Effects of AI-Generated Motion Pictograms on Comprehension and User Experience

Natsumi Okatani1[1†], Ryuji Shioya2[2††] and Yasushi Nakabayashi[3††],

Toyo University, Kawagoe, Saitama, Japan

Abstract
This study explored the potential of artificial intelligence (AI) to automate the creation of dynamic motion pictograms from static pictograms, thereby enhancing visual communication. Conventional static pictograms typically face limitations in conveying complex information. Conversely, motion pictograms can facilitate more intuitive communication through dynamic visual representations. However, the prohibitive costs required for creation. This study aimed to address these challenges and promote the dissemination of motion pictograms by automating their generation using AI. To evaluate the effectiveness of AI-generated motion pictograms, we conducted experiments comparing them with human-created motion and static pictograms. The results indicated that although AI-generated motion pictograms offer promising possibilities for visual communication, further research is required to develop more sophisticated algorithms and carefully consider design elements and motion characteristics that influence viewer interpretation. This study contributes to the literature by providing valuable insights into the potential and limitations of AI-generated motion pictograms for effective visual communication.

Keywords:
AI; pictogram; motion graphics; user experiment; visual communication..

1. Introduction

1.1 Background

1.1.1 Communicating Information through Pictograms

Pictograms are graphic symbols that convey information visually without the requirement of text information and provide visual recognition of a specific meaning [1]. Most pictograms do not require textual information; thus, they are used in numerous places, such as airports, train stations, and other public transportation systems and facilities where people from diverse countries use them to guide and warn people.

Pictograms can convey information regarding instructions, regulations, enforcement, warnings, and prohibitions. They are created to convey meaning more intuitively and quickly than words and should be simple and clear.

Pictograms can be used to warn of danger and convey information even when the user is a non-native speaker, when it is difficult to understand written information due to literacy or educational disparities, or when the user has an eyesight problem, such as the elderly. However, the number of well-known pictograms is very small, and it takes significant time for new pictograms to spread and maximize their effectiveness. In addition, due to limitations in the expression of still images, some pictograms take time to be understood, or their meanings may be misunderstood. If the meaning is familiar or known in advance, it can be understood quickly; however, if there are differences in culture or expression, it is even more difficult to understand, and in some situations, still images are insufficient to represent complex content. Thus, misunderstandings and confusion may occur, which may cause danger [2].

In recent years, "motion pictograms," which use motion to extend information, have attracted attention. Compared to still pictograms, motion pictograms can [i]express information more dynamically and convey complex information more intuitively.

1.1.2 Content Generation by Artificial Intelligence (AI)

The use of AI to automatically generate video content from still images is being actively introduced in various fields, including education, entertainment, and advertising. For example, attempts are underway to reduce the labor involved in conventional video production, such as shooting and editing, by importing photos and automatically generating video, or importing images of characters and automatically applying movement to them, thereby significantly increasing efficiency, lowering costs, and expanding the possibilities for expression. In particular, because the cost of creating still images is lower than that of

Manuscript received May 5, 2025
Manuscript revised May 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.5.20

video creation, using AI to generate video from still images enables the creation of high-quality content at low cost, and the advantages of communicating information through video content can be used in situations where there have been limitations on its use in the past.

However, some issues are associated with AI recognition, judgment, and automatic generation, such as the uncertainty of the generated results, which may return results that are not as expected. For example, it may result in physically unnatural movements or movements that differ from the desired movements. In the entertainment field, in many cases, the final product is evaluated and corrected manually before it is released as a completed product. Therefore, some uncertainty can still be allowed at the time of output by AI. However, when AI is used in fields where safety is required, such as transportation and medical care, uncertainty must be reduced as much as possible. In particular, high reliability and safety are required when the output by AI is used by users without human checks, such as in automated driving, because malfunctions can result in loss of human life.

In addition, products produced by AI may have unintended biases. For example, social biases inherent in training data can be reflected, which results in unintended biases. In the case of image generation, there is a concern that images representing a particular occupation or role may be generated with a bias toward a particular gender or race. This bias may fix social stereotypes and promote discrimination and prejudice; thus, multifaceted measures, including ethical aspects, are required to utilize the generated results more fairly and appropriately.

1.2 Motivation

The purpose of this study is to promote effective and efficient communication of information by recognizing still pictograms using AI and using the generated motion pictograms. The automatic generation of motion pictograms by AI enables low-cost and quick creation of effective motion pictograms that take advantage of video expression for still pictograms in various types. In addition to enabling more intuitive and easy-to-understand information communication, motion pictograms automate the process of creating motion pictograms, which require specialized knowledge and time, such as the use of video software and creators, thereby promoting their widespread use in a wide range of situations.

However, some issues are associated with the quality and reliability of AI-generated motion pictograms. Pictograms are important for safety assurance, for example, they indicate actions to be taken in an emergency or warn of danger. Therefore, the accuracy of the generated results is extremely important for AI-based motion pictogram generation. Although the use of AI offers advantages such as increased efficiency and wide dissemination, the reliability of the generated results must be ensured. Pictograms can lead to confusion and erroneous behavior due to semantic interpretations that differ from the intended content. Therefore, to minimize risk while maximizing the benefits of AI generation, it is necessary to ensure that the quality of the generated pictograms accurately conveys the intended information. Compared to static images or motion pictograms created manually, it is necessary to evaluate and verify whether the AI-generated results can convey information more accurately, appropriately, and effectively.

