

Bimodal conversation during on-line lectures

Russell Schick, Dept. of Computer Science, University of Toronto, Toronto, Canada.
rschick@dgp.toronto.edu

Dominique Scheffel-Dunand, French Studies, York University, Toronto, Canada.
dsdunand@yorku.ca

Ron Baecker, Dept. of Computer Science, University of Toronto, Canada.
rmb@dgp.toronto.edu

Abstract: In evaluating the impact of audio communication on student engagement and learning during online lectures, we observed a phenomenon we call bimodal conversation. While watching lectures, students were given both an audio and a text channel to communicate within their study groups, and it was found that students could engage in coherent conversations using both modes simultaneously. We present an analysis of bimodal conversations from the point of view of conversation analysis, which examines the allocation of conversational turns, and discourse analysis, which examines the context and interpretation of each utterance. Mode selection and mode switching are seen to impact conversations at both these levels, yet students are able to intuitively negotiate both modes simultaneously with few problems. The combination of audio and text modes in the context of online lectures proves to have advantages over either mode taken independently, as students can use the text channel to avoid interference with the lecture audio, but also reap the benefits of real-time audio communication.

Introduction

Online webcast lectures offer the possibility of highly distributed collaborative learning in many settings, from university undergraduate education to continuing professional development. Unfortunately distance learning systems that simply use webcasting as a scalable transmission medium can result in a disengaging experience for students (Schick, 2005). One way to combat disengagement is to provide interactive features for students that transform the plain multimedia broadcast into a more engaging experience. ePresence is an example of such a system which builds upon a consumer-grade webcasting infrastructure through the addition of interactive slides, application sharing, and communication tools that help to engage students through participation (R. Baecker, Wolf, & Rankin, 2004; R. M. Baecker, 2003).

We conducted a study to evaluate the use of additional peer-to-peer communications channels for students in distributed lectures (Schick, 2005). For this study, an experimental version of ePresence was developed which provides text chat and audio conferencing for student peer-to-peer communication within “study groups”. The audio channel was shown to contribute to student engagement by encouraging spoken discussion and social interaction, and also by providing an ambient aural “media space” within which students feel less isolated. Students also engaged in conversations that spanned both channels, resulting in a phenomenon we call “bimodal” conversation.

Although audio conferencing contributed to student engagement in our study, it must be used with caution in a lecture setting. Our study used a traditional lecture format with few opportunities for group work. As a result, inappropriate spoken conversation occasionally occurred which was disruptive for some students. Despite this, audio conferencing can be used to enhance student engagement if used appropriately and in combination with teaching methods that take advantage of student collaboration, such as team learning (Michaelsen, 2002). Future research with ePresence will investigate the use of these alternative teaching methods for online lectures.

This paper presents an analysis of bimodal conversation, with the goal of broadening our understanding of conversational behaviour in a bimodal context. We make use of techniques from conversation analysis, which examines the allocation of conversational turns, and discourse analysis, which examines the context and interpretation of each utterance.

Conversations in text

Text chat, also known as Internet-relay chat, instant messaging, or text messaging, is classified as quasi-synchronous computer-mediated-communication (Garcia & Jacobs, 1999). It is synchronous in the sense that messages are transmitted with minimal latency between two or more users who are online and actively participating in a conversation. It is not completely synchronous because users compose their messages offline before sending them, as opposed to sending each keystroke independently. Users cannot observe the message production of their peers, which impacts how participants allocate conversational turns.

The allocation of turns in conversation is the subject of conversation analysis, or CA (Sacks, Schegloff, & Jefferson, 1974). Traditional CA applies only to spoken conversation, but the underlying methodology of CA, which seeks to understand participants' shared understanding of the rules of conversation, has been successfully applied to conversations using text chat (Garcia & Jacobs, 1999). In text chat, the mechanisms for allocating turns are altered because users do not have direct control over the sequential order of messages in the text chat window. While one user is typing a response to a message, other users may send intervening messages, so adjacency pairs may be separated. This can result in "phantom" adjacency pairs and extra conversation threads that can lead to confusion (Garcia & Jacobs, 1999). Despite these problems it has been shown that participants have little difficulty interpreting text chat conversations, and can employ standard repair strategies to resolve ambiguities when necessary (O'Neill & Martin, 2003).

