Auditory and Textual Conversational Multitasking

Theoretical Introduction and Research Hypotheses

Eli Dresner
Department of Communication, Tel Aviv University, Israel

Segev Barak

Department of Psychology, Tel Aviv University, Israel

Abstract:

Conversational multitasking—the participation in several concomitant linguistic interactions—is becoming ever more prevalent, both in purely textual contexts on-line and in hybrid situations, involving both face-to-face and technologically mediated interaction. The present study examines how linguistic modality--auditory vs. textual--affects cognitive capacity for multitasking. Results show that text supports multitasking better than voice, that both the intermingling and combination of text and voice do not improve on purely textual channels, and that textual visual indications of speakers' identities do not improve auditory multitasking capacities either. In the discussion we consider the implications, applications and limitations of these results.

This research is concerned with *conversational multitasking*: The involvement (either passively or actively) in several synchronous linguistic interactions at the same time. This practice is demonstrated, for example, by users of a variety of textual Internet-based communication channels; among them are chatrooms (such as IRC, or its web-based look-alikes), textual virtual worlds (MUDs and MOOs), and instant-messaging (IM) applications. In each of these technological contexts a stream of text lines goes through a single window or several windows on the user's computer screen, and this stream of text often includes more than one conversation thread at any given time. The competent user can keep track of these simultaneous distinct threads. Multitasking is not limited to on-screen communication, though: An intermingling of auditory and textual conversation threads, for example, can be found in the office, the meeting room, the classroom, and in a variety of social situations.

Conversational multitasking (CM) has been subject to very little research to this date. One reason for this state of affairs may have to do with its very novelty: It is less of a continuation of features extant in face-to-face (f2f) communication than other aspects of CMC (Computer

Mediated Communication), and therefore it is often viewed as a breakdown of preexisting interactional standards rather than an emerging alternative standard. Another possible reason (Koolstra et al., 2009) is that communication researchers often focus their attention on a single medium of technology, while CM often involves the concomitant use of various channels.

Be the reasons for this relative lack of attention as they may, it is clearly unjustified. Multitasking in general, and conversational multitasking in particular, is becoming ever more prevalent in technology-infused societies (Foehr, 2006; Roberts et al., 2005), and hence research into the cognitive underpinnings and various kinds of implications of this phenomenon is certainly in order. In this paper we examine some perceptual and cognitive aspects of conversational multitasking in the context of CMC.

Multitasking, media multitasking and conversational multitasking. What is multitasking? Roughly put, it is doing two or more things at the same time (Baron, 2008; Koolstra et al., 2009). The term originates from computer science, where it is used to describe a computer that performs concomitantly two or more computational tasks. Even in this technical context, though, the term “multitasking” manifests some of the ambiguity that characterizes its applications to human behavior and cognition: Many computers that have single, non-parallel processors are said to be multitasking, not in virtue of their genuinely performing multiple tasks at the same time (which they cannot do), but rather due to their switching quickly back and forth among tasks.

Turning now to the human domain, we can certainly find cases of genuine multitasking (i.e., parallel performance of tasks). At the physiological level, our body is engaged in numerous concomitant processes, and at the behavioral level as well people do several things at once—for example, walk, talk and look around them. When it comes to conscious cognitive processes that require attention, on the other hand, the picture is less clear. Notwithstanding the prevalent use of the term “divided attention,” there is accumulating evidence that, in fact, attention cannot be divided (Ruthruff et al., 2003). Therefore what are seemingly cases of parallel performance of attention-requiring cognitive tasks may actually be cases of task-switching. This paper proposes to circumvent this issue and defines multitasking as the performance of two or more tasks that occupy the same time interval. This will allow us to put aside the debate over divided attention, and focus our attention on the tasks performed.

The advent of modern communication technologies, from the telephone and radio, through television, to the networked computer and the mobile phone, has given rise to the increasing prevalence *of media multitasking* (or communicative multitasking)—the engagement with two or more communication media at the same time. As elaborated by Koolstra et al. (2009), this phenomenon has variegated manifestations: People read the paper and listen to the radio at the same time, they talk on the phone while watching TV, or listen to music while doing homework. Quite often one task is primary and the other secondary, for example, not getting (full) attention at all, or only for short time intervals. Young people—so called digital natives (Prenski, 2001)—are heavy multitaskers, in accordance with their reputation, but older people do not lag much behind them (Koolstra et al., 2009). The reasons for media multitasking also vary (Koolstra et al., 2009): Media prevalence and availability is one such reason, the cultural pressure to make better use of time is another, and the dullness of some tasks may induce conjoining them with other tasks.

It is of interest and value to relate multitasking to the phenomenon of task interruption (Cutrell et al., 2001; Czerwinski et al., 2000). Task interruption occurs when the performance of a given task is interfered with by extraneous activities or stimulations. For example, the reading of a document can be interrupted by a phone call, or the work on a spreadsheet stopped in order to answer an instant message. The difference between multitasking and task interruption seems to be this: If the interruptions are each self-standing, cognitively unrelated to one another, then we have task interruption; otherwise, when the interruptions form together a coherent and continuous task (or tasks) themselves, we have multitasking. The distinction between the two cases is vague and contextual. For example, in a study of work conditions in an investment management company, González and Mark (2004) found that tasks were performed only for a few minutes, on average, before they were interrupted by other tasks. (Interestingly, in half of the cases the disruption was initiated by the task-performer herself.) If short time intervals are considered, then this seems like (a culture of) task interruption, while if a wider perspective is taken we have multitasking—the interwoven performance of many tasks during the workday.