In an experiment in a previous study [3], conducted a comparative verification of motion pictograms manually created by humans with conventional still and motion pictograms and demonstrated their effects and trends. Based on these results, the present study aims to clarify whether AI-generated motion pictograms can demonstrate the same or different effects and potentials compared to still pictograms and human-created motion pictograms.

1.3 Related Research

1.3.1 Video Generation by AI

In recent years, several AI-based video generation methods have been reported, including text-to-video (T2V), image-to-video (I2V), and video-to-video (V2V). I2V and T2I are two of the most studied fields.

I2V generates videos from still images. Although I2V can capture the features of input images and generate videos based on these features, it is difficult to control the content of the generated videos in detail.

T2V generates videos from text descriptions. For example, large models such as Soracan generate high-resolution videos with realistic motion and light effects. Text descriptions can generate videos of various styles, and long videos containing multiple scenes can be generated from a single text [4].

Conventionally, I2V and T2V primarily generate video from a single source of information, i.e., images or text, respectively. However, with the recent development of multimodal fusion technology in the field of AI, text-guided I2V (TI2V), which introduces text-based conditioning, has attracted significant attention. TI2V enables more flexible and creative video generation than I2V and T2V [5] [6].

1.3.2 Use of TI2V (Text-guided Image-to-Video)

The purpose of this study is to automatically generate and present video images as motion pictograms based on still pictograms recognized by users on their devices. Therefore, the system must be able to generate motion pictograms that convey the meaning of the target still pictogram even if the pictogram is unknown to the system. In addition, to convey the meaning of the pictogram more precisely, the advantages of text-based conditioning of TI2V are exploited.

However, TI2V requires as input a detailed textual description of the content of the image to be generated, in addition to the image to be visualized. Therefore, in this study, we assume that AI automatically estimates the meaning of a still pictogram, which is an input image, converts it into text, and uses it as input for motion pictogram generation. Therefore, the user does not need to describe the text, and the purpose of this study, which is to automatically generate motion pictograms from unknown pictograms, can be realized.

1.3.3 Still-to-Motion Pictogram Conversion

Regarding the conversion of still pictograms into motion pictograms, we conducted an experiment to compare the effectiveness of using motion pictograms and their automatic generation with the recognition of still pictograms and manually created motion pictograms. In addition, the automatic generation of motion pictograms using AI was investigated. The trends and requirements for the effectiveness of motion pictograms were clarified [3]. Several motion pictograms were more effective in improving comprehension than still pictograms and were significantly more effective for still pictograms that express actions and complex meanings and those for which the original still pictograms were not well known.

The generation of augmented reality (AR) content using recognition results from AI and image recognition has been previously used in the field of education. It was applied to a system that supports the learning of spatial graphics by recognizing spatial graphics on paper, generating a three-dimensional model based on the results, and allowing the user to use it as AR [7]. Augmenting information using AR can improve understanding. In this study, the automatic generation of motion pictograms from still pictograms and the resulting AR will be useful as an aid to intuitive understanding and action.

2. Methodology

2.1 Verification Experiment of AI-generated Motion Pictograms

To utilize AI-generated motion pictograms, it is necessary to verify whether the generated results can accurately convey information and the degree to which the information is conveyed and understood compared to still and human-generated motion pictograms. In this study, three types of pictograms were compared: still pictograms, human-created motion pictograms, and AI-generated motion pictograms. Experiments were conducted to investigate and verify the differences in the ease of communicating and understanding the meanings represented by these pictograms and the participants' subjective evaluation of the pictograms. Through these experiments, we clarify the characteristics, effects, trends, and limitations of each pictogram type and evaluate the effectiveness and limitations of AI-generated motion pictograms.

2.2 Structure of Experiments

In the experiments, the participants were asked about the meaning of three types of pictograms, namely, still pictograms, AI-generated motion pictograms, and human-created motion pictograms. The percentage of correct answers and changes in subjective perception were then compared.

2.2.1 Participants

The experiment was conducted with 129 people aged between 10 and 50 years. The participants were randomly divided into two groups, Groups A and B, to avoid bias.

2.2.2 Preparation Phase

Five pictograms were used in the experiment. Pictograms were selected from those used in public facilities and transportation systems in different regions, such as Asia and Europe, while considering their different meanings and characteristics. The pictograms used in the experiment and their meanings are as follows.

A: Caution drop
B: Connecting flights
C: Slope
D: Meeting point
E: Elevator

For each pictogram, these three types of pictograms were prepared. Details about them are shown in Table 1, including still image pictograms and prompts for AI-generated motion pictograms.

Table 1 Motion Pictograms used in Experiments

	Pictograms (Original, still)	Text entered at the prompt for generation
A	Caution drop 	man is falling down
B	Connecting flights 	he is walking from the plane on the left to the plane on the right.
C	Slope 	Going up a slope in a wheelchair
D	Meeting point 	The arrows gather in the center to form one black circle. The black circle in the center then blinks and expands and contracts repeatedly to emphasize the black circle.
E	Elevator 	The elevator is rising.

