The Effect of Audio–Video Quality on Learning Effectiveness in Distance Learning over IP Networks

Kazuyoshi Fukaya
School of Education,
Sugiyama Jogakuen University
Nagoya 464-8662, Japan

Koichiro Noda, Shuji Tasaka and Toshiro Nunome
Graduate School of Engineering,
Nagoya Institute of Technology
Nagoya 466–8555, Japan

Abstract—This paper deals with multimedia distance learning over IP networks and investigates the most suitable presentation style of contents according to communication quality by experiment. We compare three styles of contents: “text”, “text and audio” and “text, audio and video.” For the three presentation styles of contents, we examine how the learning effectiveness of the learners depends on the communication quality. The learning effectiveness is assessed in terms of the ratio of correct answer. As a result, we find that the audio–video with text raises the learning effectiveness compared to the text only if the communication quality is high. As the communication quality deteriorates, however, the audio–video degrades the learning effectiveness, which eventually becomes lower than that of the text–only case. Furthermore, we show a guideline for the selection of the best presentation style of contents for a given amount of the load traffic by means of regression analysis.

Index Terms—IP networks, distance learning, communication quality, multimedia, learning effectiveness

I. INTRODUCTION

Owing to the spread of high–speed networks and high performance terminals, many applications which rely on audio–video transmission over the Internet has become popular. Distance learning applications fall into this category.

In the distance learning, we can use a variety of media according to learning environments of learners [1]. For example, materials which combine the media such as text, audio, and video are used in Web–based applications. In the late 1990s, 84.1 % of public universities and 83.3 % of 4–year public colleges in the United States offered Web–based courses [2].

The distance learning applications can be classified into two types: real–time and on–demand [3]. The real–time type means online learning where all students “present” themselves at the same time in different places. This type commonly makes use of a TV conferencing system; a teacher and all students have to participate in the lecture at its scheduled time. The on–demand type is a mode of online delivery where students access course materials (e.g., recorded–video) on their own schedules. The students are not required to be together at the same time and can repeatedly watch the lecture. This paper deals with the on–demand distance lecture.

Many researchers have discussed the benefits of presenting information using multimedia components such as text, audio, and video [4]–[7]. In [4], Cowen shows that recall scores are generally higher for film viewers than for text readers. In [5], Mayer and Moreno describe that the student understanding is better when presenting an explanation in narrations and animations than solely in narrations in computer–based multimedia learning environments. In [6], Guan shows that spoken texts with moving pictures have a positive effect on learning efficiency. In [7], Kim and Gilman investigate the use of multimedia components such as visual text, spoken text and graphics in a Web–based self–instruction program to increase learners’ English vocabulary; the finding of the study suggests an idea that the use of visual media supports vocabulary acquisition and increases achievement scores. Note that all the studies mentioned above deal with local environments; they do not suppose distance learning.

In distance learning, teaching material is often distributed over IP networks typified by the Internet. IP networks basically provide best–effort services, and therefore the quality of communication is not guaranteed. In audio–video transmission over IP networks, packet loss and its delay are inevitable; they disturb the temporal and spatial structures of audio and those of video. The disturbance decreases fidelity of the audio–video stream [8]. Consequently, QoS (Quality of Service) of audio–video transmission cannot be guaranteed over the Internet.

References [9] and [10] show that communication quality influences learning effectiveness in the distance learning. In [9], Fujiki et al. investigate the ratio of correct answer in the test when the scrolling sentences are influenced by the network load traffic. Note that this study deals with text only. In [10], the authors have shown that text information added to video enhances learning effectiveness in audio–video transmission over IP networks. The experiment in [10] reveals that learning effectiveness with the text information is kept high even if communication quality deteriorates. Note that this study supposes audio–video transmission as the basis of the system and regards the text as an option.

The degradation of communication quality in IP networks affects the quality of media output at the receiver. However, we can find no study on the learning effectiveness of the teaching material which consists of combinations of text, audio and video under various conditions of communication quality.