Finally, we consider conversational multitasking. This is a sub-category of media multitasking, where the two (or more) communicative tasks involved are conversations, as exemplified in the opening paragraph of this paper. The note made in the previous paragraph concerning the vague boundaries between multitasking and task interruption applies in the specific case of conversational multitasking: It is hard to tell when a phone conversation is merely interrupted by instant messages on the computer screen, and when the messages add up to a unified task that is performed at the same time interval. Also, the distinction between conversational and non-conversational exposure to media is not clear cut as well: In some online contexts one may be passively exposed to a dialogue in a way that is quite similar to an exposure to mass media. Notwithstanding these qualifications, there are common, typical cases of taking part in (or following) two conversation threads (or more), which are the subject matter of this study.

Cognitive and perceptual aspects of conversational multitasking. A basic observation, made, for example, by Herring (1999), is that conversational multitasking, which is quite common in textual CMC, is seldom found in auditory f2f conversation. A real-life situation somewhat similar to a reasonably active chatroom is a cocktail party, or a large dinner table. In such situations there might go on several independent conversations at the same time, and a person involved in one of them might overhear a sentence or a word in another (e.g., her name), but even a minimally prolonged juggling of two concomitant conversation threads, although it cannot be ruled out, of course, seems to be rare.

Why is multi-tasking significantly more prevalent in textual, computer-mediated environments than in f2f situations? One kind of answers to this question may appeal to social and cultural norms: With respect to f2f interaction we have well established norms that strongly discourage us to engage in two conversations at the same time. Another type of explanation is cognitive in its orientation. Here, we suggest, the notion of cognitive load should be brought to bear on the analysis of multitasking. In particular, the difference between the visual and auditory perceptual modalities in helping accommodate such load in the context of multitasking should be acknowledged and appealed to.[1] (As is elaborated in the discussion section below, we argue that there is interplay between the normative and cognitive factors that affect multitasking practices.)

Media multitasking (and, in particular, conversational multitasking) is cognitively demanding and costly: Studies show that such multitasking gives rise to lower task performance, as compared to the performance of tasks serially (Pashler, 1994). According to limited capacity theory (Lange, 2000) one explanatory theory the demand for cognitive resources in multitasking exceeds the performer's limited capacities, which results in inferior performance. In particular, it may be that the cognitive resources that are taxed to their limit in conversational multitasking are short term resources (as opposed to, for example, long term memory, which the characteristics of CM do not seem to put special pressure on); hence the conceptual framework of cognitive load theory (CLT) may be applicable in this context (Mayer & Moreno, 2003; Pass et. al., 2004; Sweller, 1998). According to CLT, short term working memory consists in an information-processing bottleneck in many cognitive tasks, and what is required for successful performance of such tasks is (a) an efficient application of prior knowledge to the absorption of new information, and (b) an efficient integration into long term memory of new information (temporarily stored in the working memory). CLT is mainly concerned with the design of tools and environments that facilitate learning of complex cognitive tasks, and thus does not apply directly to conversational multitasking. Nevertheless, there is significant affinity between what is required in CM and what successful learning demands, according to CLT (as just described). In conversational multitasking the performer continuously takes in new linguistic information (i.e., conversation turns) into her short term memory, and has to connect it to one of the various conversation threads she is attentive to—that is, she has to (a) apply to this new linguistic information prior knowledge pertaining to the conversation it belongs to, and (b) integrate it into her long term memory of this conversation.

These observations lend support to the hypothesis (made in Herring (1999) and elaborated in Dresner (2005)) that conversational multitasking is more cognitively feasible in text than in speech, and this due to the characteristics of visual perception as opposed to those of auditory perception. Vision is inherently spatial and metric: The basic sensory input of (each of) our eyes is a two dimensional space to which are applicable such notions as above and below, near and far. This spatiality is of course of paramount importance for our daily maneuvering within the physical world around us, but it is also operative in the way we read text in general, and in our dealing with synchronous, conversational text in particular. As for traditional, asynchronous written and printed text, spatial structure gives rise to (and helps us absorb) its various characteristics that McLuhan (1962) speculated about, and that later writers argued for more convincingly (Goody, 1986; Ong, 1982) – for example, its elaborate logical and semantic structure. And as for textual CMC, this very same visual spatial structure seems to help us digest multiple conversation threads: Due to the appearance of conversation threads in visually separate windows on the screen, and thanks to our ability to refer to previous lines that still appear on the screen, we are better able to accomplish the cognitive tasks mentioned in the

[1] We are grateful here to the insightful comments of the EJC reviewers.

previous paragraph. That is, we can apply prior knowledge of the conversations involved to the intake of new information and integrate the new data into long term memory.

It should be acknowledged that the spatiality of our perception of text depends on an even more basic characteristic of the written word, namely its persistence over time (as opposed to the transience of speech). It is this persistence that allows for our spatial maneuvering through text, even in the case of the fleeting lines of synchronous CMC that remain on our screens only for a few seconds. However, this paper argues that persistence alone cannot explain the phenomena pointed to above and the results described below, and that the spatiality of vision needs to be explicitly appealed to as well. Thus, for example, Braille text is as persistent as ordinary text, but it is hard to envision a blind person using it to conduct conversational multitasking in the same ease that ordinary text is commonly used for this purpose. The difference seems to be related to the distinct perceptual modalities used in the two cases.

