MULTIMEDIA: DELIVERY QUALITY METRICS

Ian D. Benest

Abstract *— It is the author's experience that science and engineering students following a conventional degree course, overwhelmingly demand that they be taught in the lecture theatre where the immediacy of being able to question the lecturer is of prime importance. If information technology is exploited in the lecture theatre, then that material can also be made available to students twenty-four hours a day via a personal computer. If the overhead projection slides are electronically bound together with an audio narrative that is synchronised to appropriate animation, then an on-line lecture is created that can be presented as if it were a video. While a definition for quality is elusive, this paper identifies how the specification and generation mechanisms impose an element of quality on the resulting on-line lectures and identifies what characteristics of the presentation can be measured for their contribution to the overall quality.*

Index Terms — on-line lectures, multimedia quality, user-interface design, total quality management.

QUALITY

There is an instinctive understanding of what is meant by quality; subjectively it refers to goodness and worthiness [1]. If a product, such as a car, is twice the price of another, the more expensive one is generally of higher quality, though not necessarily twice the quality. But what if there are two text books on computer graphics, both of equal price and length, both written by the same author, but one treats the subject to a much greater depth than the other? Which one is of higher quality? Subjectively it must depend on the technical background of the reader and the expertise to which that reader aspires. If there are two course modules, each titled computer graphics, then the subject would be treated differently if one were written for computer scientists – it would concentrate on the mathematics underpinning the graphical techniques – and one were written for media studies where those techniques would be applied. Each module might be equally well received by those students for whom it was designed. Clearly, "fitness for purpose" [1] is of paramount importance. But "fitness for purpose" is not simply about content. How that content is delivered is also important: how it is structured, and whether students are fully engaged during delivery from the beginning to the end [2]. Delivery quality is measurable and is independent of content quality. However content that is difficult to absorb, needs special attention at delivery time and the efforts made by the lecturer can be automatically detected. Total Quality Management also demands that "value for money" is considered; is the effort expended to provide a multimedia experience worth it in terms of extra resources and time needed for its provision.

ON-LINE LECTURES

An on-line lecture [3] is the automated mimicry of a "talk-and-chalk" episode traditionally given in the lecture theatre. Electronic slides are created as if they were to be printed on to transparent slides. The "talk" is a pre-recorded narrative that is synchronised with animation and the display of any supplementary still images or video fragments. Slides are changed as if a 35mm projector were being used. Synchronised animation includes such techniques as covering up then gradually revealing the slide content, identifying specific parts of a diagram while describing its operation, and showing temporal or sequential operations like data flow. In many respects it is like a video and can be played through from start to finish with one button press. But the user is also able to reach any slide within the lecture and play just that slide; they can pause mid-sentence and the system automatically re-winds to the start of the sentence when the pause is released. The script of the narrative is displayed "beneath" each slide.

On-line lectures are designed to support and supplement the live lecture rather than supplant it. Conventional transparencies are generated for the lecture theatre as are the student slide booklets. There is no technical reason why the on-line lectures should not be used directly in the lecture theatre with the audio narrative turned off and the lecturer single-stepping through the material just like a television weather forecaster. Only the lack of suitable facilities stops this from happening in the author's own institution. Ideally, the on-line lectures should be viewed prior to the live lecture so that students benefit more in the lecture theatre. But the evidence suggests that this rarely happens and instead, if on-line lectures are consulted then it is to view those parts of the lecture that were found difficult.

Ian D. Benest, University of York, Department of Computer Science, Heslington, York, UK, ian.benest@cs.york.ac.uk

It is the author's experience that computer science students do not wish to be taught by computer and they demand live lectures. So playing the on-line lecture as if it were a video in the lecture theatre, would also not be acceptable because the students feel that they are unable to interrupt and ask a question. The fact that they do not ask questions when the the lecture is given live is irrelevant! In many respects this is a pity because during the specification of the on-line lecture the narrative is honed to say precisely and succinctly what needs to be said in a form that can be interpreted and absorbed in only one way. This is not always true when giving the live lecture without a script.