Table 2 Motion pictograms (A: Caution drop, B: Connecting flights)

		Motion-pictogram
A Caution drop	Created by manually	
	Generated by AI	
B Connecting flights	Created by manually	
	Generated by AI	

- Still pictograms
- Motion pictograms created manually by humans: These animations were created using Adobe After Effects based on still pictograms.
- AI-generated motion pictograms: Image data of still pictograms and text data about the meaning and movement of still pictograms are input to a generative AI model (Gen-2: The Next Step Forward for Generative AI [8]).

Motions, both manually created and AI-generated, ranged from 4 to 8 s each. The motion started from a state that matched the still pictograms and was presented as a looped video. As an example, Table 2 shows the motion pictograms created based on the still pictograms of A and B, which are divided into frames and lined up.

2.3 Experimental Procedure

The experiment was conducted using a web page designed to present images and videos, along with a free-text input field. The participants accessed this page using their smartphones. The experiment was conducted in a group setting, with all participants facing each other in a quiet room. However, to limit direct interaction, the participants were prohibited from conversing with each other and were required to submit their responses individually via free-text input on their smartphones.

The subjects in each group were given the task of answering the meaning represented by three types of pictograms (still pictograms, AI-generated motion pictograms, and human-created motion pictograms) in a free-writing format in the order described below. Before the start of the experiment, the participants were told how to answer the questions, how the experiment would proceed, and that the word "pictogram" and the two types of pictograms and still pictograms would be compared, and an example question and answer were presented at the beginning of the question. However, this was performed without informing the participants of the differences and intentions between the two types of pictograms. In addition, the participants were asked to respond intuitively without overthinking.

【question order】

Group A: Q1 still pictogram → Q2 AI-generated motion pictogram → Q3 human-generated motion pictogram

Group B: Q1 still pictogram → Q2 human-generated motion pictogram → Q3 AI-generated motion pictogram

In each task, the five pictograms were presented in sequence and the participants were asked to describe the meaning of each. The participants were asked to rate each pictogram on a 5-point scale to assess how their understanding of each type of pictogram and the motion pictograms in Q2 and Q3 changed compared to the still pictograms in Q1.

【Questions regarding changes in understanding in Q2 and Q3】

Question: Compared with the previous page, what is the change due to the addition of movement?

2: Much easier to understand
1: Easier to understand
0: No change
−1: More difficult to understand
−2: Much more difficult to understand

The format was such that after answering Q1, the respondents were transferred to the Q2 page, then to the Q3 page, and so on, in the following order: after answering a question, the respondents were transferred to the page with the next type of pictogram question, and so on. Although the time for presenting pictograms for each question was not set, approximately 20 min were given from the beginning of answering the question until the end of answering the questionnaire.

After the experiment, all participants were asked to complete a questionnaire using the following procedure.

- Present the correct answers to the meanings of the five pictograms used in the experiment.
- Explain that the three pictograms are still, human-created, and AI-created.
- Subjective evaluation of each pictogram format, comprehension, and opinions regarding comparisons among the formats in a free-text, unlimited-character response format.

By comparing the percentage of correct answers in Groups A and B, the effect of the order of pictogram presentation on comprehension was verified. In particular, by comparing the percentage of correct answers in Q1 and Q2, the effects of changing from still pictograms to motion pictograms on the degree of comprehension were compared according to the type of pictogram (AI-generated or human-generated). The results of Q3 were not directly used for this comparison; however, the results could be used as a reference for an overall evaluation of each pictogram format by having the participants experience the three types of pictograms (still image, AI-generated, and human-generated) and then answering the questionnaire.

3. Experimental Results

3.1 Quantitative Analysis: Average Percentage of Correct Answers and Comparison of 5-point Scale Results

The average percentage of correct responses in each task for each group was calculated, and comparative analysis was performed using the chi-square test. In addition, the mean of the 5-point rating for motion pictograms was compared to examine differences in subjective ratings for each format. In particular, we analyzed the impact of motion by comparing the results of Q1 (still pictograms) and Q2 (motion pictograms). In Q2, Group A was performed using AI-generated images, and Group B was performed using manually generated images. In this way, differences in subjects' comprehension between the AI-generated and manually created pictograms were analyzed. The results of Groups A and B are presented in Tables 3 and 4, respectively.

For each pictogram, the following conditions were set to be considered correct answers. The percentage of correct answers was calculated based on these conditions; however, the answers were written; thus, there may have been a slight error in the percentage of correct answers depending on the scorer and the scoring conditions.

A: If the word "fall" or "caution" or words indicating that such a word is included in the answer.
B: If the answer includes a reference to "transferring planes."
C: If either "slope" or "wheelchair-accessible" is mentioned.
D: When "meeting place" or "meeting someone" is mentioned
E: When "elevator" is mentioned

Table 4 Experimental results of pictogram comprehension in Group A.

