The sense of presence is crucial to evaluate the performance of audio-visual (AV) equipment and content. Previously, the overall presence was evaluated for a set of AV content items by asking subjects to judge the presence of the entire content item. In this study, the sense of presence was evaluated for a time-series using the method of continuous judgment by category. Specifically, the audio signals of 40 content items with durations of approximately 30 s each were recorded with a dummy head, and then presented as stimuli to subjects via headphones. The corresponding visual signals were recorded using a video camera in the full-HD format, and reproduced on a 65-inch display. In the experiments, 20 subjects evaluated the instantaneous sense of presence of each item on a seven-point scale under two conditions: audio-only or audio-visual. At the end of the time-series, the subjects also evaluated the overall presence of the item by seven categories. Based on these results, the effects of visual information on the sense of presence were examined. The overall presence is highly correlated with the ten-percentile exceeded presence score, $S_{10}$, which is the score that is exceeded for the 10% of the time during the responses. Based on the instantaneous presence data in this study, we are one step closer to our ultimate goal of developing a real-time operational presence meter.

**key words:** content presence, method of continuous judgment by category, audio-visual information

1. Introduction

The term “sense of presence” is used to evaluate the performance of audio-visual (AV) equipment (e.g., a television (TV) and an audio set) and the AV content such as a movie or a sports program. Although several studies have examined the sense of presence [1]–[5], the full property of the sense has yet to be clarified. Thus, a presence meter has yet to be realized. Such a meter would be useful for AV content creators because it would allow the sense of presence evoked for viewers and listeners of the created content to be estimated. Moreover, such a meter would be beneficial for purchasers of AV equipment if the degree of the presence evoked by an AV system is indicated in its specifications.

We have examined the sense of presence by dividing the sense into two aspects: **system presence** and **content presence** [6]–[10]. System presence is determined by the characteristics of the AV system used, while content presence depends on the characteristics of the contents reproduced by a system. Both types of presence must be considered to develop a presence meter. However, a set of standard content items must be prepared to measure system presence. Therefore, this paper focuses on content presence using a single recording and reproduction system.

Previously, we measured the overall presence of AV content items, which each lasted about 30 s [6]–[10], because it was assumed that the sense of presence of such a system or an item could be effectively expressed with a single number from the viewpoint of evaluating the performance of a system or a content item. However, for a presence meter, the instantaneous presence, which depends on the content item, must also be expressed. Herein the sense of presence is evaluated for a time-series. Although a previous study by Oode et al. [11] evaluated the time-series of a piece of music to examine the effects of music type and the reproduction system on the sense of presence of reproduced music, the present study examines content items valuable for broadcasting (e.g., landscapes, soundscapes, various events as well as music). Additionally, because radio and TV are currently major media, experiments are conducted under two conditions: “audio-only” and “audio-visual”. The effects of adding a corresponding moving picture to a sound are discussed.