Bimodal conversation

In our study, both text chat and audio conferencing were provided for peer-to-peer communication, which allowed students to engage in "bimodal" conversation. For each utterance, students could choose the mode they perceived as the most appropriate for that utterance, resulting in a conversation that "switches" between modes as it unfolds. The goal of this paper is to demonstrate how mode switches impact turn taking and turn construction. Related work in the area of mode selection at the conversation level (Murray, 1988) offers a taxonomy of factors influencing mode choice, and in future work we will investigate whether Murray's framework can be extended to include considerations of mode at the utterance level.

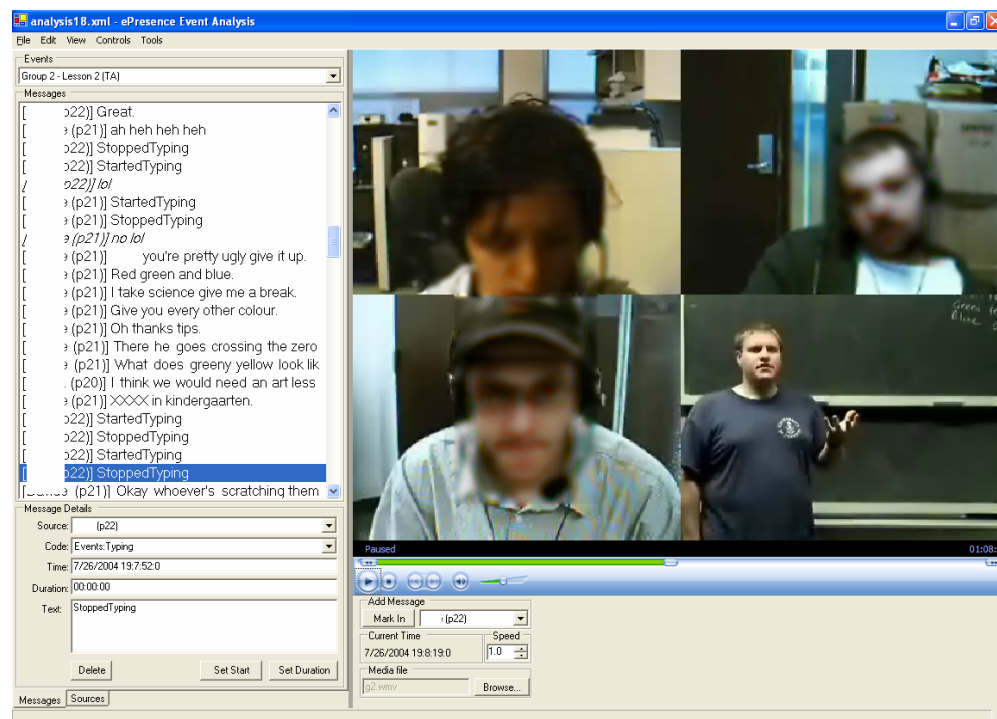


Figure 1: Bimodal playback and analysis tool

Methodology

Data for this analysis was gathered from a study investigating the impact of private student communication during online lectures. In small “study groups” of three students each, students viewed a webcast of a recorded lecture delivered to their desktop computer using the ePresence interactive webcasting system. During the lecture, students were instructed to communicate with their study group as much or as little as they desired. (Fig. 2) shows the user interface of the experimental system, which consists primarily of a large video window showing the lecturer and blackboard.