Group A	A	B	C	D	E
Pictograms	Caution, drop	Connecting flights	Slope	Meeting point	Elevator
Average Correct Percentage					
x_1 :Q1 Still	61.9%	3.2%	90.5%	14.3%	50.8%
x_2 :Q2 AI Motion	66.7%	4.8%	90.5%	12.7%	71.4%
x_3 :Q3 Manually Motion	68.3%	27.0%	92.1%	30.2%	74.6%
$\mathbf{\Delta x = x_2 - x_1}$ Increase of correct answers (Still→AI Motion)	+4.8%	+1.6%	±0.0%	-1.6%	+20.6%
Chi-Square test					
$\chi^2(1)$	0.311	0.208	0.000	0.068	5.644
p	0.577	0.648	1.000	0.794	†0.093
		*** p<.001, ** p<.01, * p<.05, †p<0.1			
Sensory comparison between still and moving pictograms (min:-2 to max:2)					
evaluation	**percentage by option**				
2 Much easier to understand	0.0%	6.5%	9.8%	6.6%	1.6%
1 Easier to understand	9.7%	30.6%	26.2%	21.3%	6.5%
0 No change	33.9%	37.1%	44.3%	47.5%	17.7%
−1 More difficult to understand	32.3%	16.1%	13.1%	21.3%	46.8%
−2 Very confusing	24.2%	9.7%	6.6%	3.3%	27.4%
Average of value					
v_1 :Value (Still → AI Motion)	+0.71	−0.08	−0.20	−0.07	+0.92
v_2 :Value (AI Motion → Manually Motion)	+1.39	+1.13	+1.37	+0.48	+1.33
$\mathbf{\Delta v = v_2 - v_1}$	+0.68	+1.21	+1.57	+0.55	+0.41

Table 3 Experimental results of pictogram comprehension in Group B

Group B	A	B	C	D	E
Pictograms	Caution, drop	Connecting flights	Slope	Meeting point	Elevator
Average Correct Percentage					
x1 :Q1 Still	72.7%	3.0%	81.8%	9.1%	43.9%
x2 :Q2 Manually Motion	77.3%	28.8%	86.4%	18.2%	75.8%
x3 :Q3 Manually Motion	74.2%	27.3%	84.8%	18.2%	75.8%
$\mathbf{\Delta x = x_2 - x_1}$ Increase in the number of correct answers (Still → Manually Motion)	+4.5%	+25.8%	+4.5%	+9.1%	+31.8%
Chi-Square test					
$\chi^2(1)$	0.364	16.366	0.510	2.316	13.903
p	0.546	*** 5.223E-05	0.475	0.128	** 0.008
		*** p<.001, ** p<.01, * p<.05, †p<0.1			
Sensory comparison between still and moving pictograms (min:-2 to max:2)					
evaluation	**percentage by option**				
2 Much easier to understand	0.0%	1.6%	0.0%	14.1%	0.0%
1 Easier to understand	4.7%	9.4%	1.6%	29.7%	3.1%
0 No change	18.8%	15.6%	51.6%	26.6%	26.6%
−1 More difficult to understand	40.6%	32.8%	20.3%	21.9%	37.5%
−2 Very confusing	35.9%	40.6%	26.6%	7.8%	32.8%
Average of value					
v_1 :Value (Still → AI Motion)	+1.08	+1.02	+0.72	−0.20	+1.00
v_2 :Value (AI Motion → Manually Motion)	−0.52	−0.91	−1.00	−1.05	−0.30
$\mathbf{\Delta v = v_2 - v_1}$	−1.59	−1.92	−1.72	−0.84	−1.30

3.1.1 Percentage of Correct Answers in Group A

Comparing the percentage of correct answers for Q1 (still) and Q2 (motion pictograms), in the group using AI-generated motion pictograms for Q2 (Group A), there was no change for C "Slope" and a −1.6% decrease for D "Meeting point." However, for the other three pictograms A, B, and E, the percentage of correct answers increased in Q2. In particular, the percentage of correct answers for E "Elevator" increased by 20.6% from 50.8% in Q1 to 71.4% in Q2. The results showed a significant trend for E (E:$\chi 2(1) = 5.644$, $p = 0.093$, $\alpha = 0.05$).

3.1.2 Percentage of Correct Answers in Group B

In contrast, in the group using human-created motion pictograms in Q2 (Group B), the percentage of correct answers increased in Q2 for all five pictograms. In particular, the increase in the percentage of correct answers for B "Connecting flights" and E "Elevator" was particularly significant, with a 25.8% increase from 3.0% in Q1 (still) to 28.8% in Q2 (motion pictograms) for B "Connecting flights." "Elevators" increased by 31.8%, from 43.9% to 75.8%. As in Group A, a chi-square test of independence was performed on the association between differences in Q1 and Q2 and correct or incorrect responses to these differences. The results showed significant associations in B and E (B: $\chi 2(1) = 0.208$, $p = 5.223E\text{-}05$, $\alpha = 0.05$; E: $\chi 2(1) = 5.644$, $p = 0.008$, $\alpha = 0.05$). In particular, in E, the p-value was extremely small, and the use of motion pictograms in Q2 significantly increased the percentage of correct answers.

3.2 Qualitative Analysis: Questionnaire Results

The free-text responses to the questionnaire were analyzed to extract the participants' subjective opinions and common evaluations of each pictogram format. In particular, the participants' positive and negative responses were focused on analyzing the effects and impressions they felt from each pictogram format. Table 4 presents excerpts from the questionnaire responses.