This paper deals with multimedia distance learning over IP networks and investigates what is the most suitable presentation style of content of the teaching material for various conditions of communication quality.
communication quality. We compare three styles of contents: “text”, “text and audio”, and “text, audio and video.” For these three presentation styles of contents, we quantitatively measure the learning effectiveness of the learners by experiment.

The rest of the paper is organized as follows. Section II describes methods of assessing the learning effectiveness. Sections III and IV show our experimental methodology and experimental results, respectively. Finally, Section V concludes the paper.

II. ASSESSMENT OF LEARNING EFFECTIVENESS

It is important how we quantitatively assess learning effectiveness of distance learning. Two kinds of methods for assessing the learning effectiveness can be used: recall–type and recognition–type [11],[12].

- Recall–type
  In the recall–type, the learner writes down answers or fills in blanks on the basis of his/her memory. The recall–type is divided into two groups: simple–recall–type and completion–type. The simple–recall–type is a method by which the learner writes down answers, while the completion–type instructs the learner to fill in blanks in sentences, expressions and figures.

- Recognition–type
  In the recognition–type, the learner selects an answer from several alternatives prepared in advance. The recognition–type comprises two–choice, multiple–choice, and matching. Note that the learner may select the right answer even if he/she does not understand the question.

In this paper, we select a recall–type to avoid answering correctly by chance. A simple–recall–type is too difficult for the learners to give the correct answer when they are shown the teaching material only once. Therefore, we utilize a completion–type where an outline is shown in an answer sheet.

The ratio of the correct answer is used as the metric of the learning effectiveness; this is defined as the percentage of the correct answer.

III. EXPERIMENTAL METHODOLOGY

A. Experimental System

Figure 1 shows the configuration of the experimental system. It consists of two routers and four PC’s, which are used as a teaching material server, a learner’s terminal, a load generator, and a load sink. Router 1 is Cisco Systems’ 2691, and Router 2 is Cisco Systems’ 7301. The link speed between the two routers is 10 Mb/s, while the others are 100 Mb/s. We have selected these link speeds so that the link between the routers can be a bottleneck.

The components of the contents of the teaching materials used in this experiment are text, audio and video. We utilized the text as the basis of the contents. We recorded audio and video of a lecturer who read the text aloud; we utilized those audio and video as components of the contents. The presentation styles of the content in the experiment are as follows.

- Text only
- Text and audio of the lecturer
- Text and audio–video of the lecturer

In this experiment, we adopted Apache 2.2.6 [13] for the teaching material server, and HTTP 1.1 was used for the transfer of the teaching material. The learners utilized a Web browser (Internet Explorer 7) on the learner’s terminal. According to the request of the learner, the teaching material server delivered a content to the learner’s terminal.

The load generator transmitted load traffic to the load sink in order to change the quality of the content received by the learner’s terminal. The load generator generated fixed–size UDP datagrams of 1472 bytes each in its payload at exponentially distributed intervals as the load traffic. We set the average amount of the load traffic to six values: 5.5, 6.0, 6.5, 7.5, 8.0 and 8.5 Mb/s.

The learner’s terminal is equipped with headphones and a 19 inch–LCD monitor. The distance between the display and each assessor was set to that in the case where an assessor usually uses a PC (i.e., approximately 50 cm through 1 m).

B. Contents

1) The content of the teaching material: This experiment was designed so that the influence of the network quality can dominate the learning effectiveness of the learners. Therefore, we made two kinds of teaching materials which are easily understood by the learners.

This experiment has prepared the following teaching materials in Japanese.

- Teaching material A:
  Itemized simple sentences to show the location information of the facilities
- Teaching material B:
  Extract of an essay

Figure 2 shows an example of content of teaching material A. Since the actual experiment was performed in Japanese, we have translated the Japanese sentences in the teaching material
For teaching material A, we have prepared two types: four short sentences and six short sentences. For convenience, we refer to the former as “teaching material A–4”, and the latter as “teaching material A–6”.