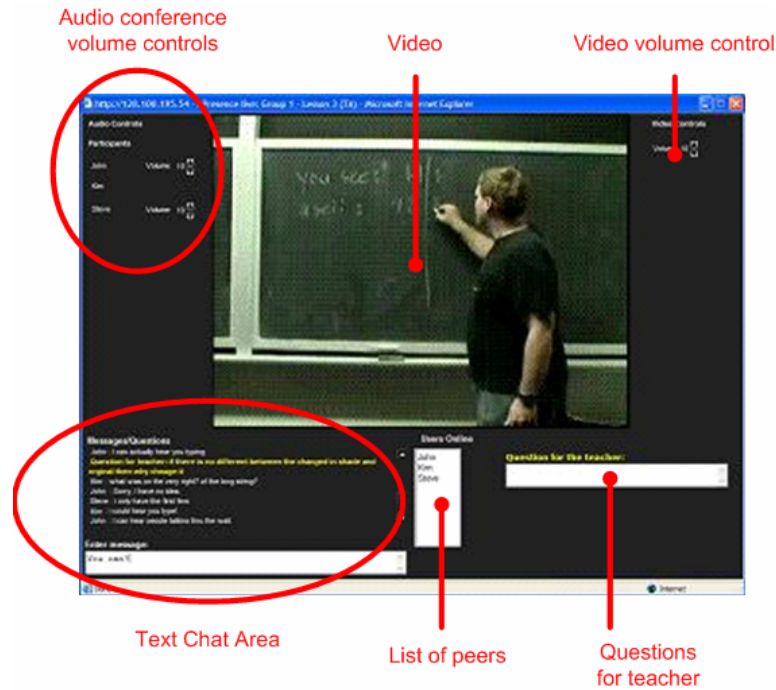


Figure 2: Experimental ePresence bimodal interface

Text messages and typing activity were automatically recorded by the ePresence system. Audio and video recordings of students were also made, and later transcribed. We developed a tool (Fig. 1) to play back audio and video in synchronization with text messages, which allowed us to make transcripts that contain both spoken and type-written utterances. We use transcription notation based on standard CA conventions (Hutchby, 1998), with separate columns for text and audio utterances. Excerpt 1 shows part of a bimodal conversation transcribed using our notation. It is a conversation between three students (with aliases Alice, Craig, and Ben) as they watch video of the lecture. This excerpt is primarily in audio, with a response in text from Craig at line 3.

Excerpt 1:

Line	Name	Text chat	Audio conference
1	Alice:		What did he say?
2	Ben:		I don't know somebody was talking to me.
3	Craig:	pixels	
4	Ben:		Her name was Alice.
5	Alice:		Oh heh heh heh pixels heh heh heh heh heh

Our analysis uses techniques of CA and discourse analysis in an attempt to understand how participants engage in bimodal conversation. We observe conversational behaviour as it appears in the transcripts and associated recordings, with the aim of identifying relevant patterns. We also analyze utterances at the linguistic level to show how changes in mode affect their construction and interpretation.

Turn construction

During the course of a bimodal conversation, participants are free to select either the text or audio mode for each utterance. Many factors may influence this choice, such as personal preference or the perceived strengths and weaknesses of each mode, such as the ability of the mode to convey emphasis or emotion. Regardless of what factors influence this choice, switching from one mode to another has consequences in terms of how utterances are constructed, due to its impact on deixis.

Deixis, or reference, is the use of deictic terms to refer to people, places and times outside the context of an utterance (Levinson, 1983). Discourse deixis refers to a previous utterance in the conversation itself, and is sensitive to mode switches because the recipient must know which previous utterance is the referent. In a bimodal environment there are several possible referents, and this introduces ambiguity that must be resolved. We observed two techniques, or “markers” of across mode reference: quotation and addressing. These markers allow the speaker to “point” from one mode to the other. Although a speaker could explicitly state that they are referring to one mode or another, we observed no examples of this. Students were able to disambiguate their across-mode deixis with the simpler tools of quotation and addressing. For example, in excerpt 2, there are two uses of across-mode deixis marked with quotation and addressing. In line 4, Alice quotes content from the text message of line 1. In line 6, Ben uses addressing to refer to Craig’s message in line 5.

Excerpt 2:

Line	Name	Text chat	Audio conference
1	Ben:	What kind of camera should I buy?	
2	Alice:		Heh heh, who // you guys are having funny questions.
3	Ben:		Oh you’re making fun of me?
4	Alice:		Yeah, the, what kind of camera should I buy.
5	Craig:	An expensive one.	
6	Ben:		Oh yeah, thanks Craig.