With auditory perception, on the other hand, things are different. As just noted, sound is evanescent—a spoken word disappears the moment its utterance has been completed, and thus is not available to reexamination (Ong, 1982). Furthermore, our ears are not designed to discriminate sounds according to the direction from which they are emitted. (Such discrimination is made, with varying degrees of accuracy, at later stages of the processing of auditory input, within our brain.) These features give rise to various characteristics of the communicative applications of sound that are discussed and researched, for example in Conversation Analysis. One is the almost complete lack of overlap in conversation turns (Sacks, H. Schegloff, E. & Jefferson, G., 1974)—there is no spatial matrix in which to place concomitant auditory inputs, and hence in a typical auditory conversation there is no overlap of such inputs. And the same features render attendance to concomitant independent auditory conversation threads more difficult, and hence less plausible. For one thing, if one is engaged in several concomitant auditory conversations some overlap between auditory inputs seems highly likely, which renders the perceptual absorption (into working memory) of what is said more difficult. And second, the auditory setup offers no help of the kind described above to the integration of new turns into their conversational contexts.

Research hypotheses and questions. In this study we examine experimentally several hypotheses concerning the effects of the auditory vs. visual linguistic modalities on conversational multitasking capacities. In accord with the foregoing theoretical discussion, our goal is not to study directly the dynamics of attention-switching and attention-division that seem to be involved in conversational multitasking. (When two chat windows are open on a user's computer screen, for example, (a) attention is switched back and forth between them, but also (b) when one window is focused on, the other is peripherally monitored for changes.) Rather, we look at various conditions where two conversation threads occupy the same time interval, and assess participants' success in the conversational tasks involved. There are numerous ways to assess the adequacy of someone's participation in a conversation—among them is the quantity and quality of his contributions to the conversation, his recall of the contents of the conversation, and his ability to apply these contents in future contexts. For reasons that will be elaborated below we chose passive content recall—the ability to follow the conversation threads without taking part in them—as our adequacy measure.

In view of the considerations raised above, concerning conversational multitasking and perceptual modality, we expect multitasking to be easier in a textual context than in auditory conversation. Thus we get our first hypothesis:

H1: Two conversation threads that are presented textually, in the form of synchronously accumulating text-lines in a chat window, will be easier to follow than when they are presented auditorily.

Note that the textual setup that is to be compared with auditory conversation consists in a single text window (presented on a computer screen), in which the two conversation threads (each involving two participants) unfold intermingled with each other. This is as opposed a scenario in which each conversation thread appears in a distinct window. The reason for this choice is that in Dresner & Barak (2006) it is shown that separate-window textual presentation is better than a single-window presentation as a vehicle for conversational multitasking. Therefore if indeed H1 will be empirically supported—i.e., the single-window presentation will be proved to be superior to auditory presentation in enhancing multitasking capacities—then separate-window presentation will thereby be proved to be superior to auditory presentation as well.

Having thus compared in H1 textual and spoken linguistic interaction from the perspective of multitasking, we may also inquire whether the combination of these two modalities together accommodates multitasking better than each of them on its own. If indeed perceptual capacities place limitations on our conversational multitasking abilities, then it is plausible that a multimodal exposure to language will make multitasking easier than when text or spoken language is presented alone. This prediction receives some indirect support from several studies of multimodal interfaces (Oviatt, 1997; Oviatt, 1999; Oviatt et al., 2000), in which it is shown that people prefer to access computer systems multimodally, especially when faced with a high cognitive load. (That is, when both speech and writing devices are available as input channels both will be used, and quite often simultaneously.) These results, which are concerned with active choices of computer users, might be applicable also to the case considered here, of multimodal exposure to language. Thus our second hypothesis is a follows:

H2: Two conversation threads that are presented both textually and auditorily—that is, read as text and heard as spoken language at the same time—will be easier to follow than when they are presented only as text or only as spoken language.

Put together H1 and H2 lead us to expect a hierarchy in which spoken language alone is lowest and combined text and speech are highest.

Having thus considered written and spoken language independently and in combination, we take a natural step further and ask to what degree the two modalities support multitasking in alteration. If one of two concomitant conversation threads is presented textually and the other auditorily, will it be easier to follow them than when both are textual (or both auditory)? There are grounds opposing expectations regarding the answer to this question. On the one hand, if perceptual separation (for example, visual separation) enhances multitasking capacities, then the separation of the two conversation lines into two distinct perceptual modalities should allow for easier attention-allocation to each one of them, and thus improve overall understanding. (Research into cognitive load suggests that task performance is improved when incoming new information is divided between the visual and auditory modalities (Mayer & Moreno, 2003). However, in the learning setups studied by cognitive load researchers, linguistically coached information is usually restricted to the auditory channel, while the visual modality is reserved mainly for pictorial or graphical information. Therefore it is not clear whether the results in CLT-induced research apply to the case at hand, when the information in both perceptual modalities is purely linguistic. On the other hand, it may be the case that receiving intertwined linguistic input from two perceptual sources is more difficult than concentrating on one such source, and that therefore inter-modal multitasking may yield outcomes that are inferior to all previously discussed cases. Because of these conflicting intuitions we refrain here from making a hypothesis, and suffice ourselves with raising the following research question:

RQ1: If one of two intermingled conversation threads is presented textually and the other auditorily, will overall understanding be superior to single-modality multitasking cases?

