

ASYNCHRONOUS VOICE: SOME DESIGN ISSUES

PHILIP MARRIOTT, University of South Australia

*School of Communication, Information, and New Media, St Bernards Rd, Magill, SA 5072, Australia.
Tel: +61 8 8302 4742, Fax: +61 8 8302 4745, Email: Philip.Marriott@unisa.edu.au*

ABSTRACT

Poor business performance and lost equity values have cast doubt for some on the future viability of electronic commerce. Most of this attention focuses on the EC channel, just one aspect of electronic commerce. The paper examines the effectiveness of the EC channel for seven types of products and infers that for four of them, the EC channel is unlikely to be the best way to sell and deliver goods and services. Much of the value to be had from this channel has already been captured and over investment in it may result in continued contraction. A more mature EC model, the Strategic Electronic Commerce Model (SECM), is proposed that provides a framework for balanced EC investments across the value chain and continued opportunities for EC investment.

We thought that it would be very convenient to use voice-based e-mail to provide quick and short replies to their ever-increasing numbers of e-mail. However, after an initial enthusiasm, we returned to using text. While we believed that voice based e-mails must be attractive, there must be differences between the two that need to be made explicit. This paper argues that asynchronous voice based communication needs further consideration. The reasons why this form of communication needs to be accommodated on the Internet and some of the social issues that act to inhibit the attractiveness of this technology are presented. It is hoped these discussions and findings will encourage designers to give more thought to developing new systems that support asynchronous communication. As a starting point, it is believed that successful

communications systems design needs to consider the underlying social forces acting upon technological design. In this case, it is the tension between oral and written communications methods.

The invention of the alphabet by ancient Greeks heralded a new form of mass messaging - asynchronous (different time, different place). Accounting, using small clay pots with inscriptions had already provided the cornerstone of trade - asynchronous records. The printing press allowed far more people to talk to others without them being present. Messages could be received at a different time from when they were sent. The next stage was the recording of voice, normally attributed to Graham Bell. Taped messages could be sent in place of letters. This was refined into the telephone answering machine and then to voice based e-mail. More recently, the voice

file (.wav) web-board has been developed which allows threaded asynchronous discussion using voice.

Asynchronicity has created a huge social and economic change. Most of us can remember how much easier international communication became with asynchronous e-mail. It may well be the corner stone of globalization. Historically, nations that mastered asynchronous communication, i.e. had a written language, tended to dominate those that did not.

Literacy. So what are the attributes of writing? Ong (1982) reports that, of the 10,000 languages used by man, only a little over 100 developed meaningful writing. As a communication system, writing includes books, scientific articles, newspapers, dictionaries, diaries, e-mails, web pages, databases, definitions, contracts, reports and encourages the attitude "if a thought is not in print, then it has not yet matured and should not be believed." Writing requires years of training; to attain a reasonable standard maybe 4 years, to be very good at it maybe 20 years for the average person. It is a system that is very good for storing our explicit memory. It helps us to remember complex things like telephone numbers and exact legal phrasing. It allows communication with people who we have never met and even after the writer is dead. In this respect, it gives durability to the ideas of writers over non-writers. Written work enables readers to work at a speed convenient to them and allows the reader to re-read or skip portions. Long and complicated arguments can be structured correctly, as editing of once written phrases is possible. However, It is not a good way to keep secrets and an unwise medium to record rumor.

The socialist movement, with its emphasis on social and economic improvement for the majority, diverted a large percentage of taxes into mass education - a 20th Century phenomenon. As a result, literacy rates have soared in the last 70 years in the more developed countries, with less developed countries catching up as they invest more to develop this communication system. It is usually assumed that economic development and literacy are correlated. Whether it is

literacy or a general education in thinking, science, history and geography is not clear. Nor is the connection apparent between literacy and technological innovation. However, because literacy has now acquired such a high political status, millions of dollars are spent encouraging writing skills over oral skills. Therefore, is it reasonable to expect literacy skills in developed countries to be more highly developed than oral skills. This may have implications for the use of voice-based web-based discussion as explored in this study.