Excerpt 3 contains an example of both quotation and addressing within a single utterance. Ben (line 2) repeats part of the original question (line 1), and at the same time explicitly addresses the utterance to its intended recipient.

Excerpt 3:

Line	Name	Text chat	Audio conference
1	Craig:	do i have your permission?	
2	Ben:		Yes you have my permission Craig.
3	Craig:		Great.
4	Ben:		Ah heh heh heh heh.
5	Craig:	lol.	
6	Ben:	no lol.	

Adjacency pairs such as question-answer or request-response often exhibit these markers when the second pair part is in a different mode from the first. Excerpt 3, lines 1-2 is an example of a question-answer adjacency pair that spans modes and makes use of both quotation and addressing to refer to the first part. Excerpt 4 contains another example of an across-mode adjacency pair that uses addressing in the second pair part (line 2) to point to the first pair part (line 1). Lisa repeats her statement in text (line 4), showing an orientation to the fact that Adam may not have been able to interpret her spoken statement due to the mode switch.

Excerpt 4:

Line	Name	Text chat	Audio conference
1	Adam:	Do you find that (the lecturer’s) speed of speaking is slower in this lecture that his speed in the last lecture?	
2	Lisa:		Not really Adam.
3	Kim:	Kind of.	
4	Lisa:	I don’t notice much difference.	

As a counter-example, consider excerpt 5 which contains two examples of within-mode adjacency pairs, one in text (lines 1-2) and one in audio (lines 3-4). Neither of these adjacency pairs makes use of quotation or addressing to restore context since there was no mode switch between the first and second pair parts.

Excerpt 5:

Line	Name	Text chat	Audio conference
1	Dave:	I'm surprised that (the lecturer) memorized the binary code for 92, 123.	
2	Mike:	hahahah yeah	
3	Dave:		It's just a random thing I guess.
4	Mike:		Yeah.

It is important to note that markers do not *always* appear following a mode switch, but they appear in a large majority of mode switches in our data, especially in the second part of an across-mode adjacency pair. This demonstrates that participants oriented to the fact that mode switching damages discourse deixis and can lead to ambiguity. Participants avoid such ambiguity through the use of addressing and quotation, and make use of other repair strategies in the event of problems, as has been demonstrated for text chat (O'Neill & Martin, 2003).

Turn allocation

The marriage of these two “language exchange systems” (Sacks et al., 1974), each with different turn allocation mechanisms results in a hybrid set of turn allocation rules. Since utterances in each mode are observed by all participants, future actions in each mode depend upon past actions in both modes. We describe the turn taking rules as they relate to those for spoken conversation: current selects next, current continues, and next self-selects. In the case of text chat we use the terms poster/reader to refer to participants that are referred to as speaker/hearer in spoken conversation.

Current selects next

In spoken conversation, current speaker can select a speaker to take the next turn by referencing them in the first part of an adjacency pair, such as a question or a request. In bimodal conversations there are two possible mode transitions that can occur: text to audio, and audio to text. Excerpt 3 contains an example of a text to audio mode switch, from line 1 to line 2. Because the first part is in text, there is a noticeable delay before Ben responds since he needs to first notice the text message in the chat window. In the bimodal environment next speaker or next poster selection is limited to selecting a *future* speaker and/or a *future* poster, as with pure text chat conversations (Garcia & Jacobs, 1999).

Current speaker continues

A current speaker or current poster can continue if no next speaker self-selects, or by continuing without delay after a transition relevance place. Excerpt 6 show the current poster continuing in audio. Since his text comment (line 1), which contains an “emoticon”, elicits no response from the other students, he clarifies his intent in audio (line 2).

Excerpt 6:

Line	Name	Text chat	Audio conference
1	Ben:	-_- zzzzzz	
2	Ben:		Heh that's supposed to be a little face that I drew.
3	Alice:	can someone wake me when it's over???	