Figure 3 shows an example of content of teaching material B whose sentences have been translated from Japanese into English by the authors. The sentences are taken from a Japanese language problem given by “National Center for University Entrance Examinations” for the National Center Test in 2007 [14].

In this experiment, the learners assess the contents of the teaching material in a variety of communication quality for the presentation styles. Then, considering the learners’ burden, we have prepared 102 contents whose sentences are different from each other; each learner assesses them. The number of contents is 36 each for teaching materials A–4 and A–6. On the other hand, we have set a smaller number of contents for teaching material B; the number of contents of teaching material B is 30 because long sentences in teaching material B need longer time for assessment.

The experiment for assessment was designed so that each content can be presented to a learner just once. This is because the learner may remember contents which have been presented before.

A completion–type question was made from a content. Therefore, 102 questions have been prepared. Each question has been printed on an A4 paper; it is presented to each learner in the experiment. All learners answered the same questions.

The question for the content in Fig. 2 and that in Fig. 3 are illustrated below.

**Teaching material A:**
- There is a department store in the ( ) of the high school.
- There is a laboratory in the ( ) of the high school.
- There is a ski resort in the ( ) of the high school.
- There is a used bookstore in the ( ) of the high school.
- There is a palace in the ( ) of the high school.
- There is a church in the ( ) of the high school.

**Teaching material B:**
- If it is ( ) of the ( ) to ( ) ( ), can we regard the ( ) style ( ) , where ( ) are ( ), as ( )?
- This seems to be produced and ( ) at least on the basis of an ( ) different from the ( ) ( ).

2) Specifications of the media: The file size of the content of the teaching materials of the text–only case is 17 through 28 kbyte. All the fonts of the text are “MS P Gothic”. The font size in teaching material A is 32 pt, and that in the teaching material B is 28 pt. Table I shows the specifications of the transmitted audio and video with text.

Table II shows the presentation time of the content for three kinds of the teaching material. Contents of teaching material A have an approximately constant presentation time. On the other hand, that of teaching material B varies. This is because all sentences in teaching material A have approximately the same length, while the length of a sentence in teaching material B

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>SPECIFICATIONS OF AUDIO AND VIDEO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Audio</td>
</tr>
<tr>
<td>coding scheme</td>
<td>LinearPCM 48kHz 16bit stereo</td>
</tr>
<tr>
<td>image size [pixel]</td>
<td>–</td>
</tr>
<tr>
<td>picture pattern</td>
<td>–</td>
</tr>
<tr>
<td>average bit rate [kb/s]</td>
<td>1536</td>
</tr>
<tr>
<td>average frame rate [frame/s]</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 2. An example of content of teaching material A

Fig. 3. An example of content of teaching material B

If it is essence of the art to produce invariable forms, can we regard the Japanese style gardens, where changes are vital, as art?

This seems to be produced and exist at least on the basis of an art idea different from the European one.
differs from sentence to sentence.

3) The learning method of the teaching material: In this experiment, the assessors learn each teaching material in the following steps.

Step 1) The assessor clicks the link of the WEB page to play the content of a teaching material. The text, audio and video of the lecturer are set to start at the same time. The text content of the teaching material appears character by character automatically. The display of the characters is almost synchronized to the audio and the video. The displayed characters do not disappear for a while. All characters of a content disappear simultaneously about one second after they have been displayed. The text is displayed in the upper part of the monitor. When we display the video of the lecturer, it is put in the lower-right of the text.

Step 2) After the content disappears, the assessor looks at the printed question of the content.

Step 3) The assessor answers the question of the content without time constraint.

Step 4) Return to Step 1), if the next content of the teaching material exists.

For convenience, we refer to the case of presenting “text only” as “case 1”, “text and the audio of the lecturer” as “case 2”, and “text and the audio–video of the lecturer” as “case 3”. Figure 2 corresponds to “case 3”, and Fig. 3 to “case 1”.

4) Assessment of learning effectiveness: The number of assessors in this experiment is 17; they are Japanese males in their twenties. Each assessor (i.e., learner) assesses the contents displayed on the learner’s terminal. The contents displayed on the learner’s terminal are regarded as the stimuli to be presented to the learners.