Asynchronous communication, because of its loss of context and immediacy, encourages thinking about 'essences' (or averages) rather than situational thinking. Ong (1982) provides an example using the set: a hammer, a saw, a log and a hatchet. The question is to identify the odd one out? To those of us brought up to thinking with Plato's essence, we may identify the saw, hammer and hatchet as tools, so the log stands out. Ong reports on oral people saying that, if they had to do without one of these objects, then they would discard the hatchet because the saw did a better job. When it was pointed out that the log was not a "tool", the response was that the others would be no use without the log and the log would become a tool once fashioned. The oral person is focused on relative use, i.e. situational thinking and may be a more useful way for innovators to think.

Ong (1982), in his study of orality vs. literacy, points out that writing tends to encourage conservatism. Oral cultures do not have history books, dictionaries, and operating manuals. Further, writing encourages standardised learning both in terms of process (because teachers all read the same "how to do it" books) and in terms of content because once an explanation of a phenomenon is written down; there is a tendency not to think about any background reasoning. This conservatism and standardisation of thinking is expected to save time by not "reinventing the wheel" but it acts against innovation. Ong (1982) also points out that text cannot be questioned, or change its mind. If "bad" knowledge is written down, it can be corrected only if all the copies are destroyed, else it will "infect" readers for years to come. Writing

also encourages the “objectivity” of knowledge, problem solving and decision-making, while allowing interpersonal skills to decline and thus discourages a ‘people’ view of technical problems - something that has plagued IS research (Butler 2000).

ORAL/AURAL

Oral and aural are bound together, speaking needs hearing – together they mean communicating by listening and talking to others. Those who prefer to use this sense like to rely on phone calls, talking it through, discussion, argument/debate, telling stories, hearing good speakers, they enjoy television, videos, films, riddles and meetings. This form of communication is usually associated with tacit knowledge, which makes use of body language and simple diagrams during discussion. It is also usually connected to a preference for kinaesthetic rather than reading. Kinaesthetic means learning from experience, doing, having a go, touching, trial and through error and practice (simulated or real). Ong (1982) argues that oral communication requires a different and more extensive use of thinking which is more “immediate” compared to text. He also argues orality is additive rather than subordinative; aggregative rather than analytic; makes for redundancy rather than copious; conservative, participatory rather than objectifies; situational rather than abstract. While some of these qualities do not suit science, they do suit community learning in its broader sense.

Anyone who has tried to have a detailed debate, using keyboard based “chat rooms” on the Internet, will appreciate how slow and limiting written words are in expressing ideas quickly. Media richness theory, plus every e-mail user’s experience, reinforces the limitations of this form of communication. Tacit skills like trade skills, horses riding and dealing with relationships are hard to make explicit (written down). The industrial revolution occurred with very little reading and writing around. Socrates was concerned that books did not take enough account of the context of the audience. Rather he, like Hegel, believed that knowledge creation required the

interaction of interested parties. Legal arguments in a court work in the same fashion.

Most cultures have taboos or secrets that cannot be written down. These taboos are knowledge that only certain people are thought to be able to handle. This occurs in business. There are many conversations, practices and trade secrets where a written record is not thought to be appropriate. Governments discourage the publishing of information about bomb and drug making - they have censorship laws. These types of knowledge must be communicated carefully. Writing makes controlling access difficult. Secrets are better managed in the oral form.

So, in the competition between the orality and literacy, both have their advantages and disadvantages; a balance seems appropriate. Ong (1982) sees this shift from oral to literate as a shift in emphasis from use of the ears to use of the eyes. Yet we literates complain about excessive noise pollution and tired eyes from watching too much TV (a non-literate activity). Our senses are most likely getting more of everything as population and communication technology increases. However, given the recent development of voice based e-mail and web boards, it may be more useful to compare synchronous and asynchronous. It is suggested that the extensive discussion about rich communication to exchange knowledge and meaning is referring to synchronous communication. It may be hard to have meaningful dialogue at the level of tacit knowledge sharing with people we have never, and may never meet. Synchronicity has large ramifications for knowledge sharing. Asynchronous communication is not expected to assist Socratic questioning, dialectic argumentation or any other form of Hegelian/Habermanian inquiry systems.