Excerpt 3, lines 3-5 shows an example of the current speaker continuing in text (with intervening laughter), demonstrating that current speaker/poster self-selection can span both modes.

Next speaker self-selects

Next speaker can self-select when the current speaker issues a first pair part that is unaddressed, or gives up the floor. In excerpt 7 there is a transition from text to audio. Alice's comment (line 1) is not specifically addressed to anyone, so Ben self-selects in audio by offering an empathetic remark (line 2).

Excerpt 7:

Line	Name	Text chat	Audio conference
1	Alice:	I was taking notes but stopped...	
2	Ben:		Are you serious?
3	Alice:		Yeah

Excerpt 1 contains an example of an audio to text transition with an unaddressed question. Both potential next speakers self-select, one in audio (line 2) and one in text (line 3). Notice that this allows a form of conversational parallelism similar to pure text chat, where two people can generate utterances simultaneously without interfering with each other. This type of behavior can lead to multiple threads, which are discussed in greater detail in the next section.

Repairing multiple threads

Text chat conversations have a tendency to split into multiple threads (McDaniel, Olson, & Magee, 1996; O'Neill & Martin, 2003). This is a result of the separation of message production and transmission, where participants are unable to observe turn construction and thus have a difficult time "holding the floor". Excerpt 8 shows an instance of thread confusion between two students, which arises because of the simultaneous transmission of the first two messages (lines 1-2). Mike attempts to give the floor to Dave in the audio mode, using a request for clarification (line 3). The attempt fails, and Dave takes the floor in the text mode (line 4). Mike's request for clarification is repeated in the text mode (line 6), and eventually elicits the desired response (line 8). Although this attempt at correcting multiple threads was unsuccessful, this example demonstrates that the audio channel is a possible site for repairing the thread confusion that can arise in text chat.

Excerpt 8:

Line	Name	Text chat	Audio conference
1	Dave:	Remember the joke that (the lecturer) said in 148?	
2	Mike:	I want to ask how do the binary numbers get written down on that silicon chip.	
3	Mike:		What joke?
4	Dave:	hahaha	
5	Dave:	Good question.	
6	Mike:	What joke?	
7	Mike:	Yeah why do we care?	
8	Dave:	You know the one that talks about the things we need to write and the things that we don't need to write.	

Conclusions and future work

Conversations across two modes

In the context of webcasting, there are advantages and disadvantages to both audio and text communication that suggest combining both modes will benefit students. Audio can convey ambient noise and paralinguistics, yet it risks interfering with the webcast audio if used inappropriately. Text chat avoids interference but has a looser set of turn taking rules that prevents fine control over message sequencing. They also differ in their ability to convey certain types of information or intentions. Audio can convey emphasis and emotion because of its ability to convey prosody, but text allows offline composition of messages which promotes more integrated and reflective communication. Students in our study successfully constructed conversations in both modes, choosing modes based on their perceived advantages and disadvantages in context. We feel that due to the increasing use of text and audio services in mobile and desktop environments, further study of bimodal conversational behaviour is warranted both within and outside the current distance education context.

Preference for previous mode

In bimodal conversation, although mode switches may occur, we have demonstrated that there is a preference for the previous mode due to the limitations of across-mode deixis. Mode switches result in a loss of context which impacts the ability for an utterance to implicitly reference the first part of an adjacency pair. Switching modes necessitates non-minimal utterances that require explicit referencing to repair the loss of context associated with the mode switch. Since they result in non-optimal utterances, further investigation into the reasons for mode switches is necessary.

Mode selection may be partially explained by Grice's cooperative principle, which serves as a guide for the content of contributions to a conversation (Grice, 1979). This principle consists of conversational maxims of quantity, quality, relevance, and manner. Choice of mode can be considered within this framework by considering its impact on the maxim of manner. Under this maxim fall directives such as: avoid obscurity and ambiguity, and be brief and orderly. Since mode switches may contribute to ambiguity, and require non-minimal utterances in order to restore context, avoiding mode switches is a means of adhering to cooperative principles. Further study of bimodal conversation data is necessary in order to determine if these principles do indeed play a role in mode selection.