We explain how the stimuli have been made for teaching materials A-4 and A-6. The number of stimuli that we have prepared for each teaching material is 648 (= 36 × 18). We represent the stimulus by $C_{i,j}$, where the subscript $i = \{1, 2, \ldots, 36\}$ implies the content number, and the subscript $j = \{1, 2, \ldots, 18\}$ the value which represents the manner of producing a stimulus. We have 18 manners of producing stimuli because of three presentation styles of a content (namely, “case 1”, “case 2” and “case 3”) and the six kinds of the average amount of the load traffic interfering with the media traffic as already mentioned in Subsec. III-A.

From among the 648 stimuli, we used 612 (= 36 × 17) ones in the experiment; i.e., 36 stimuli for each learner. This is because we have so arranged that each learner assessed a content only once; that is, the learner was never exposed to the same content with different presentation styles and different load traffic.

The stimuli a learner assesses are $C_{1,j}, C_{2,j}, \ldots, C_{36,j}$; the value of $j$ is different from learner to learner. The value for the first learner is $j = 1, 2, \ldots, 18$, that for the second learner is $j = 2, 3, \ldots, 18$, and 1, and for the $n$-th learner, $j = n, n+1, \ldots, 18, 1, \ldots, n-1$; the value of $j$ is changed in a cyclic order. As there were 17 learners, the stimuli $C_{1,18}, C_{1,1}, \ldots, C_{17,17}$ were not used.

We have made the stimuli for teaching material B in a similar way. The number of stimuli is 450 (= 30 × 15). Again, let us denote the stimulus as $C_{i,j}$, where the subscript $i = \{1, 2, \ldots, 30\}$ implies the content number, and the subscript $j = \{1, 2, \ldots, 15\}$ the number which represents the manner of producing a stimulus. In teaching material B, we set the average amount of the load traffic to five values; the 5.5 Mb/s is excluded. Therefore, each content can make 15 stimuli.

We used the 450 stimuli in the experiment of teaching material B. Each learner assessed $C_{1,j}, C_{2,j}, \ldots, C_{30,j}$. The value of $j$ for the first learner is $j = 1, 2, \ldots, 15$, and for the second learner, $j = 2, 3, \ldots, 15$, and 1. That for the $n$-th learner is set to $j = n, n+1, \ldots, 15, 1, \ldots, n-1$. The 16th and the 17th learner assessed the same stimuli as that of the first learner and the one of the second learner, respectively.

It took about 80 minutes for a learner to finish all assessment.

IV. Experimental Results

We show the experimental results of teaching materials A–4, A–6 and B in Figs. 4, 5 and 6, respectively. Each figure plots the ratio of the correct answer versus the average amount of load traffic for the three cases.

In order to clarify the tendency in the experimental results we utilize simple linear regression analysis, where the average amount of load traffic is employed as the independent variable and the ratio of the correct answer as the dependent variable. Let us represent the estimate of the ratio of the correct answer by $S_{X(Y)}$ [%]; the subscript $X = \{A-4, A-6, B\}$

<table>
<thead>
<tr>
<th>Teaching material A–4</th>
<th>Teaching material A–6</th>
<th>Teaching material B</th>
</tr>
</thead>
<tbody>
<tr>
<td>About 14 seconds</td>
<td>About 18 seconds</td>
<td>22 through 28 seconds</td>
</tr>
</tbody>
</table>

Fig. 4. Ratio of correct answer versus average amount of load traffic (teaching material A–4)
implies “teaching material A–4”, “teaching material A–6” and “teaching material B”, respectively. Similarly, the subscript \( Y = \{ \text{case1, case2, case3} \} \) specifies the presentation style of content. In addition, let \( L \) [Mb/s] denote the average amount of load traffic. Then, Eqs. (1) through (9) show simple linear regression lines thus obtained.