Asynchronous. The advantages of asynchronous communication include being able to communicate effectively with someone who is unavailable at the same time you wish to communicate. This will be particularly the case for people living in different time zones. Having time to think about the content of a reply makes asynchronous communication attractive. It also appears non-imposing on the

receiver, which may be important to the sender if there is a power difference between the receiver and sender. Moreover, asynchronous communication requires something physical be supplied, thus allowing it to double up as a formal and accurate record of a communication. This is important with contractual messages. Asynchronous communication allows the receiver to better time-manage when to listen to messages. A telephone call requires your immediate attention, but you can deal with a letter at your convenience. Asynchronous can also be used when the sender is trying to avoid a face-to-face confrontation, maybe to avoid awkward questions or embarrassing retaliation. When it is in text form (as opposed to a recording), it can be very extensively crafted prior to delivery. Moreover, it can mask identity (age, gender) or speech problems (accent, mute). This is particularly relevant with long messages.

Kock (2001) undertook a study of asynchronous use of group support systems. He found a difference between synchronous and asynchronous. First, he found that, while synchronous increased the quantity of ideas, asynchronous improved group process efficiency especially with 'group set-up' costs. There was little time wasted organising meetings, in informal social interaction and in introductions when asynchronous methods were used. While there was less volume of ideas generated by the group, the responses were more reflective and considered. Shirani et. al. (1999) have reported similar results. Though there was less number of ideas generated using e-mail over face-to-face meetings, the ideas were more considered, more in-depth [inferential]. They argue that asynchronous communication has the advantage of a deeper analysis of problems. Therefore, they considered a need for the appropriate mix of synchronous and asynchronous communication. In both these studies the asynchronous communications were in text format.

IDENTIFYING PROBLEMS

From the preceding discussion we find that it is worthwhile trying to develop both

synchronous and asynchronous voice based communication via the Internet. Apart from technical issues, new communications methods made users nervous. For example, e-mail made some people nervous when it was first introduced but given the history of letter writing for most people this soon passed. Yet, protocols for e-mails are still being developed. The same was true of mobile phones. Taking calls in a public space still make some people reluctant to use mobile phones. It is assumed the same is true of asynchronous voice based communications. Many people complain about feeling uncomfortable about leaving messages on answer phones, although this does seem to be passing.

An opportunity arose to explore the strength of this type of nervousness in response to asynchronous voice based communications using a voice-based web-board being used with a large first-year university subject. These were 'communications' majors studying modes of Internet communications. The subject was run over 5 months, and initially the students were required to post voice messages. Later, this was relaxed to allow the choice between voice and text. The 600 students were required to post a summary of their readings each week to a threaded online voice/text forum. Forums varied in size from 20 to 80 participants. The intention was to provide an incentive for students to complete required readings before each weekly tutorial and to stimulate discussion. Postings were an assessed part of the course.

Wimba Voice Board software (www.wimba.com) was used to facilitate the forums. This software has both a client and server component. The server component was installed on a server at the university. Participants then connected to the server through a web browser and the client software automatically downloaded and ran embedded in their web browser. The Wimba software installed easily on the server and the client software installed seamlessly into most participants' web browsers.

Figure 1 is an example of the Wimba Voice Forum interface.

This experiment was conducted in a context where participants are computer literate (to at least a basic level), predominantly young, and motivated to participate. Most participants used computers in ‘public’ computer labs with only a few making postings from home computers. Participants could choose to make contributions to the forums as text or voice or a combination of text and voice. Text was input through the keyboard. Voice was typically input using a microphone/headphone headset through the computer’s sound card. Participants could make contributions from any location on the Internet. For most participants the choice was between using computers at university or using a computer at home. All participants were required to purchase headsets prior to commencing the course.



Figure 1: Wimba Voice Forum Interface

Information on the usage of the forums was gathered by examination of server logs and a questionnaire survey of participants at the completion of the course.

We noticed that, as participants used the forums, over time, the frequency of text postings increased with a proportional decrease in the frequency of voice postings.

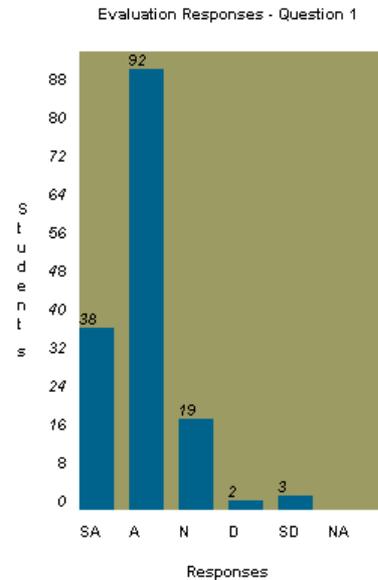
We wanted to explain this difference and this provided the means of exploring issues of personal social tension from using the new technology.