2. Experimental Methods

2.1 AV Content Items and a Recording-Reproduction System

Table 1 shows the 40 content items used in the current experiment, where each item lasts about 20 to 40 s. These items are the same as those in our previous studies [8], [9], which indicated that the movements of sound and visual images are important to evaluate the sense of presence. In Table 1, the items are arranged into four groups according to the movements of sound and visual images. Here movement does not mean a moving picture, but a change in position relative to the observer (a dummy head in the actual situation or the subject in the experiment). For example, the visual image of the view of a waterfall is part of a stationary visual image. Table 1 indicates the movement direction, while Table 2 ex-
Table 1  AV content items used as stimuli and changes in presence from audio-only to audio-visual conditions. Major sound sources are included in the sample name. If sound sources are not indicated in the sample name, they are indicated in parentheses [A ( )]. Arrows in parentheses [ ] indicate directions of visual- and/or sound-image movements as a time sequence (i.e., ←: Left to right, →: Right to left, ↑: Low to high, and ↓: High to low, °: drawing apart, ⊙: drawing close, and φ: no movement). A, V, and AV mean audio, visual, and audio-visual, respectively. Mark by each item indicates the statistical significance of change by adding the corresponding moving picture in 1-s intervals: ▲ – statistically significant increase, ▲ – marginally significant increase (p < 0.1), (no mark) - no significant change, ▼ – marginally significant decrease (p < 0.1), and ▼ – statistically significant decrease. Significance levels are denoted with marks of * (p < 0.05) and ** (p < 0.01). Underline indicates the results are shown in Fig. 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>▲ Fallen leaves on a playground [A (wind roar); ←; AV (two-person flirtation); →]</td>
<td>V First-person view facing an office desk</td>
</tr>
<tr>
<td>▲ Fireworks (large quantity) [AV: ↑; ↑; ↑; ↑ (at least 30); A (talk, claxon); φ]</td>
<td>[A (footsteps, talk); ←; →]</td>
</tr>
<tr>
<td>▲ Fireworks (small quantity) [AV: ↑; ↑; ↑; ↑; ↑; A (talk); φ]</td>
<td>▲ Loudspeakers reproducing music [A: →; →; →; →; →; →; → (7 reciprocal panning)]</td>
</tr>
<tr>
<td>▲ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>V First-person view from a moving car</td>
</tr>
<tr>
<td>▲ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>[A (falling from a height, a stimulus was reproduced and the volume of the headphone was increased); →]</td>
</tr>
<tr>
<td>▲△ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>▲ Passing roller coaster B (rear) [A: →; →; →; →; →; → (8 reciprocal panning)]</td>
</tr>
<tr>
<td>▲△ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>A (announcement, background music) φ</td>
</tr>
<tr>
<td>▲△ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>Scene of playing Japanese archery [A (flying arrow)]; ←</td>
</tr>
<tr>
<td>▲△ Passing roller coaster A (front) [AV: ↓; ↓; ↓; ↓; A (background music); φ]</td>
<td>View of a university building [A (car, motor-bike); φ; φ; φ; A (tweet); φ]</td>
</tr>
<tr>
<td>**△ Scene of a basketball game [AV (bouncing ball, footsteps); →; ←; A (cheer); φ]</td>
<td>Activities in a park [A (babble of voices of about 50 children); φ]</td>
</tr>
<tr>
<td>** Scene of an indoor batting cage [V (ball); →; ←; ←; ←; ←; ←; ←; ←]</td>
<td>Babbling brook</td>
</tr>
<tr>
<td>** Scene of a basketball game [AV (bouncing ball, footsteps); →; ←; A (cheer); φ]</td>
<td>Brass band concert</td>
</tr>
<tr>
<td>** Scene of an indoor batting cage [V (ball); →; ←; ←; ←; ←; ←; ←; ←]</td>
<td>Choral singing</td>
</tr>
<tr>
<td>** Scene of a basketball game [AV (bouncing ball, footsteps); →; ←; A (cheer); φ]</td>
<td>Fountain at night (lighted)</td>
</tr>
<tr>
<td>** Scene of an indoor batting cage [V (ball); →; ←; ←; ←; ←; ←; ←; ←]</td>
<td>Night scene with chirping insects</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Quiet corridor in a building</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Quiet forest</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Raining night scene of lakeside</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Ringing mobile phone</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Scene of cooking</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Scene of keyboard typing</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Scene of open-air food stalls</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>Trees in breeze</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>View of a merry-go-round</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>View of a waterfall</td>
</tr>
<tr>
<td>▲△ First-person view from a moving car [A (engine noise, tire noise); φ]</td>
<td>View of lakeside</td>
</tr>
</tbody>
</table>

clarifies the situation for four sample items. If a content item contains several sound sources, whether the source moves or is stationary is determined by the movement of the adopted title (e.g., the item passing roller coaster A (front) contains a moving coaster and stationary background music as shown in Table 2).