In the feedback, 27.8% of the respondents mentioned the animated pictograms. Evaluations of the animated pictograms were generally positive, with 81.3% of the respondents stating that the pictograms were "easier to understand" or "got the message across," whereas 18.8% of the respondents gave negative evaluations such as "difficult to understand" or "didn't understand."

Table 5 Feedback (partially)

About Pictograms
Meeting Point: ・Meeting point pictograms were the most difficult to understand ・Meeting point pictograms were difficult to understand because they lacked the element of gathering people ・Meeting point pictograms in Europe were particularly confusing
Other: ・The ramp and elevator pictograms were very easy to understand in the human-created video ・Pictograms are a means of communicating across language barriers, and some pictograms are widespread and easy to understand as symbols; however, some pictograms cannot be understood using only still images.
Design Quality: ・The design of a pictogram can make a significant difference in comprehension ・A simple, easy-to-understand design is important ・It is difficult to create a design that all users can understand
About motion Pictograms
Motion Pictogram ・I thought it was so much easier to understand by having movement ・It was very easy to grasp when it was displayed as a video ・Some concepts were easier to understand with movement
Human-created Motion: ・Videos created by humans are more natural and easier to understand ・Elevator videos created by humans were more natural and easier to understand ・Human-created videos were easier to convey intent than images
AI-Generated Motion: ・AI-generated videos are difficult to understand ・AI-generated motion pictograms do not accurately represent the intent ・Videos generated by AI typically contain unnatural movements ・AI-generated videos are difficult to understand because of their unnatural movements ・Videos generated by AI were difficult to read and understand the intention

Regarding AI-generated motion pictograms, 32.3% of the respondents mentioned them, with 5.4% providing positive evaluations and 94.6% providing negative opinions, with the majority of the evaluations being negative. Conversely, manually generated motion pictograms were mentioned by 47.0% of the respondents, with 96.3% and 3.7% providing positive and negative ratings, respectively.

4. Discussion

4.1 Percentage of Correct Answers and Responses in Group A

Regarding the change in the average percentage of correct answers from Q1 (still image) to

Q2 (motion), Group A, which used AI-generated motion pictograms for Q2, exhibited the most significant increase (20.6%) for E “Elevator”; A “Caution drop” also exhibited an increase of 4.8%, whereas B “Connecting flights” exhibited a slight increase, and C “Slope” exhibited no change. D “Meeting point” exhibited a slight decrease of 1.6%.

4.1.1 A “Caution drop”

In the “Caution drop” pictogram, the answers of 12.7% of the participants changed with the pictogram changing from still to motion. Notably, the answers of 9.5% of the participants changed from incorrect to correct. These participants changed their answers from “fall” or “slip” in the still pictogram to “fall” or “fall” in the motion pictogram. This suggests that the dynamic concept of “falling” represented by the pictogram was conveyed more clearly and was better understood by the motion pictogram. This result is consistent with the finding that 32.3% of the respondents answered “easier to understand” and 24.2% answered “very easy to understand” in a five-point evaluation of the pictogram for A. The correct answer was not given due to the motion-enhanced pictogram. Conversely, the answers of 3.2% of the participants changed from correct to incorrect because of the motion pictograms. These participants, despite correctly perceiving “fall” in the still image, incorrectly interpreted the motion pictogram as “get stuck in a swamp” or “step off.” Based on the feedback analysis, the slower speed or less movement of the motion pictogram, combined with the posture of the original pictogram, may have caused this misinterpretation.

4.1.2 B “Connecting flights”

In the B “Connecting flights” pictogram, the answers of 14.3% of the participants changed with the pictogram changing from still to motion. However, the effect on the percentage of correct responses was limited, with the answers of only 1.6% of the participants changing from incorrect to correct. These participants changed answers such as “I will board a plane” to answers such as “changing planes,” indicating that the motion pictogram made it easier to convey the action of “transferring” represented by the pictogram. Conversely, the answers of the majority of the participants remained incorrect even in Q2, indicating that the effect of the motion pictogram was limited. These results are consistent with the fact that the mean value of the 5-point evaluation for the pictogram in B was −0.08, and the majority of the participants (37.1%) answered “no particular change.”

4.1.3 C “Slope”

In the pictogram of C “Slope,” the answers of 4.3% of the participants changed with the pictogram changing from still to motion. Specifically, we observed cases in which the participants changed their answers from a specific use such as “a slope for the handicapped” to an uncertain expression such as “not sure” or from an alert expression such as “Please consider the elderly and the handicapped” to an objective description such as “There is a slope.” The results of the survey were similar to those of previous surveys. However, no change was observed in the average percentage of correct answers because the numbers of changes from correct to incorrect answers and from incorrect to correct answers were equal. These results suggest that the impact of motion on pictogram comprehension is limited. The results are also supported by the fact that most participants (44.3%) responded with “no particular change” on a 5-point scale.