RESULTS

Our first concern was whether participants found the software difficult to use. We wanted to know whether the design of the software was a significant influence on the decision to post in text or voice. Most participants found the software easy to use with no participant claiming the software was difficult to use (see Figure 2).

Question : I found the Wimba Voice Forums easy to use

* Total number of students submitted the responses = 154
 * Mean = 1.961



SA = strongly agree, A = agree, N= No strong opinion, D = disagree, SD = strongly disagree

Figure 2: Survey question 1

We asked participants whether they preferred to post text messages. Most participants indicated that they preferred to post text messages. However, a small number had a preference for voice (see Figure 3).

We asked participants whether they preferred to post voice messages. A small number strongly supported voice postings

Question : I prefer to post text messages to Wimba

* Total number of students submitted the responses = 154
 * Mean = 2.130

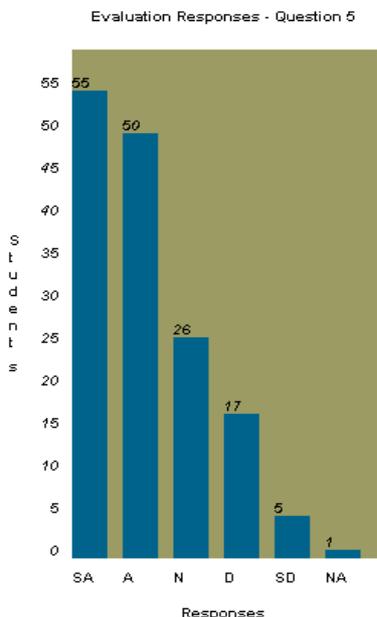


Figure 3: Survey question 2

while most participants were evenly divided between having no strong opinion to strong disagreement (see Figure 4).

Table 1, shows the selected free text responses to questions probing preference for voice/text. These comments are explored further in the discussion section of this paper.

DISCUSSION

It is clear from the students' comments that the majority of participants, by the end of the trial, preferred to post text messages. This is strongly indicated. Even though the server logs show that the majority of messages had a voice component, it would seem that, while participants have predominantly made voice postings during the trial, the majority drifted back to text. It should be remembered that they were under some pressure to use the voice at the start, but this requirement was relaxed over time. What then was wrong with posting voice messages?

The most common comment was on the embarrassment of talking to a computer in a public place like the computer lab. Participants

Question : I prefer to post voice messages to Wimba

* Total number of students submitted the responses = 154
 * Mean = 3.594

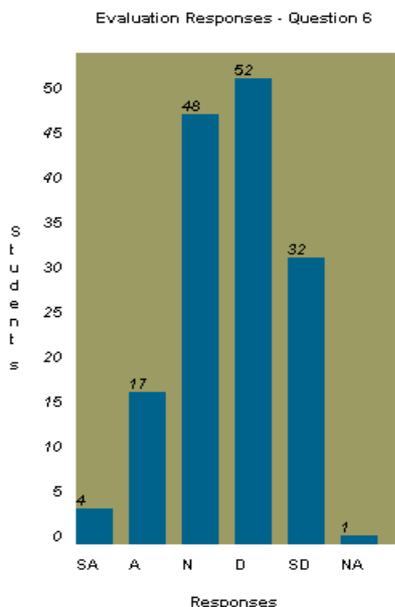


Figure 4: Survey question 3

said they felt self conscious when talking to an unresponsive computer and many voice postings are, in fact, spoken in such a quiet voice that they should be termed whisper postings. It is interesting to note that some of the participants, while reluctant to speak publicly into their computer, will happily talk into a mobile phone in the same setting. Perhaps with time and familiarity, the act of talking publicly into a computer will be commonplace and will not cause embarrassment. It is not clear if it was really a concern for fellow lab users or whether it was because they had to talk into a void - there is no one at the other end, no feedback, recording is like talking to yourself. It is suspected that the students also felt uncomfortable apparently talking to themselves in public.

Participants also report that they do not like hearing their own voice when they played back their message. It is expected that this is a transitional problem; people who often record their voice seem to get accustomed to the sound.

Table 1: Selected Comments.