See Table 1 for a summary of experimental conditions and research hypotheses/questions for the first experiment (H1 through RQ1).

Finally, we turn to a set of considerations and resulting hypotheses of a more advanced nature. If indeed text proves to be a superior vehicle for multitasking than spoken language, in accord with H1, then this may be due to several contributing factors. One such factor, already considered in the foregoing theoretical discussion, is that in typical synchronous written communication settings lines of text remain on the computer screen for a short period of time, and therefore newly appearing lines can be related to older ones and thus more easily connected to their thread and better understood. (In auditory conversation this is obviously impossible.) This hypothesis has been given support in Dresner & Barak (2009), where the beneficial effect of window-separation on multitasking was shown to be positively correlated with the amount of text present at each given moment on the computer screen: When conversation threads are presented in separate windows, old lines in each window belong to the same thread as new lines that appear in the window, and therefore the more old lines there are, the more useful window separation proved to be.

An additional reason for the advantage of text over spoken language may be that in textual synchronous conversation, lines are associated with the people who pronounce them in an explicit way, namely, through the appearance of the name of the (textual) speaker right before what he says. In conversations that are accessed only through hearing we identify the speaker only on the basis of recognizing her voice, which may require more cognitive resources—resources that are stretched to their limit anyway in the cognitively demanding situation of multitasking. Thus it may be hypothesized that a visual aid for speaker-recognition may enhance multitasking capacities in the auditory context. One kind of such aid is obvious: Auditory conversations could be turned into audio-visual ones. Another, related, possibility is to associate spoken linguistic turns with constant, unmoving pictures, graphic avatars, or simply printed names. We chose the last of these options, in order to retain as much continuity in the experimental setup as possible, and thus avoid the inadvertent introduction of new unknown variables. Thus we get the following hypothesis:

H3: Two conversation threads that are presented textually auditorilywill be easier to follow when each speaker's name appears on the screen in the beginning of his linguistic turn than without this visual aid.

Furthermore, the said textual aids can be used not only for speaker recognition, but also for thread separation. That is, the names appearing on the computer screen during the spoken conversations can be graphically grouped so as to assist the separation of the auditory input into threads. Thus we get our final hypothesis:

H4: Two auditorily presented conversation threads that are accompanied by a textual speaker identification that visually separates conversation threads will be easier to follow than similar such auditory threads accompanied by a textual speaker identification that is not so separated.

Table 2 summarizes the experimental conditions and research hypotheses for the second experiment (H3 through H4).

A computerized chat-simulation program and voice recording of these chats were used for testing the hypotheses experimentally. Two independent conversation threads were concomitantly presented to each participant. To test H1, H2 and RQ1, in Experiment 1 the two dialogues were presented in one of four conditions: textually; auditorily; both textually and auditorily; and one thread presented textually and the other auditorily (modality split). To test H3 and H4, in Experiment 2 the two dialogues were presented auditorily as in Experiment 1, while the name of the speaker was presented on the computer screen, either intermingled in one window, or spatially separated (in two windows, each indicating one of the two threads). A control group received no name-presentation. After being exposed to the dialogues participants completed a multiple-choice test, assessing their recognition of factual details that were mentioned during the conversations. (It should be acknowledged that the task the participants were faced with—namely, passive recall—is artificial, in that it is not accompanied by active participation in (at least one of) the conversations. See the discussion below for the rationale behind this choice and an analysis of its limitations.)

Our hypotheses predicted that in Experiment 1, participants will show gradually declining performance when presented with combination of text and voice presentation, then text presentation, then voice presentation (H1 and H2). This can be represented as follows:

← Increasing performance   Decreasing performance →

text and voice combined > text alone > voice alone

In Experiment 2, we predicted that participants will show gradually declining performance when the auditory dialogues will be accompanied by visual presentation of names of the speakers spatially separated into two windows, then intermingled in one window, then with no name presentation. This can be represented as follows:

← Increasing performance   Decreasing performance →

speaker names in two separate windows > speaker names in one window > no speaker names

Method

Participants

For **Experiment 1**, participants included 102 undergraduate and graduate students (69 females, 33 males).

For **Experiment 2**, participants included 92 undergraduate and graduate students (75 females, 17 males).

Overall, the participants were 18-36 years old (mean age 23.24), and were all native Hebrew speakers.

Apparatus

Computerized chat simulation. The hypotheses were tested experimentally through the use of a computerized chat-simulation program. chat-simulation program (Bright-Aqua Technologies LTD., Israel), that can be found on-line at http://spirit.tau.ac.il/comTwo 20-lines long Hebrew conversation threads were presented to all participants. In one of them two persons (a female and a male) discuss football teams and food preferences, and in the other two persons (again a female and a male) discuss Israeli politics. Both dialogues were fabricated, and are not claimed to be representative of typical chat conversations; the cognitive phenomena that are being examined in this study are of a general nature, and therefore it was not required that actual chat conversation be exactly mimicked or reproduced. (This is not to say, though, that we do not acknowledge the effects of content and context on multitasking; see the discussion below.)