4.1.4 D “Meeting point”

In the D “Meeting point” pictogram, the answers of 11.1% of the participants changed with the pictogram changing from still to motion. However, although the answers of 1.6% of the participants changed from correct to incorrect, the answers of no participants changed from incorrect to correct. Some participants who correctly answered “meeting place” in the still image provided incorrect interpretations in the motion pictogram, such as “this indicates that there is a hole ahead.” This may have been because the motion pictogram emphasized the gathering of people toward the center, which made it difficult to convey the intent as a gathering place and may have caused the misunderstanding. These results are consistent with the finding that the percentage of participants who answered “no particular change” was the highest in the 5-point evaluation of the D pictogram compared to the other pictograms.

4.1.5 E “Elevator”

In the pictogram for E “Elevator,” the answers of 22.2% of the participants changed from incorrect to correct with the pictogram changing from still to motion. This is the highest percentage compared to the other pictograms, suggesting that the change to motion pictograms significantly improved comprehension. Participants who incorrectly interpreted “toilet” or “window” in the still images correctly recognized “elevator” in the motion pictograms. In addition, a significant trend was observed between the change from still pictograms to motion pictograms and the change in the percentage of correct responses (E:$\chi2(1) = 5.644$, $p = 0.093$, $\alpha = 0.05$), suggesting statistical significance. These results suggest that motion communicates the dynamic concept of “vertical motion” represented by pictograms and significantly improves participants’ comprehension.

4.2 Percentage of Correct Answers and Responses in Group B

In Q2, Group B, which used motion pictograms created by humans, showed an increase in the percentage of correct answers for all pictograms. In particular, the most significant increase was observed for E “Elevator,” which showed a 31.8% increase in the percentage of correct answers. Next, B “Connecting flights” showed a 25.8% increase, a significant improvement compared with Group A’s 1.6%; D “Meeting point” showed a 9.1% increase; A “Caution drop” and C “Slope” each showed a 4.5% increase.

4.2.1 A “Caution drop”

In the “Caution drop” pictogram, the answers of 6.2% of the participants changed from incorrect to correct, whereas the answers of 1.6% of the participants changed from correct to incorrect as the pictogram changed from a still pictogram to a motion pictogram. Participants who interpreted the static images as “fall” or “slip” arrived at the correct answer by visually perceiving the fall movement depicted by the motion. Conversely, participants whose answers changed from correct to incorrect may have correctly understood “fall” in the still image; however, the stumbling behavior before the fall movement in the motion pictogram led to an incorrect interpretation of “fall.” These results demonstrate that the motion pictogram communicated the concept more clearly and contributed to improved comprehension. However, the motion pictogram was not always correctly understood by all participants and may have deepened the misunderstanding in some participants. In particular, note that the detailed representation of motions can induce unintended interpretations.

4.2.2 B “Connecting flights”

In B “Connecting flights,” the answers of 26.2% of the participants changed from correct to incorrect, and all incorrect answers changed to correct answers. Specifically, ambiguous answers such as “airport” and “getting on a plane” were changed to “changing planes” and “Connecting flights,” suggesting that the intent represented by the pictogram was conveyed more clearly by the visual representation of the motion-based changeover behavior. A statistically highly significant association was found between this change in correctness and the changes in Q1 and Q2 (B:$\chi2(1) = 0.208$, $p = 5.223E\text{-}05$). This result suggests that motion pictograms may provide higher comprehension than still pictograms, especially when representing dynamic concepts such as “transit.” In other words, the motion pictogram may have allowed the participants to more easily understand the meaning of the pictogram and select the correct response by visually representing the dynamic concept that the pictogram represents.

4.2.3 C “Slope”

In the pictogram for C “Slope,” the responses of 6.2% of the participants changed with the pictogram changing from still to motion, with all correct answers changing to incorrect answers. Analysis of these subjects’ responses revealed that they misunderstood the still pictograms as representing a wheelchair, such as “wheelchair” and “wheelchair going up a slope,” and did not understand that the pictograms indicated a wheelchair-accessible “slope” or “hill.” It was assumed that the visual representation of the wheelchair moving down a slope by the motion pictogram helped the participants understand more clearly that the pictogram represented a “slope.”

4.2.4 D “Meeting point”

For the D “Meeting point” pictogram, the answers of 10.8% of the participants changed from incorrect to correct as the pictogram changed from still to motion, whereas the responses of 1.5% of the participants changed from correct to incorrect. Analysis of the responses of participants that changed from incorrect to correct revealed misconceptions such as “pitfall caution” and “central location.” This may be due to the difficulty in expressing the abstract concept of "meeting place" in still images. However, the visual representation of people gathering toward the center through motion clearly conveyed the intent of "meeting place" or "meeting place," and the percentage of correct answers was considered to have increased.

In contrast, the responses of participants that changed from correct to incorrect suggest that the detailed information contained in the motion pictogram led to misunderstanding. In particular, incorrect interpretations such as "a place where businessmen often gather" were observed. This may have been because a person was carrying a bag in the motion pictogram, which was associated with the image of a businessman. These results suggest that motion communicates the concepts it represents more concretely and improves comprehension.

However, we also observed that including excessive information in motion led to unintended information being communicated and caused misunderstanding. In particular, detailed information such as a person's attributes and actions included in a motion pictogram may induce associations that deviate from the main purpose of the pictogram and lead to misunderstandings. Therefore, when creating motion pictograms, it is important to focus on the information to be conveyed and eliminate unnecessary information as much as possible.