The Wimba voice forums allow postings to be made as text or voice. In this course some students made only voice postings, some made only text postings, and some made a combination. Which did you prefer to post, text or voice or both?	Why did you prefer one type of posting (voice / text)? Give reasons.
Voice	Easier
Text, definately!	I feel silly speaking to the computer! I don't mind doing voice in class, but when I'm in the public pools I really feel uncomfortable doing it. But that's not a problem for the Uni, but for the students to deal with.
Text postings.	I found that typing allowed you to delete mistakes more easily and gather thoughts, while with the voice postings you really had to think and plan and concentrate on what you were going to say. I also think that voice postings are a little embarrassing!
I preferred to do voice postings only.	While speaking you can simply say your thoughts, but with writing you have to be careful of grammar and things like that. Talking was much easier.
I prefer text	Because I feel stupid talking to a computer, no matter how many people are doing it.
I prefer text messages	Because I can think of more things to say. When you post a voice message you 'fumble' and don't say what u want to
I prefer to post text	Voice postings (unless utilized at home) are a disruption in computer pools whilst other people are trying to do their work. Text postings are just as effective without the disruption
voice postings with PS text postings.	I like the sound of my own voice. Unfortunately this is true. I do however think I get more across through voice as it adds another dimension to discussion topics - more cues than text. I do think that a combination would be very helpful in many situations. A dot point summary in the text with an elaborated version on the voice posting. I am looking forward to using this facility more and more. All I have to do is train associates to use it. How did we get on Wimba is it expensive at home?
Voice and text	I liked to do both for the practice. However, I was very tempted not to do voice postings, as it was quite embarrassing often to be the only person in the room who appeared to be talking to themselves!
I preferred text postings.	For one I don't like the sound of my recorded voice and secondly I felt a bit self conscious sitting there chatting away seemingly to myself.
Text - voice is embarrassing in a room full of people	see above - also you can proof read and edit text and take time to write exactly what you want to say.
I always preferred to post in text, only one of my posts were voice postings.	I preferred post in text, because it's uncomfortable to do a voice post when people are trying to do quiet study

Table 1: Selected Comments, Continued.

Text messages were my choice of messaging.	Talking to your computer in a packed computer pool room can be slightly embarrassing! Which is why I opted for text messages.
I much preferred to post text messages as oppose to a voice message.	I find it more comfortable to type rather than speak despite it taking a little longer.
text	With text, you don't have o talk loudly to yourself in a very quite computer pool environment.
I prefer to use text because it is more convenient.	Because text contributions are easier to do if you don't have a computer where you can plug your headphones into.
I preferred the text messages because I do not like the sound of my voice. Also with the text messages I was able to see what I was writing and as a result was able to change what I was saying at any time that I pleased.	I preferred the text messages for mainly the above reasons that I have already mentioned.
i preferred to post text rather than voice	because you can get across a lot more clearly what you want to say when you can revise it properly before posting it. When listening to a lot of the voice postings, you tend to hear people saying a lot of umms and ahhs and not really concluding with a point. it is also a lot less embarrassing to write a point of view in a silent classroom than have to try and whisper it into your mic.
I preferred to post text messages rather than voice	I like text because I really don't like the sound of my voice. Text is easier as well because you don't have to talk into your mic in a quiet classroom and feel stupid
both - depends on topic	text is useful for longer entries voice is a bit like speaking in public, difficult to do
I used a combination of both, but preferred voice.	I enjoyed it as something different to do as opposed to typing, and I did not feel embarrassed about speaking in computer pools as I know many other students did.

Another common comment concerns the technical difficulty of making a voice recording. While the Wimba software performed flawlessly, the participant was required to connect a microphone and headphone/speakers. This added complication was enough to deter a number of participants from making a voice contribution. Whereas it is likely that most computers will allow text input, participants were not confident that the voice input on an unfamiliar computer would work. As new multimedia computers are better designed, it may be that audio input and output will be built into the PC so sending a voice message will then not require any special actions over sending text.

Editing a voice contribution prior to posting required the whole contribution to be re-recorded from the beginning, regardless of how small the error. In contrast, an error in a text entry could be fixed easily. It is reasonable to expect the likelihood of making an error in a voice contribution would increase proportionately with the duration of the recording. Therefore, a long voice contribution would take an increasingly long time to get error free. This would seem to encourage shorter voice postings or suggest that longer voice postings contain a higher percentage of errors. This is reinforced by participants complaining that the postings of their colleagues contained a number of annoying speech errors. This may begin to explain the fact that participants preferred to consult text

rather than voice postings of other participants. We therefore have a situation where most participants also preferred to receive text messages.

