Persson, Per (2003) ExMS: an Animated and Avatar-based Messaging System for Expressive Peer Communication, Proceedings of *GROUP 2003 conference*, Nov. 9-12, 2003.

ExMS: an Animated and Avatar-based Messaging System for Expressive Peer Communication

Per Persson
Nokia Research Center
Itämerenkatu 11-13, 00180 Helsinki, Finland
per.persson@nokia.com
(Work reported here was partly conducted while at Swedish Institute of Computer Science, Kista, Sweden)

ABSTRACT
While many *synchronous* computer-mediated communication systems have failed to encourage users to make use of the expressive capabilities of their avatars, *asynchronous* systems may hold better chance. This paper reports on the design and user study of a message system that allows users to concatenate and annotate avatar animations and send them to peers. During three weeks, a group of 11 17-year-olds exchanged 222 animated messages in their everyday life environment. The interplay between text and animation allowed users to create significantly expressive messages. Many messages told micro-stories about fictitious and real events. Users identified with their avatars and were proud of their embodied representation. The content of messages deepened during the course of the study.

General Terms
Design, Human Factors.

Keywords
Animation, computer-mediated communication, multi-media authoring tools, avatars, expressiveness

1. INTRODUCTION
In face-to-face situations, verbal and non-verbal communication provide a rich multi-modal environment. Language is there, of course, but also body, face, hands, voice, gestures, and gazing of the speaker. Our species and its cultures have developed sophisticated systems of how to express and interpret non-verbal behavior [[1], [14]]. Prosodic features such as pitch (level, range and variability) loudness and tempo can carry expressive value. Facial expressions and gestures – including gazing behavior - are primary gateways to what a speaker means, feels, senses or perceives, especially in short range. Bodily movements, gestures and postures are equally expressive. Hands are also important in pointing, giving layouts of objects, emphasize aspects of spoken discourse, or indicate relationships between discourse segments (e.g. "On the other hand....").

Although rich, face-to-face communication is restricted to time and space. Speaker and listener must cohabit the same space and time. Quite a few computer-mediated communication systems (CMC systems) have distributed communication over space and time, while trying to maintain the communicatory richness and "embodiedness" of face-to-face communication.

GROUP'03, November 9–12, 2003, Sanibel Island, Florida, USA.

1.1 Embodied Computer-Mediated Communication
The most natural solution perhaps and also most heavily researched is video systems in which a video image is broadcast together with sound or text [[9]]. Although these systems are technically challenging and may involve slight usability problems (such as confusing gaze directions), the major reason why business conferencing is the only sucessful application is probably found elsewhere: true physical appearance is not always desirable in conversations with others. For one, it does not provide anonymity, which could be preferred if talking with non-acquaintances [[18]]. Second, some users may prefer to use body proxies to communicate with others, since it allows them to be more expressive, different and imaginative.

The use of such proxy representations of the user – so called *avatars* – in the digital realm has gradually been introduced in various systems. In the textual worlds of MUDs (Multi-user dimensions) [[5], [6]] the user is represented with a name label. The basic way in which user could add non-verbal communication to their verbal expressions was to describe them in words ("kissing", "waving"). But MUDs also introduced the first image-like avatar representation. Emoticons – alphanumeric symbols that in certain sequences and layout receive a certain iconic and image-like structure - were used to indicate a reaction, stance or gesture by the user to things happening in the text field [[15]]. This practice migrated into e-mail communication and today it is widely used, for instance using smileys to mark irony and "just kidding" [[13]]. Today, several e-mail and chat clients substitute such typed smileys with simple graphically simple and colored emoticons (e.g., www.incredimail.com).

Visual chats such as www.palace.com and V-chat [[17]] provide some expanded support for embodiment since users are somewhat free to choose or create the image representing the avatar. In terms of embodied *behavior*, however, these environments provide little support: some spatial proxemics [[10]] and the standard palette of 5-10 expressions.

Collaborative Virtual Environments (CVE) and Immersive Virtual Worlds provide 3D worlds, often with a much more elaborate avatar, capable of facial expressions, actions, gestures in addition to movement (e.g., DIVE [[7]], MASSIVE [[8]], and www.activeworlds.com). These kinds of system have been used for collaborative work, chat and on-line games. (Some of these systems allow a single user to interact with synthetic and artificial agents. Although research within interface agents and gaming has

investigated embodied behavior, I will disregard them in the current study. My focus is human-to-human communication.)