The lines of text accumulate in typical chat mode—i.e., each new line begins with the name of its author (followed by a semicolon), and appears below the previous line; when the window becomes full new lines push older ones out of sight.

In **Experiment 1**, the dialogues were presented in four different conditions:

Textual presentation: The conversation threads were displayed on the computer screen, intertwined within a single window (450 pixels high and 400 pixels wide – 20 text lines long). A new line of text was added every 8 seconds. The window closed 8 seconds after the conversations ended.

1. *Auditory presentation*: The conversation threads were played auditorily using headphones, intertwined as in the textual presentation. The lines of the four different "persons" on the conversations were read by four different voices (two males and two females, according to the gender associated with the names in the textual presentation). The pace of the auditory presentation was synchronized with that of the textual presentation.

2. *Textual+auditory presentation:* The conversation threads were concomitantly presented both on the earphone and on the computer screen as in the auditory and textual presentations described above.
3. *Modality split*: The conversation threads were split according to modality, so that one conversation was textually displayed on the computer screen, whereas the other was auditorily played by earphones. The textually and auditorily presented threads were counterbalanced between participants, so that some participants were presented with the first thread auditorily and with the second thread textually, and other participants received the opposite presentation. There was no significant difference between the scores of these two participants' groups. The textual presentation was in a single window (450 pixels high and 400 pixels wide – 20 text lines long). A new line of text was added every 16 seconds, to allow the auditorily played lines to be played 8 seconds after the presentation of a new text line. Thus, the interval between auditory lines was also 16 seconds. The window closed 8 seconds after the conversations ended.

In **Experiment 2**, the dialogues were presented auditorily with visual presentation in three different conditions:

1. *One window with names*: This condition was identical to the third condition of Experiment 1, except that instead of presenting the lines of the conversations on the screen, only the name of the speaker was displayed.
2. *Two windows with names*: As in the previous condition, the auditory presentation was accompanied by the name of the speaker displayed on the screen, except that the two threads were spatially separated, each presented in its own window (each window was 225 pixels high and 400 pixels wide – 10 text lines long). A new speaker name was therefore added every 8 seconds on the left or right window, alternatively. The location of each of the two conversations on the screen (i.e. i.e., left or right) was counterbalanced; there was no significant difference in the participants' scores between left and right location). The windows closed 8 seconds after the conversations ended.
3. *One window control:* This condition was identical to the first condition (in the present experiment), except that each time a line was auditorily played, the presentation on the screen was "XXXXX" instead of name of the speaker.

Thus, in all the experimental conditions in both experiments, the duration of presentation was the same (approximately 350 seconds). The total number of names presented at any given time was identical, whether one or two windows was used.

Multiple-choice test. The participants' ability to follow the conversations was measured using a computer-displayed multiple-choice test that was completed immediately after the conversations ended. The multiple-choice examination was the same for all conditions, and consisted of 20 questions, 10 regarding each conversation thread (grouped together under distinct headers—first 10 questions about one dialogue and then 10 questions about the other). Each question had 5 answers, 4 of which were distracters and one was the correct answer. All questions were factual, concerned with specific details that were mentioned during the conversations (e.g., "Who was taking Efrat to watch soccer? a) Her brother b) her uncle c)

Her father d) Her grandfather d) none of the above"; "According to Michal, what action of the following should Israel make? a) Retreat from Hebron b) Retreat from the west bank c) Build a fence between Israel and the west bank d) Establish a Palestinian state e) All of the above"). The internal reliability of this test (Cronbach's alpha), was 0.99 and 0.86 in Experiments 1 and 2, respectively. Total test scores ranged from 30% to 95% correct answers with a mean of 66.13% (SD=15.84) in Experiment 1 and from 25% to 90% correct answers with a mean of 62.39% (SD=14.93) in Experiment 2.

Procedure

Participants were seated by the experimenter in front of a computer screen, where the instructions for the experiment were presented. If the experimental condition included auditory presentation, the participant was instructed to wear earphones. The instructions preceding each condition described what the participant should expect. When the participants completed reading the instructions he/she pressed a button, after which the window(s) with the textual conversations that they were instructed to follow appeared on the screen and/or the conversations were played auditorily through the earphones. Each participant was randomly assigned one of the experimental conditions (n=25-27 and n= 30-32 per each group in Experiments 1 and 2, respectively). The order of the conditions was random. The participants were passive observers/listeners of the textual/auditory conversations – they could not take part in them.

When the conversations ended the chat-window(s) closed, and then a message appeared, prompting the participants to fill the multiple-choice test examining their understanding of the textual conversations they viewed. After completing the multiple-choice test, participants completed general information forms, including age, gender and experience with computers and Internet chat.

Results

Experiment 1: Effects of modality of presentation on recall of two conversations

Table 3 presents the means and standard deviations of the proportion of correct answers of participants presented with *textual*, *textual and auditory*, *auditory* and *modality split* presentations. As can be seen, in agreement with H1 but in contrast to H2, there was no significant difference between the *textual presentation* and *textual and auditory presentation* conditions, but participants in these two conditions performed better than participants that were presented with *auditory presentation*. In addition, the *modality split presentation* also did not differ from the textual or textual+auditory presentation, but yielded higher scores than *auditory presentation*. One-way ANOVA yielded significant main effect of presentation condition [F(3,98)=6.82; p=0.0003]. Post-hoc comparisons (Fisher PLSD) showed significant differences between the *auditory presentation* condition and the three other conditions: *textual presentation (p=0.0052;* Cohen's d=2.31*)*, *textual and auditory presentation (p<0.0001;* Cohen's d=1.28*)* and *modality split presentation (p=0.0017;* Cohen's d=0.89*)*.