Those participants who preferred voice claimed it was because it easy and provided them with a richer means of communication that contained verbal cues and an emotional context. Why was it easier for these students? Are these participants better trained / more comfortable in giving monologues? These questions need to be followed up.

Some participants used both voice and text together. Typically, the text was written as a series of dot points and the voice component elaborated on these points. This hybrid might be the appropriate use of the technology.

The people involved in this trial were a group of undergraduates, reasonably competent in text-based communications. They have never been trained in voice-based work in the way actors and media presenters are. They did not have any disabilities like blindness. All of them could type on a computer keyboard. In order to send voice-based messages they had to plug in a headset and talk into a void, both of which can be embarrassing. They were under no operational pressure to use voice-based messages, such as having to work with people who could, or would not, write English. Nor did they gain much from any added emotional depth provided by the voice rather than text as the students were not emotionally attached to each other, as might be the case with family communications. But still there did seem to some demand for voice based asynchronous communication, a minority seemed to prefer it.

CONCLUSION AND IMPLICATIONS

The first half of this paper argued for voice based communications while the second half has tried to reveal the social factors that influence use of available technology in voice based communications. A majority of people do not wish to communicate by writing is evident by high usage rates of telephones. However, these people are often limited to synchronous communication, which may be what they seek. But, in some cases

asynchronous communication may be preferable both because of time differences and because of what is to be communicated. Appropriate technologies need to be designed.

The basic technology itself is emerging with Wimba is an example. The finer details of personal preferences, styles and habits need to be considered by the designers. The main social factors influencing the use of Wimba type technologies were that voice files could not easily be edited after mistakes and that they felt embarrassed by having to 'talk to the computer'. As with telephone answering machines, one has to learn to record a message with no feedback. If asynchronous voice based communication is to complement text based communication, then these two issues need to be overcome. The issue of being embarrassed by new communication technologies may simply be transitional. New means of communicating do seem to disrupt social habits.

The troublesome technical issue is that one cannot easily edit voice messages. If it were assumed that if one could, then some of the attributes of oral communications such as developing different thinking skills would also alter. However, there does still seem to be a demand for easily finding a 'word' in the middle of a taped message. There would appear to be at least two possible solutions to this. First, the development of wave recognition software so that you could repeat a word and the program locates that spoken word in a tape message. The second would be to use parallel tracking to allow flags to be tagged along the length of a spoken message, maybe by sentence or in 10-second slots. Then the program could display the 'tag name' for whatever section of the spoken message was being replayed. The users could then ask the program to replace one section with a revised comment.

This is an exploratory study. It will be interesting to study the adoption of asynchronous voice based communication technology of people who not familiar with text-based communications and have a preference for a learning style that is oral. It will also be interesting to study if this technology is appropriate for people with

typing impairment. The role of setting or context (social or business communications) in influencing adoption of asynchronous voice based communications technology needs to be studied. However, it seems reasonable to the authors that any form of communication that adds to our diversity should be encouraged rather than dismissed. While the Internet started as a facilitator of primarily text based communication, the new multimedia opportunities need encouragement.

REFERENCES

- Butler, T., "Making Sense of Knowledge", ECIS Proceedings, Vienna, July, 2000.
- Checkland, P., and S. Holwell, *Information, Systems and IS*, Wiley, New York, 1998.
- Ong, W. J., *Orality and literacy: the technologizing of the word*, Methuen, New York, 1982.
- Shirani, A., *et al*, "A Comparison of two Technologies for Synchronous and asynchronous Group Communication", *Information & Management*, vol. 36, 1999, pp. 139-150.
- Kock, N., "Asynchronous and Distributed Process Improvement", *Information Systems Journal*, vol. 11, 2001, pp 87-110.

With thanks to Wimba.com

THE AUTHOR



Philip Marriott (M.Ed.St, Gdip Systems Analysis, B.Ed) is a Lecturer in Information Management and Electronic Communication. He has completed studies in Education and Computer Science and is completing a PhD in the area of online education. Philip teaches and conducts research in online publishing on the Internet. He manages and develops multimedia and online software in a wide variety of commercial and academic contexts. [contact: philip.marriott@unisa.edu.au]