1.2 Drawbacks of Synchronous Communication

However, as Salem and Earle [[16]] point out,

"Despite several years as a research topic, avatars in existing collaborative systems play little role in the communication that takes place. Communication is achieved almost entirely on the exchange of text or in some cases voice messages, while in the real world communication is far richer than this."

This claim is confirmed by a large-scale user-study of Microsoft's V-chat [[17]]. With seven different gestures at their disposal, the approximately 35 000 users visiting the chat space over a period of 119 days, used in average only one gesture every two minute.

There are probably several causes for this discouraging result. For instance, in many of these environments typing text normally implies a halt to the control of the avatar. Also, some of these systems provide graphically poor avatars unable to display rich facial and gestured behavior [[16]]. In addition, unlike textual communication, most embodied computer-mediated communication system do not maintain a history (with the exception of [[11]]). If a user missed a gesture or behavior of other avatars in the CVE or chat, there is no way to recover it and replay it. Assuming that users interleave other activities with the chat (other applications, eating, reading newspaper, taking care of children), it is fair to assume that substantial amount to embodied behavior passes unnoticed. Subsequently, if there is nobody watching there is no real reason to make use of the embodied capabilities of the chat.

There are probably also more deep going reasons for the absence of embodied communication in these systems. The keyboard/mouse interface to control the avatar is so radically different to the "interface" we have vis-à-vis our own bodies or even puppets, marionettes and dolls. Controlling the behavior of the avatar probably consume major parts of the user's attention, leaving the content of the communication behind. Since these environments are *synchronous* communication environments - calling for immediate response if you have been spoken to - the cognitively and tactile challenging avatar control mechanism probably create a "conversational stress". In is likely that most users would choose to drop the avatar body control in this situation.

Of course, much avatar research has been devoted to finding support for users in conversational stress. One proposed solution is that the system takes over some responsibility for the control of the avatar. In this way, the user need not micro-managing every aspect of animating the human figure, but can focus on the conversation itself. For instance, on the basis of how humans initiate and terminates conversations, take turns and display involuntary reactions, *BodyChat* automatically chooses appropriate behaviors in such situations [[3]]. Another example is *Comic Chat*, which on the basis of the text entry automatically creates comic panel compositions, what avatars to include in such panels, the placing of word balloons and emotional expression [[11]].[1] Although automation relives the user of micro-managing, it also makes the user loose control of the avatar. If the avatar does or displays things to others "behind the user's back", the user might feel disconnected and socially uncomfortable. As the artificial 'intelligence' in automation systems seldom is able to handle situated and often widely shifting goals and communication needs, automation may be experienced as frustrating.

Another solution to the conversational stress problem is to allow real time manipulation of avatar limbs and face using the participant's real body connected to sensors [[12]] or some other equipment [[2]]. Normally, however, such equipment is expensive and rare for a great majority of users.

Another reason why CVE and visual chat users discard embodied communication could be because most communication in those systems takes place between strangers and non-acquaintances. Non-verbal communication needs context and situation to be meaningful and properly understood. Groups of peers and good friends that share memories and know each other's preference, taste and habits, are able to refer to and play upon such shared experiences. The shared context can be taken for granted in communication between peers, which means that very subtle communication means (words or gestures) can gain significant amount of expressiveness.

This contrasts sharply with the situation in most chats and CVE: communication takes place between people who never met in real life and do not share background knowledge of each other. Of course, the system administrators try to engage users in activities that might provide such a context: making or buying artifacts, building houses, or making career and family (e.g. the recently released *SimsOnline*[2]). The question is, however, if this is enough to encourage users to engage in embodied communication.

2. EXMS – AN AVATAR-BASED MESSAGING SYSTEM

By shifting focus from *synchronous* to *asynchronous* communication modes, our ExMS system proposed a somewhat different approach to the above-mentioned problems. Although we strongly believed in the value of embodied CMC, we decided to make a *messaging* system, rather than a chat or CVE. Since an *asynchronous* message system stores and then forwards messages when recipient requests for them, it does not call for immediate reply. This means that there is ample of time to compose, ornament, and edit a response before sending it. This will decrease the conversational stress, and thereby leaving time to play around with embodied behavior attached to a verbal message. The absence of conversational stress also means that recipient can re-play an avatar-based message multiple times in order to understand its meaning and expression. In addition, a sender can review his or her composed message before sending it. In contrast to a CVE or graphical chat, the sender can thus better control and ponder the social consequences of a particular expression or gesture in the message.