Code switching as mode switching

In the related phenomenon of code switching (Gumperz, 1971), bilingual speakers make use of two languages, or codes. Alternating codes can be used to add emphasis or to convey additional meaning to an utterance. Similar use might be made of mode switching, and this suggests that factors influencing code choice might also influence mode choice. Our future investigation on the perception of acoustic elements will allow us to describe how such data can be used by listeners, together with other linguistic and non linguistic information, to reconstruct or clarify the meaning intended by speakers in the audio mode.

Auditory cues in bimodal conversation

Many aspects of speech and non-speech perception have been ignored in this paper. To study the role played by the audio channel in bimodal communication, we intend to investigate the process of auditory perception and describe how acoustic data can be used by listeners, together with other information, to reconstruct the meaning intended by the speaker. Knowledge about the process of auditory perception indicates that listeners and learners perceive according to probabilities which they use to generate expectancies. In other words listeners perceive according to the manner in which they have prepared their perceptual system to approach the task (Chamot, Küpper, & Impik-Hernandez, 1988; Noizet, 1980; O'Malley, Chamot, & Küpper, 1989; Sanders, 1977).

References

- Baecker, R., Wolf, P., & Rankin, K. (2004). The ePresence Interactive Webcasting and Archiving System: Technology Overview and Current Research Issues. *World Conf. on E-Learning in Corp., Govt., Health., & Higher Ed., 1*, 2532-2537.
- Baecker, R. M. (2003). A principled design for scalable internet visual communications with rich media, interactivity, and structured archives. *Proceedings of the 2003 conference of the Centre for Advanced Studies conference on Collaborative research*, Toronto, Canada, 16-29.
- Chamot, A. U., Küpper, I., & Impik-Hernandez, M. V. (1988). *A study of learning strategies in foreign language instruction: Finding of the longitudinal study*. Rosslyn, VA: Interstate Research Associate.
- Garcia, A., & Jacobs, J. (1999). The eyes of the beholder : Understanding the turn-taking system in quasi-synchronous computer mediated communication. *Research on Language and Social Interaction*, 32(4), 337-367.
- Grice, H. P. (1979). Logic and Conversation. *Communications*, 30, 7-72. from the Linguistics and Language Behavior Abstracts database.
- Gumperz, J. J. (1971). Bilingualism, Bidialectalism, and Classroom Interaction. *Language in Social Groups: Essays by John J. Gumperz* (pp. 311-339). Stanford: Stanford University Press.
- Hutchby, I. (1998). In Wooffitt R. (Ed.), *Conversation analysis : Principles, practices, and applications*. Malden, Mass.: Polity Press.
- Levinson, S. C. (1983). *Pragmatics*. New York: Cambridge University Press.
- McDaniel, S. E., Olson, G. M., & Magee, J. C. (1996). Identifying and analyzing multiple threads in computer-mediated and face-to-face conversations. *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Boston, Massachusetts, United States, 39-47.
- Michaelsen, L. K. (2002). Team Learning in Large Classes. In C. A. Stanley, & M. E. Porter (Eds.), *Engaging large classes : strategies and techniques for college faculty* (pp. 67-83). Bolton, MA: Anker.
- Murray, D. E. (1988). The Context of Oral and Written Language: A Framework for Mode and Medium Switching. *Language in Society*, 17(3), 351-373.
- Noizet, G. (1980). *De la perception à la compréhension du langage : Un modèle psycholinguistique du locuteur*. Paris: Presses Universitaires de France.
- O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension in second language acquisition. *Applied Linguistics*, 10(4), 418-437.
- O'Neill, J., & Martin, D. (2003). Text chat in action. *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, Sanibel Island, Florida, USA, 40-49.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest semantics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- Sanders, D. A. (1977). *Auditory perception of speech. an introduction to principles and problems*. Englewood Cliffs: Prentice - Hall.
- Schick, R. (2005). *A study of student conversation in text and audio during webcast lectures*. Master's thesis, University of Toronto.

Acknowledgements

This research was funded in part by a grant from Bell University Labs.