Lectures rapidly impart information, but the absorption of that information is by no means total and so lectures can only "oil" the learning process which must take place outside the lecture theatre.

SPECIFYING ON-LINE LECTURES

On-line lectures are specified using a schematic paste-up mechanism in which symbols are obtained, dragged and dropped on to an electronic slide. Simple cover-and-reveal animation is obtained by inserting on to the slide a long thin rule-off symbol between, for example, two bullet points. More complex animation is achieved by creating successive layers of slides. Layers are copied and dragged to the next layer in the same way as for a symbol. The registration from one layer to the next is achieved by "page turning" to and fro between the layers. Page turning is invoked using two of the mouse buttons and the third button is used for placing the copy on the new layer. Page turning is inherent in the interface that is used for the specification and is the main mechanism for browsing the information space.

The narrative is scripted, being broken up into sound bites of usually one to four sentences; in effect, an "audio paragraph". The script is practised, recorded on to audio tape in quiet surroundings, and then copied into individual computer files so that each file contains an audio paragraph. Sometimes the sound bite can be just one word long depending on how the animation is to be shown. A dynamic hyperlink to each audio paragraph provides a symbol that can be dragged and dropped on to the slide – the order of play being left-to-right, top-to-bottom as physically placed on the slide. Like the rule-off symbols, the dynamic hyperlink symbols are removed from the generated slide, so they are not seen during the final delivery. Music can be specified to break up parts of the presentation. When they abut an audio paragraph they are joined with an appropriate fade of the music before or after the audio paragraph.

Supplementary images and movies are dynamically hyperlinked and placed appropriately on the slide so that they appear at the right time in the presentation.

CHALK-AND-TALK LECTURES QUALITY GUIDELINES

The lecture should have structure and that structure should be made clear to the students [4]. A lecture should have the following pattern (similar to that suggested by Koumi [5]).

1. Engage the students' attention by telling them what they are going to learn and why they need to know the information;
2. remind them of what they should know in order to cope with this new knowledge; then,
3. teach the material; then,
4. tell them what they have been taught; and
5. connect this material to what follows by identifying what they will now be able to assimilate given this new knowledge, in preparation for the next lecture.

Lecturing continuously for an hour ensures a low take-up of information. Human beings are just not designed to maintain such a prolonged period of attention and if a point is made that is unclear requiring considerable internal inferencing before absorption then the point that follows is likely to be lost. Ideally "learners should frequently experience 'delighted revelation' as they suddenly grasp a difficult concept" [5], but this is unlikely to apply to everyone in the auditorium. So there are a number of techniques which, if employed, help to maximise the amount absorbed and therefore increase the students' learning efficiency.

- During the lecture, key points should be repeated with pointers back to the main theme (or story-line) of the lecture. This is to help those wandering minds get back on track.
- Frequent changes in presentation help to keep the student attentive. For example: changes in revelation technique on each slide; one slide can contain bullet points, the next can be a diagram with salient features identified and useful animation, the next slide can be a movie fragment and the next a still image, and so on.
- Pausing [4] in the narrative is a useful assistant for absorbing a complex idea prior to the arrival of the next one. Of course the pause must not be "filled with distracting or interfering stimulation" [4]. The system described in this paper ensures that the narrative, delays, and animation do not occur in parallel although it is possible to specify a slide show with background music.
- The spoken narrative can also be varied, from an exposition of facts, to the explicit pose of a question or even the more powerful implicit pose of a question invoked by the partial revelation of an answer. But these types of changes are currently impossible to detect automatically.