These AV content items were recorded using a HD (high-definition) video camera (Sony, XDCAM PMW-EX1R). Moving pictures were recorded in the full-HD format (1920 × 1080/60i). For audio signals, the binaural technique [12] with a dummy head (Koken, SAMRAI) was used to accurately record the original sound field. Two microphones (Aco, Type 7013) with the C-type couplers were mounted in both ears of the dummy head. The output signals from the microphones were amplified with measuring amplifiers (Ono Sokki, LA-6100) and recorded as the audio tracks using the video camera. Consequently, the audio and visual signals were synchronized. The audio signals 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 item was reproduced in an office room as a stimulus in the experiments. Video stimuli were presented with a 65-inch full-HD display (Sharp, AQUOS LC-65GX5), while auditory stimuli were reproduced through headphones (Sennheiser, HD650). 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 [13]. Prior to the reproduction, the inverse characteristics of the headphone transfer function, which is the transfer function from the input terminal of a headphone to the output of the microphone placed in the outer ear of the dummy head, was convolved to the recorded signals [12]. The reproduced sound pressure level (SPL) was calibrated via the following procedure. First, the headphones were worn on the dummy head. Second, a stimulus was reproduced and the volume of the headphone amplifier (Yamaha, AX-9) was adjusted so that the
measuring amplifier indicated the same SPL as the original sound field.

2.2 Procedures

Experiments were conducted under two conditions: audio-only and audio-visual. In the audio-only condition, only sound stimuli were presented, while both sound and the corresponding moving pictures were presented in the audio-visual condition. First the audio-only experiment was conducted, and then the audio-visual experiment was conducted at least one-month later. Prior to the experiments, the sense of presence was defined as “a feeling that you are actually in the situation/location”. Twenty undergraduate students (Female: 7, Male: 13) participated in the experiments.

In the experiments, an item was randomly presented to the subject, and the subject responded instantaneously to the sense of presence using the method of continuous judgment by category [14], [15]. Namely, the subject pressed one of seven keys (1–7) on a computer keyboard, where 1 indicates “not feeling presence at all” and 7 indicates “large feeling of presence”. The pressed key was recorded with a sampling frequency of 10 Hz. The subjects were instructed not to press any key until he or she could evaluate the presence after the stimulus began.

After each item was presented, the subject was asked to indicate the overall presence using a seven-point Likert scale for the term “I felt the sense of presence”, where 1 and 7 represent “strongly disagree” and “strongly agree”, respectively.

3. Results and Discussion

3.1 Variation in Instantaneous Presence

Figure 1 shows four examples of the experimental results (one example from each group, which is underlined in Table 1). Dotted and solid lines represent the average scores of all subjects’ responses to the audio-only and audio-visual conditions, respectively. Because the latency, which is the time interval between the start of stimulus presentation and the start of response, differs among the subjects, the averages were calculated after half (ten) subjects began their responses. Although there is a fine fluctuation in the averages, individual data shows that most of the responses lasted at least a few seconds.

Table 2 explains the four items from the viewpoints of the dummy head position, time-course of the event, and movements of sound images. Figure 2 shows the time-courses of the loudness for these items estimated by Zwicker’s method[16] in 500-ms intervals. Comparing Figs. 1 (a) and 2 as well as the descriptions in Table 2, an increase in loudness is obvious when the roller coaster passes by the observer. In our previous study [17], a psychoacoustical experiment investigated the movements of the sound images using 10 other subjects, in which two examples of a passing roller coaster A (front) and a view of a waterfall were examined. Figure 3 demonstrates the movements of the sound images in the azimuth for these two examples, where ~90 and 90 degrees mean left and right, respectively. Movements of the sound image of the coaster are clear for a passing roller coaster A (front), while the sound image is stationary for a view of a waterfall. Because the left-right movement is perceived robustly in binaurally reproduced sound, left-right movements are shown as arrows in Fig. 1. In a similar manner, schematic diagrams of the movements are presented in Fig. 1 (right panels) for the four items. In Fig. 1 (a), the visual movements of the coaster are denoted as numbers corresponding to the track in the still picture.

Comparing these figures and the table reveals that the sense of presence significantly increases when movements in the visual and/or sound images are large (e.g., 20 s or later in Fig. 1 (a) and 5 s or later in Fig. 1 (b)) and when the loudness increases (e.g., around 5 s in Fig. 1 (c)). These increases agree well with the property shown in the previous report on music pieces [11]; the sense of presence is evaluated higher when there is movement in a sound image and the dynamic range of reproduced sounds is large. On the other hand, the sense of presence is steady when the visual and audio images are stable (Fig. 1 (d)).
(a) *Passing roller coaster A (front)* in the group of (Visual image: moving, Sound image: moving). Still picture is the scene around 21 s.

(b) *Passing roller coaster A (rear)* in the group of (Visual image: stationary, Sound image: moving).