4.2.5 E "Elevator"

In the pictogram E "Elevator," the answers of 32.3% of the participants changed from incorrect to correct as the pictogram changed from still pictograms to motion pictograms, whereas the responses of 1.5% of the participants changed from correct to incorrect. This increase in the percentage of correct responses was the most significant among the five pictograms in this experiment. The results of the chi-square test for the association between the change from static to motion pictograms and the change in correct and incorrect responses showed a statistically significant association ($\alpha = 0.05$, E:$\chi 2(1) = 5.644$, $p = 0.008$, $\alpha = 0.05$). This suggests that motion contributes significantly to understanding the concept of "elevator."

Analysis of the responses of participants that changed from incorrect to correct revealed that these participants misunderstood the concept to be different from that of an elevator, such as "restroom" or "unisex." The visual representation of the elevator cage going up and down by the motion may have helped the participants understand the concept of "elevator" more accurately.

Conversely, the responses of participants that changed from correct to incorrect suggest that the detailed information contained in the motion pictogram led to misunderstanding. Specifically, in addition to the fact that the still pictograms in the motion showed male and female pictograms in the elevator basket, which could have easily been misunderstood as "toilets," when the motion pictogram demonstrated the elevator rising, the full body of the human figure, which was hidden in the still picture, became visible in the motion pictogram. The pictograms of a man and a woman were emphasized, which may have caused the viewers to strongly perceive the image of a "restroom." This suggested that the interaction between the elements in the pictograms was altered by the motion, creating a new interpretation.

4.3 Advantages and Disadvantages

These results suggest that motion pictograms can communicate dynamic concepts such as "elevator" more effectively than still pictograms. However, the detailed information contained in motion pictograms can induce unintended interpretations and deepen misunderstandings. Therefore, when creating motion pictograms, it is important to focus on the information to be communicated and eliminate unnecessary information as much as possible. In addition, when depicting a person in a pictogram, it is necessary to carefully consider how the person is to be understood and to design the pictogram in such a way that it does not cause misunderstanding.

5. Summary

In this study, experiments using five types of pictograms were conducted to compare the effectiveness of AI- and human-generated motion pictograms. The results suggest that AI-generated motion pictograms are more effective in improving comprehension than human-generated pictograms. Nevertheless, there is room for improvement.

Specifically, in the AI-generated motion pictograms, the most significant increase in the percentage of correct answers was observed for E "Elevator," suggesting that the participants understood the concept of "elevator" more accurately through motion.

Conversely, in the case of A “Caution drop,” the speed and movement of the motion were misleading in some cases. In addition, in B “Connecting flights,” C “Slope,” and D “Meeting point,” the effect of the motion was limited, and in some cases, it deepened the misunderstanding. In contrast, the human-created motion pictograms demonstrated an increase in the percentage of correct answers for all pictograms, especially for E “Elevator” and B “Connecting flights.” This may be because the human-created motion pictograms expressed more natural and intuitive motion, and the participants accurately understood the intent.

These results suggest that although AI-generated motion pictograms are not as effective as human-generated motion pictograms, they may contribute to improving comprehension of certain pictograms. However, the quality of AI-generated motion pictograms is highly dependent on the performance of the generation model and the quality of the training data, which leaves room for further improvement.

To employ AI-generated motion pictograms without human evaluation, it is essential to establish a mechanism to eliminate potential misunderstandings and misinterpretations caused by the influence of motion on pictogram understanding. Although motion can be an effective means of visually communicating dynamic concepts that are difficult to express with still images, it can also induce misunderstandings and misinterpretations, depending on how the motion is expressed. For example, in this study, a pictogram of A “Caution drop” was misinterpreted if the motion was too slow or moved too little.
There are two possible approaches to solving this problem.

The first approach involves developing an automatic checking system: it would be effective to develop a system that automatically detects potentially misleading elements for AI-generated motion pictograms. Specifically, generated motions should be analyzed based on multiple evaluation criteria to identify movements and expressions that may induce misinterpretation.

The second approach is the combined use of multiple AI models. It is conceivable to introduce a mechanism to generate multiple motions and compare them to select more appropriate motions. In addition, developing an AI model that understands the meaning of pictograms through deep learning and verifies that the generated results allow the generation of more accurate motions.

Through these efforts, we expect to improve the quality of AI-generated motion pictograms and increase the percentage of correct answers. In future research, we plan to employ AI to automatically generate optimal motion pictograms without human intervention.

The immediate recognition and visualization of still pictograms as motion pictograms, which can then be displayed as AR, will promote more intuitive and accurate understanding and facilitate information complementation. Specifically, it would enable intuitive communication of meaning independent of language and more effectively express information that also contains complex meanings. For example, even if a user cannot understand a written warning or instruction, the given information can be easily understood using motion pictograms. In the future, we plan to develop a more versatile AR platform by researching systems that can convert unknown types of pictograms into motion pictograms.

In conclusion, the use of AI-generated motion pictograms can significantly contribute to pictogram efficiency and advancement. In today’s multilingual and globalized society, it is expected to contribute to the realization of a more inclusive society by reducing barriers to communication. However, a multifaceted approach, including the development of automatic checking systems, the advancement of generative models, and the continuation of human evaluation, is essential to realize this goal.