Boyd [6] and Yorke [7] recommend a narrative speed for news broadcasts to the general public, such as those produced by the British Broadcasting Corporation (BBC), to be at a rate of 180 real words per minute. The actual pace should be one that is comfortable for the lecturer, clear to the student, and suits the complexity of the lecture course. The BBC World Service, broadcasting to an audience whose first language is not English, aims for a speed of between 140 and 160 words per minute [6]. So when the lecture is being introduced or summarised or when any of the new concepts are being summarised, the BBC News rate could apply. Given that the central material in an on-line lecture conveys complex concepts (to those students) it ought not to be delivered with the same urgency as that of news; the author recommends a delivery speed that is closer to that which is appropriate for the World Service. "Slower speeds with longer pauses" permit "superior comprehension" [4].

While pausing plays a role in concept absorption, pauses also play a sub-conscious role in either linking together or dissociating topics [5]. For example where there is to be a topic change, there should be a long pause after the narrative and before the new topic is revealed. Once revealed there should be a short pause followed by the start of the next sound bite. Where the topic is unchanged, but a new point is being made, a short pause follows the end of the narrative, the new point is displayed, and there is a further long pause before the narrative continues. Incorrect pacing leads to sub-conscious messages conflicting with the lecture content in a similar way to the conflicts that occur when the word "but" for example, is followed by a positive statement reinforcing the previous point instead of a negative (limiting) statement that the "but" word was signalling. The system described in this paper automatically installs the correct pausing within a slide (associating) and between slides (dissociating).

Where the narrative contains reference to the speaker, to the audience, or to people in general, it gives the material a human interest flavour that revives flagging interest. This can herald the speaker's opinion, provide information about how other people perform or think, or identify personal experiences that are often enlightening. Part of a lecture can include a simulated question-and-answer session where a person quizzes an expert colleague to extract their personal views. Such a lecture would give a high human-interest rating. While human interest is measured, there are no lectures in the corpus that include such a question-and-answer session.

Hot links for viewing supplementary material are automatically installed by the on-line lecture generator ensuring consistency within and across lectures.

MEASURING CHALK-AND-TALK LECTURE QUALITY

Chalk-and-talk lectures are traditionally evaluated using questionnaires identifying student opinions but not the effectiveness of lectures. If it is completed at the end of a set of lectures, the scores will be influenced by the last one or two presentations. If the module is a compulsory one and particularly if the content is regarded as difficult or is unpopular, the ratings may not be as high as for an optional module with relatively straightforward content. A twenty-five percent return rate can be expected. Are these completed by students with "an axe to grind", and should the views of the silent seventy-five percent indicate complete satisfaction? While student feedback plays an essential role in quality assurance, it is not without bias; a more objective analysis available before delivery to students would be useful. On-line lectures can be objectively analysed.

BASIC MEASUREMENTS

Visual complexity is simply defined as the ratio of the number of black pixels to white pixels; a slide heavy with elaborated bullet points will be "darker" and is usually less acceptable than one with large white "open spaces". All symbols included in a slide are checked for their visual complexity measure, and if within the range of 30% to 70% they are taken to be pictures (not graphical symbols) and in themselves they are not regarded as complex. Each such picture is discounted (including the occupying area) from the slide's visual complexity measure.

How easy the narrative is to understand can be gauged by Flesch's readability yardstick [8] that is a function of both the complexity of a sentence (by measuring the number of real words per sentence) and the complexity of the words (number of syllables). Flesch offers an interpretation of this measure (Table 1).

TABLE 1
FLESCH'S READING EASE

Value	Interpretation
0 – 30	very difficult
30 – 50	difficult
50 – 60	fairly difficult
60 – 70	standard
70 – 80	fairly easy
80 – 90	easy
90 – 100	very easy

Human interest is measured by the number of words that refer to people (e.g. I, I'm, her, me, we, your, etc.), as a percentage of the total number of words in the narrative for that slide. The interpretation of the measure is given in Table 2; this is a variation of that recommended by Flesch [8].

TABLE 2
PERSONAL WORDS

% of personal words	Interpretation
0 – 2	dull
2 – 4	quite interesting
4 – 7	interesting
7 – 11	really interesting
11 – 17	highly interesting
17+	dramatic

With the script of the narrative and the actual spoken version of that narrative available, it is possible to compute the lecturer's speaking rate. This rate includes only those pauses that naturally occur within an audio paragraph.