(c) *First-person view on a playground slide* in the group of (Visual image: moving, Sound image: stationary). Still picture is the scene around 8 s.

(d) *View of a waterfall* in the group of (Visual image: stationary, Sound image: stationary)

**Fig. 1** Examples of experimental results. Left and right panels show a still picture and instantaneous presence values, respectively. Error bars indicate ±1 standard errors calculated every 1 s. Arrows show schematic diagrams of the movements of visual (V) and sound (A) images where the direction and length of arrow denote the direction and the period of movement, respectively.
image movement, our previous studies indicated that besides these two features, some features of sound and a moving picture affect the overall presence [8], [10]. To construct a presence meter, further examinations on the relationships between the instantaneous presence and other features are necessary.

### 3.2 Effects of Addition of Moving Pictures on Instantaneous Presence

To examine the effects of adding moving pictures on the instantaneous presence, differences in the average scores between the audio-only and audio-visual conditions were calculated at one-second intervals. Because the subjects’ responses were recorded with a 10-Hz sampling frequency, the 10 most recent scores were averaged and subtracted to calculate the one-second difference. This data processing was carried out to diminish the effect of noise in the observed data, such as missing data or flashing errors in the transition of pressing keys. The maximum (absolute value) among the calculated differences was adopted as a result for each content item. The statistical significance of the difference was examined by the Wilcoxon signed-rank test because the raw data were categorical. Table 1 shows the results where the following marks indicate the change upon adding moving pictures: ▲ – statistically significant increase, ▼ – marginally significant increase ($p < 0.1$), (no mark) – no significant change, ▽ – marginally significant decrease ($p < 0.01$), and ▼ – statistically significant decrease. Significance levels are denoted with marks of * ($p < 0.05$) and ** ($p < 0.01$).

For the four content items in the group of [sound image: stationary, visual image: moving], the sense of presence increases more than 0.5 points upon adding moving pictures, and two items are statistically significant (Table 1). In these items, the sound sources move with the observer because the sound sources are the engine and tires while riding a first-person view from a moving car, the legs and hip of the observer in a first-person view on a playground slide, an announcement in the gondola in a first-person view from a ropeway gondola, and the footsteps of the observer in a first-person view walking down a quiet corridor. It is hypothesized that the subjects do not feel movement in the audio-only condition because the sound source moves with the observer, but the addition of the corresponding picture allows the subjects to understand movement. Thus, adding a picture allows the subjects to feel a higher sense of presence.

As for the six items in the group of [sound image: moving, visual image: stationary], the addition of moving pictures decreases the sense of presence more than 0.5 points, and two are statistically significant (Table 1). An example is a passing roller coaster A (rear) in which the roller coaster is moving to the rear of the observer, but the subject cannot see movement (Fig. 1 (b)). Although the audio signals allow the subjects to understand movement, the inconsistency between the audio and visual information decreases the sense of presence.

For items in the groups of [sound image: moving, visual image: moving] and [sound image: stationary, visual image: stationary], the effects of adding moving pictures depend on the item. Thus, examining the increase or decrease in presence upon adding moving pictures is not fruitful. An alternate explanation is that the audio-visual presence score is similar to the higher value of the audio-only or visual-only presence, which was shown in our previous studies [8], [9]. To examine this explanation, Fig. 4, which is similar to the figures in the previous studies but is drawn using current data, shows the overall sense of presence of each content item. The previous studies [8], [9] involved experiments to evaluate the overall sense of presence in three conditions (audio-only, visual-only, and audio-visual) using the same content items but using 21 other subjects. The correlation coefficients between the previous and current results are 0.92 and 0.95 for the audio-only and audio-visual conditions, respectively. Because the results herein are consistent with previous studies, we assumed that the data for the visual-only condition can be compared to the current results (Fig. 4). The rule that the audio-visual presence score is similar to the higher value of the audio-only or visual-only presence also holds in this study.