Experiment 2: Effects of presentation of speakers' names on recall of two auditory conversations

Table 4 presents the means and standard deviations of the proportion of correct answers of participants presented with the *one window with names*, *two windows with names*, or *one window control* presentations. As can be seen, there was no difference between the three conditions. This was supported by one-way ANOVA [F(2,89)<1].

Proportion of correct answers did not correlate with level of computer or internet use, nor with chat or instant-messaging software (IMS) experience, in either of the two experiments.

Discussion

The confirmation of H1 by our findings provides empirical support to the wider theoretical framework in the context of which this hypothesis was raised. As may be recalled from the introduction to this paper, according to our theoretical framework multitasking should be cognitively more feasible in textual contexts than in auditory conversation, and this is due to the former medium’s visually perceived spatial structure. However, the claim that conversational multitasking is more feasible textually than auditorily is not presented here as ruling out other grounds for the relative prevalence of multitasking in textual CMC. Indeed, in Dresner and Barak (2009) we suggest that there is interplay between the cognitive and normative levels: Because multitasking in purely auditory contexts seems perceptually problematic conversational norms have been developed that all but rule it out. In on-line textual communication, on the other hand, multitasking can be carried out much more easily, and therefore engaging in it is seldom perceived to be offensive. The support given to H1 in this paper coheres with these considerations and supports them.

A reservation that can be raised against the use of recall-measurement in support of H1 is that improved recall may characterize reading (as opposed to listening) in general, and not only in the context of multitasking. If this is indeed the case, how can it be said that *multitasking* was shown to be more feasible in the textual context? What the data show is only that you remember what you read better than what you hear.[2] Our response to this reservation is as follows. As suggested in the theoretical discussion in the first section, some of the advantages of visually perceived text over auditory language may be of a wide scope, but this does not undermine their applicability (and the interest in this applicability) in the context of multitasking. As may be recalled, our interest here was in the cognitive outcomes of attending several threads of conversation that occupy the same time interval—in particular, in the effects of perceptual modality on these cognitive outcomes. Recall is one such significant outcome, and hence it is of value and interest to examine whether and how it is affected by change in language modality. This having been said, the reservation points to a viable and interesting direction for future research. The difference in recall between the textual and auditory modalities may be measured when a single conversational task is involved, and also when two are. The comparison of these two measures can tell us whether the beneficial effect of the

[2] We are grateful to the referee of the EJC for drawing our attention to this possible reservation.

textual modality on content recall is enhanced in the context of multitasking, or whether it is similar to what is found in the single-task case. We propose to pursue this line of research in the future.

An interesting middle ground between textual and auditory multitasking, which our findings only touch upon, involves conversational multitasking situations in which one interaction thread is auditory and the other textual. The auditory interaction could be either face-to-face, or, for example, conducted through the telephone, and the concomitant textual conversation may be conducted on the computer screen, through a hand-held computing device such as a tablet PC or cell phone (SMS messages). Situations involving multitasking of this kind become ever more prevalent in technologically infused societies, and questions regarding both cognitive capacities and social norms should be raised with respect to interactions of this kind as well.

Our own results take a first small step in addressing this interesting set of hybrid cases. RQ1 is concerned with a hybrid situation, involving an auditory thread and a textual one. Our results show that indeed multitasking in a hybrid auditory-textual situation is superior to the purely auditory case. This finding is in accord with the fact (already noted in the introduction) that bi-modal multitasking is indeed prevalent in situations combining face-to-face auditory interaction and mediated textual communication. It should be noted, though, that according to this study bi-modal communication accommodates multitasking approximately to the same degree as a single-window textual interaction. As already noted above, it is shown in Dresner and Barak (2006) that this kind of textual setup is inferior (with respect to multitasking) to other textual setups, where conversation threads are separated spatially (i.e., presented in different windows) or through color. Thus our current findings place bi-modal multitasking (that involves auditory and textual threads) above auditory multitasking and below optimal textual multitasking. Of course, much more work is needed in order to verify whether this result does indeed apply in general to the hybrid cases considered above. It is possible (and maybe plausible), for example, that the juxtaposition of a textual thread with an audio-*visual* one will yield better multitasking results (in our experiment text was intermingled with pure audio). Also, immersion in a face-to-face situation (as opposed to attendance to it through computerized mediation) may make a substantial difference. It seems both theoretically interesting and practically important to pursue these research directions.

Apart from the normative issues considered above, our results raise a series of questions with respect to the role of perception in multitasking situations in particular, and in the intake of language in general. Thus the refutation of H2 shows that no enhancement of multitasking capacities arises from modality redundancy, for example, concomitant exposure to identical linguistic input in text and voice. One possible explanation of this result is based on the finding that control over the encoding phase of a memory task helps alleviate the cognitive load that arises from multitasking (Craik, Naveh-Benjamin, Ishaik, & Anderson, 2000). It may be hypothesized that control is responsible also for the advantage of text over aural language as a vehicle for conversational multitasking: Text allows for such control, in virtue of its persistence on the screen and our ability to maneuver around it, while the transient nature of spoken language implies that we have little control over its intake. The addition of aural language to the text does not enhance control, and therefore no improvement in recall was observed.