The time spent pausing on each slide is calculated by summing the number of delays imposed by the automated generation process and adding the pauses between sentences within each audio paragraph. Speech files are automatically indexed during the lecture generation process to provide the pausing and roll-back feature and so the delays are extracted from these index files. The total run-time for each slide is the sum of all the run-times of each constituent audio paragraph, the pauses inserted between them and the long pause at the end of each slide. The time spent pausing as a percentage of total run-time for each slide gives the pause-time measure.

RESULTS AND DISCUSSION

Fifteen on-line lectures, containing 237 slides covering a module on digital circuit design for first year undergraduates were created by the author and analysed. Each lecture was designed to be given as a live fifty-minute presentation. Where that presentation did not fill the time available, the author provided appropriate in-class questions which were first attempted by the students before the correct answer was explained by the author; the first ten lectures required these questions. In general, the live lecture took almost twice as long to complete as did the on-line version. The aggregated results are given in Table 3.

The average run-time for each lecture was 22 minutes, so giving lecture 12 live, is quite difficult to fit into fifty-minutes. MacManaway [9] reports that most students (84%) think that 20 to 30 minutes is "the maximum time during which they could tolerate uninterrupted lecturing and note-taking during a lecture period" [9]. So on average the set of fifteen on-line lectures ought to be acceptable. On average, 90% of a slide is white space. In the author's judgement, 83% white space is the minimum if the slide is to remain clear.

The average Flesch readability measure is interpreted as "standard" which is equivalent to the writing in Time Magazine [10]. This measure compares favourably (in the author's view) with the Communications of the ACM (43.19) and, surprisingly, the IEEE magazine Computer (28.42) which are measured to be "difficult" and "very difficult" respectively [10]. Over all the lectures, the human interest rating is "quite interesting"; in the author's view a rating of "interesting" would be more appropriate.

The average speaking rate is that suitable for the BBC's World Service, which to the author, seems to be appropriate. On average, a slide's run-time is 1 minute 20 seconds of which about 25% is taken up with pausing. Given that the live lecture takes almost twice as long as the on-line version, the author feels that the pause time ought to be nearer 50%. But during the live lecture, the presenter is moving around, changing slides, and so on, giving students the opportunity to absorb the latest information. What should the on-line lecture do? Such delays might invoke the normal empathic reaction to something in apparent distress, perhaps causing people to think that the computer had malfunctioned, because of its inactivity, thus distracting them from absorbing the information?

So it is possible to measure on-line lectures objectively and compare with published guidelines. The large standard deviations for most measures suggests a variety of attention-holding techniques are in place.

Total Quality Management demands an assessment of "value for money". Assuming that slides have to be produced by some means: either handwritten or with the aid of a desktop publishing package, and that lecturers have to structure their presentations and make notes to accompany their slides, these time-consuming activities can be discounted when comparisons are being made between on-line lecture production and live lecture production. The additional time consuming activity is concerned with recording and installing the narratives. For this, several practice sessions are necessary before taping, and then each audio paragraph has to be captured with a sound editing tool, edited so as to have no blank periods at the beginning and end, and then be inserted in the on-line lecture specification tool. Care needs to be taken to ensure that the right words are used during animation; visual directions are a particular problem. For example, "and now you see.." followed by a long sentence should not be expressed when the student only sees "it" once the sentence has finished. The change has to take place before the word "now" is played. The author has not timed this production process, but a very rough estimate would be that it takes at least ten times the audio run time and perhaps adds three hours to the preparation of each lecture.