\[
S_{A-4(\text{case1})} = 78.6 + 0.420L \\
S_{A-4(\text{case2})} = 115.7 - 4.78L \\
S_{A-4(\text{case3})} = 137.1 - 7.83L \\
S_{A-6(\text{case1})} = 66.0 - 0.770L \\
S_{A-6(\text{case2})} = 88.6 - 3.92L \\
S_{A-6(\text{case3})} = 122.4 - 8.93L \\
S_{B(\text{case1})} = 44.7 - 0.105L \\
S_{B(\text{case2})} = 46.4 + 0.062L \\
S_{B(\text{case3})} = 86.8 - 5.92L
\]  

Figures 4 through 6 also plot the estimates given by the regression lines.

First, we examine how the three presentation styles of contents influence the ratio of the correct answer. In Figs. 4 through 6, we notice that the ratio of the correct answer in “case 1” tends to be independent of the average amount of load traffic. As the text–only teaching material contains a small amount of data, it is hardly influenced by the load traffic. If the average amount of load traffic is low, the values of the ratio of the correct answer in “case 2” and those in “case 3” are higher than those in “case 1”. This is because the lecturer’s audio and video reinforce the memory of the text \([4]–[7]\). However, as the average amount of load traffic increases, the ratio of the correct answer deteriorates in “case 2” and “case 3”. This is because the degraded auditory quality and/or visual quality interfere with learners’ concentration and memory of the text.

Next, we compare the teaching materials in terms of the ratio of the correct answer.

In Figs. 4 and 5, we notice that the ratio of the correct answer of teaching material A–6 is lower than that of teaching material A–4 on the whole. This is because the learner has a difficulty of memorizing the sentences in teaching material A–6, which consists of six short sentences. In particular, in “case 3”, as the average amount of load traffic increases, the ratio of the correct answer of teaching material A–6 rapidly decreases in comparison with that of teaching material A–4. This is because the degraded auditory quality and visual quality greatly interfere with learners’ concentration and memory of the text as the contents of teaching material A-6 are more complicated than A–4.

Referring to Figs. 4 and 6, we compare the ratio of the correct answer of teaching material A–4 to that of teaching material B. We then find that the ratio of the correct answer of teaching material B is lower than that of teaching material A–4. This is because it is difficult for the learner to memorize the content of the teaching material with long sentences. On the other hand, the decreasing rate of the ratio of the correct answer of teaching material B as a function of the average amount of load traffic is lower than that of teaching material A–4. This is because the contents of teaching material B are related sentences. The context helps the assessors fill in the gaps in the sentences.

Using Eqs. (1) through (9), we can show a guideline for the selection of the best presentation style of contents for a given amount of the load traffic in each teaching material. In teaching material A-4, the intersection of the regression lines given by Eqs. (1) through (3) corresponds to approximately 7.1 Mb/s (see Fig. 4). Therefore, if the average amount of load traffic is less than 7.1 Mb/s, we should utilize “case 3”; otherwise, “case 1”. In teaching material A-6, the intersection given by Eqs. (4) through (6) indicates approximately 6.9 Mb/s, below which “case 3” is the best; otherwise, “case 1” is an appropriate selection. The threshold load traffic for teaching material B is about 6.7 Mb/s, below which “case 3” is the best; otherwise, we should choose “case 2”. 
V. CONCLUSIONS

In this paper, we investigated the most suitable presentation style of contents according to communication quality in multimedia distance learning over IP networks by experiment. As a result, we observed the following. Audio–video added to text raises the learning effectiveness for learners when the degradation of communication quality is low. However, the degradation of the received audio and/or video decreases the learning effectiveness as the communication quality becomes lower. This is remarkable in complicated teaching materials to memorize. Furthermore, we show a guideline for the selection of the best presentation style of contents for a given amount of the load traffic.

We have some issues to be studied as our future work. For example, in our experiment, the text of the teaching material appears character by character automatically; we need to treat other manners of text display. We will also investigate the learning effectiveness with other teaching materials and the effect of the learners’ education background or skills on the ratio of correct answer.

REFERENCES