First, the relation of the differences in the one-second interval presence and the overall presence is discussed between the audio-only and audio-visual conditions. In Fig. 4, the statistical significances of the difference in the overall presence upon adding moving pictures are indicated as
Fig. 4 Average of the overall presence score. Items are arranged in descending order of the audio-visual presence. Data of the visual-only condition are taken from our previous reports [8], [9]. Open and filled arrows denote that the differences between the audio-visual and audio-only presence are statistically significant ($p < 0.05$) and marginally significant ($p < 0.1$) by the Wilcoxon signed-rank test, respectively. Open and filled triangles are from Table 1. Underline indicates the results are shown in Fig. 1.

Next, the relation between the triangle directions and the differences in the overall presence is considered for the three audio-visual conditions. If a triangle in Fig. 4 is in the upward direction, the visual-only presence is higher than the audio-only presence. A typical result is a view of a waterfall in which the visual-only presence score is about 1.3 points higher than the audio-only presence, and the audio-visual presence is close to the visual-only presence. Although there are two exceptions, fireworks (large quantity) and passing rowboat, these differences between the audio-only and visual-only presence are not statistically significant by Mann-Whitney’s U-test. This might suggest a possibility that, if the difference of presence between audio-only and visual-only conditions is small, a cross-modal effect occurs to enhance audio-visual presence. In general, the addition of a moving picture, which has a relatively high visual-only presence or almost the same presence as audio-only presence, results in a change indicated by an upward triangle.

On the other hand, if the triangle is in the downward direction, the visual-only presence is lower than the audio-only presence. For example, for a passing roller coaster A (rear), the visual-only presence is about 2.8 points lower than the audio-only presence, and the audio-visual presence is slightly lower than the audio-only presence.

Based on the aforementioned discussion, we conclude that the change in presence upon the addition of moving picture is due to the fact that the audio-visual presence does not exceed the audio-only or visual-only presence; the audio-visual presence score is similar to the higher value of the audio-only or visual-only presence.

Incidentally, in the audio-only condition, the subjects were not provided a prior explanation on the sound sources or situation because preliminary information about the situation (e.g., a moving picture of the situation) affects the presence perception [18]. Besides, the first author previously examined the effect of a prior description on the perception of sound and observed an interesting effect; without a description, the sound of a waterfall is perceived as noise, but the
same sound is perceived more comfortably with the literal information of sound of a waterfall [19], [20]. Therefore, the evaluation results in this experiment might be different if the sounds are described beforehand. In actual listening conditions (e.g., a radio program), a prior description of the content item is usually given. In the future, the effect of prior information of the content item on its presence should be examined.

3.3 Relation between Overall and Instantaneous Presence

Here the relation between overall presence and instantaneous presence (not a one-second interval average but the raw score) is discussed. First, the maximum value of the time-series data of instantaneous presence was derived and compared to the overall presence score. The maximum values tend to be larger than the overall values: The errors averaged over the 40 content items are 0.23 and 0.14 for the audio-only and audio-visual conditions, respectively, where the error is defined by the difference when the measured overall presence value is subtracted from the maximum value. Second, the mean values of the time-series data were compared to the overall presence. The mean values tend to be smaller than the overall values: The average errors are −0.24 and −0.25 for the audio-only and audio-visual conditions, respectively. These errors indicate that the maximum value is too large to estimate the overall value, and the mean value is too small.

Thus, the ten-percentile exceeded presence score, $S_{10}$, which is the score that is exceeded for the 10% of the time during the responses, was examined. Figure 5 shows that the relation between the overall value and $S_{10}$ agrees well.

The average errors are 0.14 and 0.03, while the correlation coefficients are 0.96 and 0.97 for the audio-only and audio-visual conditions, respectively. When the five-percentile exceeded presence score, $S_5$ is adopted, the average errors are 0.16 and 0.07, while the correlation coefficients are 0.96 and 0.97 for the audio-only and audio-visual conditions, respectively. This means that $S_5$ is also an appropriate measure to estimate the overall presence. In the future, the best percentile should be considered with many items.

A previous study has suggested that the overall loudness is not the average of instantaneous loudness at every moment, but that it is based on the instantaneous loudness above a certain level [14], [15]. Moreover, the German standard for loudness of time-varying sounds adopted the five-percentile exceeded value of loudness data calculated in 2 ms intervals [21]. These findings suggest that the relationship between overall presence and instantaneous presence is similar to the loudness evaluation.