An altogether different explanation of the failure of H2 may be based on the specifics of the auditory input stream that was added to the textual one. No conscious efforts were made (by the actors reading the conversations) to use the additional aural modality in order to stress some aspects of the text (e.g., through intonation, or shouting), and thereby improve its intake. It is possible that if voice is used this way to support text bi-modal presentation will indeed improve upon a single, textual channel.

Finally, we turn to H3 and H4, which were not confirmed. Visual textual cues indicating whose conversational turn it is did not prove beneficial, nor did similar cues indicating also which conversation thread is currently active. The question, in this case as well, is why? One possible answer is that the theoretical premises that motivated these two hypotheses are wrong. That is, it might very plausibly be the case that the visual indication of speakers' identities does not make any contribution to the advantage of text over voice as a vehicle for conversational multitasking. In this case this advantage is explained by the possibility of content-matching that is available in text and not in voice—as described in the introduction to this paper and supported in previous experiments (Dresner & Barak , 2009)—as well as by other factors that still need to be unearthed.

Alternatively, it may be the case that the specific way in which the identities of the speakers were indicated in our experiments was detrimental to their function. Note that in the said experiments, as is the case in most contemporary chat interfaces, identity is designated textually, i.e., *linguistically*. This is as opposed to face-to-face interaction, where we identify the speaker visually, but through non-linguistic means such as face recognition, bodily traits and clothes. (In both face-to-face interactions and our experiments (as well as in telephone conversations), voice is another factor that helps us identify the speaker. The question raised here, however, is whether and how this auditory identification can be supported through visual means.) It can therefore be hypothesized that non-linguistic visual aids to speaker- and thread-identification may prove to be more beneficial for multitasking purposes. Such aids may take the form of simple graphic non-textual markers (i.e., avatars) on the one hand, or the presentation of complete audio-visual conversations on the other hand. An empirical examination of these hypotheses seems to be of both theoretical and practical interest, in that it may lead to better understanding of the interplay between linguistic and non-linguistic perception on the one hand and possibly to improvement in multitasking interfaces on the other.

The various limitations of this study should be acknowledged. As acknowledged above, and argued, for example, in Bentley (1997), recall is only one among many measures that could be used in order to assess effective listening (in our case – to a textual dialogue), and arguably not the best one. In Dresner & Barak (2006) the choice of this measure is supported through an appeal to the difficulty of the cognitive tasks involved and our aim to avoid a flooring effect, but certainly other measures should be appealed to and utilized. In particular, multitasking capabilities should certainly be tested in a context where participants take active part in at least one of the conversation threads—as already acknowledged in the introduction, such involvement seems to characterize most actual textual interactions. Also, the interaction between content type and multitasking capacities is certainly not denied here—cognitive and affective aspects of content, for example, may affect multitasking, and we hope to be able to examine how they do so in the future. Finally, our appeal to the spatial aspects of visual perception in order to motivate our research and explain its findings should be eventually supported through eye-movement tracking experiments (Rayner 1994), through which the details of perceptual maneuvering during conversational multitasking can be uncovered.

Thus we conclude that this research settles a few interesting questions and opens up many others, in a domain of inquiry that brings together considerations from a variety of fields. A better understanding of the phenomenon of multitasking has arguably been demonstrated to be of interest and importance.

References

Baron, N. (2008). **Always on**. Oxford: Oxford University Press.

Bentley, S. (1997). Benchmarking listening behaviors: Is effective listening what the speaker says it is? **International Journal of Listening**, **11**, 51-68.

Craik, F., Neveh Benjamin, M., Ishaik, G., & Anderson, N. (2000). Divided attention during encoding and retrieval: Differential control effects? **Journal of Experimental Psychology: Learning, Memory and Cognition**, **26** (6), 1744-1749.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In M. Hirose (Ed.), **Human-Computer Interaction: INTERACT ’01** (pp. 263–269). Amsterdam: IOS Press.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging and interruption: Influence of task type on performance. In C. Paris, N. Ozcan, S. Howard, & S. Lu (Eds.), **Proceedings of OZCHI 2000** (pp. 356–361). Sydney, Australia: Academic Press

Dresner, E. (2005). The topology of auditory and visual perception, linguistic communication, and interactive written discourse. **Language @ Internet** 2/2005. Retrieved November 10, 2009 from http://www.languageatinternet.de/articles/161/

Dresner, E. & Barak, S. (2006). conversational multi-tasking in interactive written discourse as a communication competence. **Communication Reports, 19**, 70-78.

Dresner, E. & Barak, S. (2009). Effects of visual spatial structure on textual conversational multitasking. **Communication Quarterly, 57**, 104-115.

Foehr, U. G. (2006). **Media multitasking among American youth: Prevalence, predictors and pairings.** Kaiser Family Foundation. Retrieved July 10, 2010 from http://www.kff.org/entmedia/7592.cfm

González, V., & Mark, G. (2004). "Constant, constant, multi-tasking craziness": Managing multiple working spheres. In E. Dykstra-Erickson & M. Tscheligi (Eds.), **Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems** (pp. 113–120). New York: ACM Press.