TABLE 3
BASIC MEASUREMENTS – AGGREGATED RESULTS

Lecture Number	No. of slides	Run Time	Visual Complexity (%)		Flesch Readability		Human Interest (%)		Speech Rate (words/min)		Time on Slide (seconds)		Pause Time (%)	
		(min:sec)	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
1	13	22:41	8.54	1.75	69.53	12.82	2.77	1.97	149.33	12.84	104.69	37.58	26.67	5.64
2	12	19:19	7.40	2.18	68.20	16.30	2.92	1.89	146.27	24.12	96.61	57.55	35.03	17.60
3	13	17:02	8.59	2.00	70.98	10.67	3.15	2.38	159.63	20.77	78.59	46.90	35.42	17.75
4	14	14:43	9.45	1.64	70.28	12.79	2.64	2.87	154.89	14.76	63.09	34.10	23.56	4.76
5	15	12:25	8.75	1.81	63.32	14.92	3.20	2.61	156.87	18.48	49.64	28.01	28.29	7.22
6	18	19:43	10.85	1.53	61.75	13.16	1.17	1.07	155.98	13.65	65.72	29.88	27.63	8.39
7	10	13:42	9.95	2.05	61.89	16.24	2.70	1.68	158.52	14.74	82.21	62.68	24.29	4.44
8	19	18:34	9.95	1.54	63.87	13.75	2.21	1.67	151.10	8.89	58.63	23.95	25.91	5.53
9	10	13:09	10.02	2.10	53.96	14.27	3.00	1.67	170.39	12.10	78.92	44.66	24.32	6.77
10	17	22:09	9.79	2.19	62.28	13.75	1.53	1.29	156.27	17.14	78.18	22.43	22.55	6.05
11	20	30:44	10.32	3.03	57.22	13.80	0.85	1.06	143.61	9.51	92.22	39.13	21.71	6.75
12	20	41:23	10.67	1.66	58.70	9.89	0.80	0.68	153.39	13.86	124.15	64.65	20.17	7.65
13	20	25:54	10.02	1.93	65.78	10.81	2.00	1.70	146.53	12.45	77.70	42.53	22.35	5.60
14	16	30:08	9.97	2.15	67.49	14.00	2.81	1.38	142.94	13.85	112.97	49.25	26.16	6.37
15	20	25:18	12.33	3.09	56.11	14.00	1.75	1.41	155.60	12.37	75.91	44.81	23.27	6.38

SEMANTIC ANALYSIS

It is argued below that two quality aspects can be identified from the measurements taken. These are based on the guidelines in the literature. They are: the quality of presentation and the difficulty of the material.

Quality of Presentation

The factors below contribute to presentation quality. In all cases, extreme variations invalidate the rules.

- The less time spent on the slide the better.
- The less cluttered the slide the better.
- The slower the narrative the better.
- The more pausing the better.
- The higher the reading ease the better.
- The more personal words the better.

The absolute measures are ordered so that for example the least cluttered slide in a lecture is number one, the next least cluttered is number two. For each slide the resulting six factors are multiplied together and divided by the number of slides raised to the power of six. This presentation quality measure has a maximum possible value of one (the worst case). The measure for each slide provides a relative value: which is the best and worst presented slides and all others are rated in between. The same computation may be applied to each lecture to provide the best through to the worst presented lectures; in this case a normalising division is unnecessary. These results are shown in Table 4 where the best presentation is 1 and the worst is 15.

Difficulty of Content

The literature identifies the following factors that would contribute to the difficulty of the content.

- The more cluttered the slide the more difficult is the content because there is a need to give a lot of information in order to maintain a conceptual whole.
- The slower the narrative the more difficult the lecturer thinks is the material.
- By pausing more, the lecturer is deliberately (though maybe sub-consciously) giving students the opportunity to absorb difficult material.
- The lower the reading ease the more difficult the topic: scientific and technical journals are more difficult to read than general interest magazines.

Again these can be ordered at the slide level and a difficulty of content computed. The measure can also be computed at the lecture level; the results of which are shown in Table 4 with the most difficult rated 1 and the least difficult rated 15.