Incidentally, we used “not feeling presence at all (1) – large feeling of presence (7)” to measure the instantaneous presence because this type of scale is common in experiments using the method of continuous judgment by category [14], [15]. Moreover, because a previous work by Oode et al. [11] also used this type of scale, we can compare the results. On the other hand, we used the Likert-type scale of “strongly disagree (1) – strongly agree (7)” to measure overall presence because we wanted to compare the current results to those in our previous studies, which used a Likert-type scale. We assumed that the data are comparable between these two types of scale. However, the equality of the categories’ distance should be verified in a future work.

![Fig. 5](attachment:image.png)

(a) Audio-only condition. (b) Audio-visual condition.

**Fig. 5** Relation between the overall presence and 10-percentile exceeded presence score, $S_{10}$, for 40 content items.
3.4 Relation between Latency and Presence

Latency depends on the content item. To examine the relation between latency and instantaneous presence, the median value of the latencies among the subjects was derived for every content item. Then the presence score at the beginning of response was derived and averaged among the subjects for every item. Figure 6 shows that latency and the presence score are negatively correlated, indicating that when the sense of presence is low, the latency tends to increase.

Figure 6 also suggests that latencies are longer for the audio-only condition as a whole, which may be due to difficulty imagining the situation in the audio-only condition. For the audio-only condition, content items with latencies of more than 4 s are view of lakeside (5.2 s), first-person view walking down a quiet corridor (5.1 s), quiet corridor in a building (4.8 s), ringing mobile phone (4.7 s), and first-person view facing an office desk (4.6 s). On the other hand, only the quiet corridor in a building (4.1 s) has a latency beyond 4 s for the audio-visual condition. Latency should be considered when developing a presence meter because the meter should reflect human responses.

4. Conclusion

In this study, experiments to evaluate the instantaneous presence of 40 AV content items were conducted. At the movements of sound and visual images as well as loudness increase, the sense of presence becomes higher, which is consistent with previous studies on the overall presence.

For items with different presence scores between the audio-only and audio-visual conditions, a simple rule can explain the difference; the audio-visual presence does not exceed the higher value of the audio-only or visual-only presence. This rule is confirmed under the condition of dummy-head recording. Because this recording method is not the ordinary technique to produce television (TV) programs, further studies are needed in which actual TV content items are used as stimuli.

Additionally, the overall presence can be estimated with ten-percentile exceeded presence score of the instantaneous presence data, $S_{10}$.

The ultimate goal of our research is to develop a presence meter. We have developed neural-network-based model to estimate the overall presence using physical features of AV signals [8], [10], which will be expanded based on the instantaneous presence data obtained in this study to realize a real-time operating meter. Moreover, although this study focuses on content presence, basic research to clarify the property of system presence is necessary to develop a presence meter.

Acknowledgment

This study was funded by the National Institute of Information and Communication Technology (NICT), Japan. The authors thank Mr. Kazutomo Fukue at the University of Yamanashi for his assistance in collecting the AV content items.

References


---

Kenji Ozawa received the B.E., M.E., and Ph.D. degrees in Electrical Communication from Tohoku University in 1986, 1988, and 1994, respectively. He is currently a professor of the University of Yamanashi. His research interests include psychoacoustics, signal processing of audio signals, and Kansei (Emotion and Sensibility) information processing.

Shota Tsukahara received the B.E. degree in Computer and Media Engineering from the University of Yamanashi in 2013. He is currently with a private company. He engaged a research on the Kansei evaluation and modelling of the sense of presence when he was with the University of Yamanashi.

Yuichiro Kinoshita received the B.Eng. and Ph.D. degrees in computer science from Ritsumeikan University in 2002 and 2007, respectively. He is currently an associate professor in the Department of Computer Science and Engineering, University of Yamanashi. His research interests include Kansei modelling, Kansei information processing and human-computer interaction.

Masanori Morise received the Ph.D. degree in engineering from Wakayama University in 2008. He was a JSPS Research Fellow (DC1) in 2006–2008, a postdoctoral researcher at Kwansei Gakuin University in 2008–2009, and an Assistant Professor at Ritsumeikan University in 2009–2013. He is currently a Project Assistant Professor at University of Yamanashi. His research interests include speech analysis/synthesis and speech perception.