Glossary

- Pictogram: A graphic symbol that conveys information visually without the need for text.
- Motion Pictogram: A pictogram that uses motion to convey information more dynamically and intuitively.
- AI-generated Motion Pictogram: A motion pictogram created using artificial intelligence techniques.
- Text-to-Video (T2V): An AI technique that generates videos from text descriptions.
- Image-to-Video (I2V): An AI technique that generates videos from still images.

・ Text-guided Image-to-Video (TI2V): An AI technique that generates videos from still images guided by text descriptions.

References

[1] Y. Ota. Research on Pictogram Design of International Safety Signs. 2003 Joint Research Grant Research Report. 2004.

[2] C. Tijus, et al. The design, understanding and usage of pictograms. In G. Rijlaarsdam (Series Ed.) and D. Alamargot, P. Terrier, & J.-M. Cellier (Vol. Eds.), Studies in Writing, Vol. 21, Written Documents in the Workplace (pp. 17–31). 2007.

[3] N. Okatani, et al. Development of Automatic Generating System of Motion-Pictograms from Still-Pictograms, International Journal of Computational Methods, 2024.

[4] R. Sun, et al, From Sora What We Can See: A Survey of Text-to-Video Generation, arxiv, 2024.

[5] K. Niu,. et al, Conditional Video Generation Guided by Multimodal Inputs: A Comprehensive Survey, ResearchGate, 2024.

[6] F. Fan. et al, AIGCBench: Comprehensive Evaluation of Image-to-Video Content Generated by AI, arxiv, 2024.

[7] N. Okatani, et al. Development of 3D Modeling System using AI and AR. Transactions of the Japan Society for Computational Engineering and Science, vol. 2022, p. 20220009. 2022.

[8] Runway ML Research Team. Gen-2: The Next Step. Forward for Generative AI. Manuscript in preparation. Available from: https://research.runwayml.com/gen2 [Accessed 2023].

[9] Ministry of Land, Infrastructure, Transport and Tourism. Graphical Symbols for Guidance (JIS Z8210). Japan. 2019.

[10] M. Ommiya, & M. Matsushima, Introduction to statistical analysis: Concepts of estimation, testing, and ANOVA. Morikita Shuppan Co., Ltd., Japan. 2015.

[11] O. Uchida, Ready-to-use EXCEL survey tabulation and analysis. Tokyo Shoseki Co., Ltd., Japan. 2009.

Natsumi Okatani received the M.E. degree in Information Sciences from Toyo University, Japan. Currently a Ph.D. candidate in the Graduate School of Information Sciences and Arts, Toyo University, Kawagoe, Saitama, Japan. A member of the Shioya Laboratory (Supercomputing Research Laboratory). Research interests include artificial intelligence, multimedia, image recognition, visual representation, and their applications in VR/AR, UI/UX, and interactive systems.

Ryuji Shioya holds a Doctor of Engineering degree from the University of Tokyo and is a specialist in computational science and engineering. His areas of expertise include Computational Mechanics, Computer-Aided Engineering, Parallel and Web Computing, Network and Heterogeneous Computing, Simulation Engineering, Massively Parallel Computing, the Finite Element Method, Domain Decomposition Methods, and Structural Analysis. His research spans a wide range of informatics fields, including Computational Science, Software, Human Interfaces and Interactions, and Database Science, as well as Nuclear Engineering. He has held various academic and research positions, including Visiting Researcher at the School of Engineering, Cardiff University (2023–2024), Research Fellow at the Japan Society for the Promotion of Science (1994–1996), and Visiting Researcher at the University of New South Wales (1994). He completed his graduate studies at the University of Tokyo's Graduate School of Engineering. Prof. Shioya is an active member of several academic societies, including the Information Processing Society of Japan, the Japan Society for Computational Engineering and Science, the Japan Society of Mechanical Engineers, the International Association for Computational Mechanics (IACM), and the Japan Society for Simulation Technology.

Yasushi Nakabayashi graduated from the Department of Quantum Engineering and Systems Science, Faculty of Engineering at the University of Tokyo. He then proceeded to the University of Tokyo’s Graduate School of Engineering, where he earned a Master’s degree in Systems Quantum Engineering and later a Doctorate in Information Engineering. In 1999, he was awarded a Ph.D. in Engineering. During his doctoral studies, he served as a Research Fellow of the Japan Society for the Promotion of Science (JSPS) from 1996 to 1999. Following this, he worked as a Research Associate at the Graduate School of Frontier Sciences, the University of Tokyo, from 1999 to 2002, where he conducted research in fields such as fluid engineering and computational science within a cutting-edge academic environment. In 2002, he joined Toyo University as a Lecturer in the Faculty of Engineering. In 2009, he transitioned to the Faculty of Information Sciences and Arts, where he served as Lecturer and Associate Professor before being appointed Professor in 2017.His areas of expertise span a wide range, including Computational Fluid Dynamics (CFD), Computational Mechanics Systems, and Network Computing. He has made significant contributions to the development of Computer-Aided Engineering (CAE) systems and computational science. His work is particularly recognized in the fields of fluid engineering within manufacturing technology and foundational theories in informatics.