BT.710-4, "Subjective assessment methods for image quality in high-definition television" (1998).
- 12 Y. Hida and H. Asada, *Dictionary of Current Adjective Usage* (Tokyodo Publishing, Tokyo, 1991). (In Japanese)
- 13 *rockin'on*, Feb. 2010 (rockin'on, Tokyo, 2010). (In Japanese)
- 14 *DVDvision*, Jan. 2010 (Hinode publishing, Tokyo, 2010). (In Japanese)
- 15 J. Kawakita, *The KJ Method* (Chuokoron-Shinsha, Tokyo, 1986) (In Japanese)