TABLE 4
SEMANTIC ANALYSIS RESULTS

Slide Number	Quality of Presentation		Difficulty of Content	
	measure	order best = 1 worst = 15	measure	order most = 1 least = 15
1	9000	4	4550	12
2	672	2	1080	6
3	420	1	2730	9
4	256000	9	12320	15
5	1152	3	3456	11
6	640640	11	400	3
7	171990	8	6318	13
8	127008	7	3024	10
9	108000	6	720	5
10	769824	12	9240	14
11	856128	13	336	2
12	3685500	15	1260	7
13	411840	10	2600	8
14	17550	5	462	4
15	2515590	14	198	1

CONCLUSIONS AND FUTURE WORK

The aggregated results provide an overview of each lecture; space does not permit the provision of results at the slide level and of course without access to the on-line lectures it is difficult to relate the results to each lecture and slide. Furthermore, while the results given in Table 3 are compared with published guidelines there is no grounding of the results in Table 4 with student opinions, so as to provide an absolute measure for comparative purposes. Is the lecture or slide well presented, or simply the best of a mediocre set?

Diagrams are often used to simplify the material being conveyed and can be a sign that it is conceptually difficult. Since the specification is a schematic (symbols interconnected with lines and annotated with text), the presence of one or more diagrams on a slide can be automatically detected. By adopting animation (other than simple revelation) the lecturer is using the technique to explain a difficult topic. Neither the presence of diagrams nor animation currently influence the difficulty-of-content measure.

Lectures should be topped and tailed; such slides might be identifiable by a faster narrative with a more difficult reading ease. Would the same characteristics identify a summarising mid-slide? These two characteristics would indicate the presence of motivating cues for learning, and support for absorbing difficult concepts. The rest of the material would be the lecture detail. Being able to detect these aspects could also assist corpus search mechanisms to offer lecture overviews, summaries, and content detail.

The analysis so far takes each slide as an independent artefact and does not identify how each fits with the others in the lecture. If the quality-of-presentation improves through the lecture, then it will help to retain attention until the end. If the difficulty-of-content increases through the lecture, then the absorption of new material will be less complete as the student tires. Ideally the difficulty-of-content through the lecture should reflect an inverted U or even an M if there is a mid-lecture review.

For an analysis to be useful to a lecturer, it is necessary to provide a verbal interpretation that offers constructive criticism which identifies the good and bad points. Since some concepts are difficult to convey simply, and some slides are necessarily packed, compromise is essential. A set of rules need to be applied to each slide and to each lecture identifying the good and bad aspects and ignoring the "reasonable compromise" slides.

REFERENCES

[1] Green, D, "What is Quality in Higher Education? Concepts, Policy and Practice", In: Green, D, (Ed) *What is Quality in Higher Education?*, The Society for Research into Higher Education, 1994, pp. 3-20.

[2] Melia, T, "Inspecting Quality in the Classroom: An HMI Perspective", In: Green, D, (Ed) *What is Quality in Higher Education?*, The Society for Research into Higher Education, 1994, pp. 38-45.

[3] Benest, I, D, "The Specification and Presentation of On-Line Lectures", *Innovations in Education and Training International*, Vol 34, No 1., 1997, pp. 32-43.

[4] Bligh, D, A, *What's the Use of Lectures*, University Teaching Methods Unit, London, 1971.

[5] Koumi, J, "Narrative Screenwriting for Educational Television: A Framework", *Journal of Educational Television*, Vol 17, No 3., 1991, pp. 131-148.

[6] Boyd, A, *Broadcast Journalism: Techniques of Radio and TV News*, Heinemann, Oxford, 1988.

[7] Yorke, I, *The Technique of Television News*, Focal Press, London, 2nd Edition, 1987.

[8] Flesch, R, "A New Readability Yardstick", *Journal of Applied Psychology*, Vol 32, No 3., 1948, pp. 221-233.

[9] MacManaway, L,A, "Teaching Methods in Higher Education – Innovation and Research", *University Quarterly*, Vol 24, No 3., 1970, pp. 321-329.

[10] Lemos, R,S, "Rating the Major Computing Periodicals on Readability", *Communications of the ACM*, Vol 28, No 2., 1985, pp. 152-157.