Goody, J. (1986). **The logic of writing and the organization of society**. Cambridge: Cambridge University Press.

Herring, S. (1999). Interactional coherence in CMC. **Journal of Computer Mediated Communication**, **4**. Retrieved December 10, 2009 from http://jcmc.indiana.edu/vol4/issue4/herring.html

Koolstra, C., Ritterfeld, U. & Vorderer, P. (2009). Media choice despite of multitasking? In T. Hartmann (Ed.), **Media choice: a theoretical and empirical overview** (pp. 234-246). London: Routledge.

Lang, A. (2000). The limited capacity model of mediated message processing. **Journal of Communication, 50**(1), 46-70.

Mayer, R. and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. **Educational Psychologist**, **38**, 43-52.

McLuhan, M. (1962). **The Gutenberg Galaxy**. Toronto: Toronto University Press.

Ong, W. (1982). **Orality and Literacy**. London: Methuen.

Oviatt, S. (1997). Multimodal interactive maps: Designing for human performance. **Human Computer Interaction**, **12**, 93-129.

Oviatt, S. (1999). Mutual disambiguation of recognition errors in a multimodal architecture. **ACM SIGCHI Conference on Human Factors in Computer Systems.** Pittsburgh, PA: ACM Press, 576-583.

Oviatt, S., Coulston, R. & Lunsford, R. (2000). When do we interact multimodally? Cognitive load and multimodal communication patterns. **Proceedings of IEEE International Conference on Multimodal Interfaces**. Pittsburgh, PA: ACM Press, 129-136.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. **Psychological Bulletin, 116**(2), 220-244.

Pass, F., Renkel, A. & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. **Instructional Science**, **32**, 1-8.

Prensky, M. (2001). Digital natives, digital immigrants. **On the Horizon, 9**(5), 1-6.

Rayner, K. and Pollatsek, A. (1994). **The Psychology of Reading**. Hillsdale, NJ: Laurence Erlbaum Associates.

Roberts, D. F., Foehr, U. G., & Rideout, V. (2005). **Generation M: Media in the lives of 8-18 year-olds.** Kaiser Family Foundation. Retrieved July 10, 2010 from http://www.kff.org/entmedia/7251.cfm

Ruthruff, E., Pashler, H. E., & Hazeltine, E. (2003). Dual-task interference with equal task emphasis: Graded capacity sharing or central postponement? **Perception & Psychophysics, 65**(5), 801-816.

Sacks, H. Schegloff, E. and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking in conversation. **Language, 50**, 696-735.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. **Cognitive Science**, **12**, 257-285.

Tables

Table 1. Summary of experimental setups and research hypotheses/questions for experiment 1.

Experimental condition	Condition description
1. Textual presentation	Two temporally intermingled textual conversation threads are presented in a single chat window on a computer screen.
2. Auditory presentation	Two temporally intermingled spoken conversation threads are heard by the participant through earphones.
3. Textual and auditory presentation	Two temporally intermingled conversation threads are concomitantly presented to the participant both visually (as text in a chat window on a computer screen) and auditorily (as speech heard through earphones).
4. Modality split	Two temporally intermingled conversation threads are presented to the participant—one textually, in a chat window on a computer screen, and the other auditorily, through earphones.

Hypothesis 1: Condition 1 (textual presentation) yields better recall than condition 2 (auditory presentation).

Hypothesis 2: Condition 3 (textual and auditory presentation) yields better recall than both condition 1 (textual presentation) and condition 2 (auditory presentation).

Research question 1: Will condition 4 (modality split) yield better or worse recall than condition 1 (textual presentation) and condition 2 (auditory presentation)?

Table 2. Summary of experimental setups and research hypotheses for experiment 2.

Experimental condition	Condition Description
1. One window with names	Two temporally intermingled spoken conversation threads are heard by the participant through earphones, while a synchronized presentation of the current speaker's name at each moment appears in a single chat window on a computer screen.
2. Two windows with names	Two temporally intermingled spoken conversation threads are heard by the participant through earphones, while a synchronized presentation of the current speaker's name at each moment appears in two chat windows on a computer screen, each window for one of the conversation threads.
3. One window control	Two temporally intermingled spoken conversation threads are heard by the participant through earphones, while a synchronized presentation of non-informative visual input ('xxxxx' strings) appears in a single chat window on a computer screen.

Hypothesis 3: Condition 1 (one window with names) yields better recall than condition 3 (one window control).

Hypothesis 4: Condition 2 (two windows with names) yields better recall than condition 1 (one window with names).

Table 3. Effects of modality of presentation on recall of two conversations. Means and standard deviations of the proportion of correct answers of participants presented with *textual, textual and auditory*, *auditory* and *modality split* presentations.

Experimental condition	Textual presentation	Textual and auditory presentation	Auditory presentation	Modality split presentation
Mean (standard deviation)	67.12 (17.50)	73.54 (10.48)	55.40 (17.38)	68.52 (11.67)

Table 4. Effects of presentation of speakers' names on recall of two auditory conversations. Means and standard deviations of the proportion of correct answers of participants presented with the *one window with names*, *two windows with names*, or *one window control* presentations.

Experimental condition	One window with names	Two windows with names	One window control
Mean (standard deviation)	61.10 (15.23)	62.67 (17.00)	63.50 (12.61)