[1] While these solutions aim to smoothen conversational turns and determine the placement and layout of comic panels and speech balloons, they do not deal with the expressive and emotional gestures of the avatar.

[2] www.ea.com/eagames/official/thesimsonline/

Equally important was to explore how embodied CMC was picked up and used by peers (not strangers) and how their shared context played a part in their embodied CMC. In this respect, a messaging system also made more sense (not many tight group of peers synchronously gather in CVEs or visual chats in a natural way).

Despite the shift of attention to asynchronous communication, ExMS still aimed to retain the notion of avatar. Although there would be no digital space in which the avatar "lived" or moved (other than in messages), we wanted to study how users developed a relationship and identified with one single and unique character during longitudinal usage. "Taking pride" in the uniqueness of one's character vis-à-vis others', we argued, could act as a motivator for using the system. This differentiates ExMS from many other character-based messaging services on the web, in which characters are easily exchangeable (e.g., www.fjallfil.com or *Moviemaker* at http://mymovie.sierraclub.org/mm).

Our two avatar animation designers were given some requirements. At least one avatar should be known by the general public (it turned out to be Calvin in *Calvin & Hobbes*). There should be equal distribution of gender among the avatars, and possibly some genderless figure. We encouraged them to work with both animals and humanoids and design for a target group of 16-17 year olds. These open requirements enabled the animation designers to tap into their own creativity (both were professional illustrators). We ended up with 6 avatars in various styles and personality (fig 1).

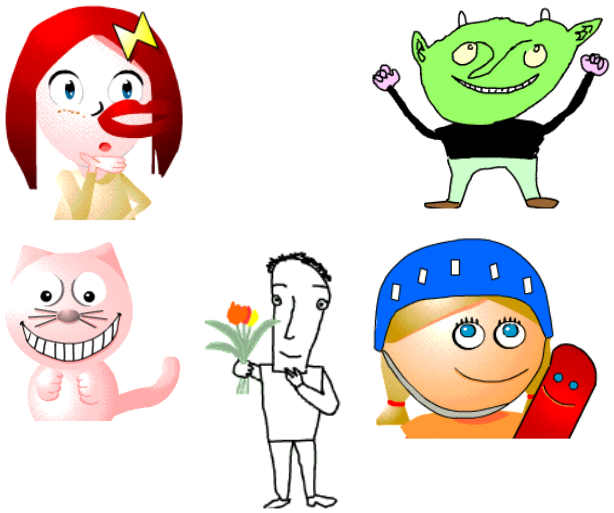

Figure 1. From topleft to bottom right: Manga, Monster, Cat, Mr. Y. and Skate Girl. For copyright reasons, Calvin is excluded. If this figure does not appear in color, check http://www.nokia.com/exms/ for correct color scheme.

More important was the design of the animation (or clip) set connected to each avatar. In contrast to animation libraries of CVE and visual chat system, the ExMS system did not focus on how avatars could move through space, interact with objects or smoothen conversation. Instead, *expressiveness* was emphasized. This meant the ability to express *feelings* (positive and negative, extrovert and introvert), *reactions* (calm and aggressive) and *moods* (ups and downs). We provided the animation designers with a list of 21 expressions that tried to capture all of these aspects with various strengths. In addition, the animation designers were encouraged to muster up approximately five "surreal", "freaked out" or "bizarre" behaviors, not necessarily connected to expressions of feelings. The purpose of varying the expressive tone of the behavior in the clips was to promote a multifaceted personality of each avatar. Although some avatars, such as the Cat or Calvin made a 'cute' first impression, they contained aggressive and even violent behaviors. Vice versa, although the Monster at first looked scary and evil, he had cute and nice behaviors too (e.g., "blushing"). With a varied set of behaviors we hoped that users would have approximately the same kind of "expressive potential" in the avatar-based messaging.

Since face is important conveyor of expressiveness and screen space was limited, the style of 'big head and small body' had to be employed, although the body also conveyed expressiveness. We also instructed our animators that characters could move towards and away-from "the camera", promoting close-ups as well as long-shots. Each clip, however, had to be one self-contained behavioral unit (no equivalents of "camera cuts" were allowed inside the clips).

We ended up with 27-32 clips for each avatar (available at http://www.nokia.com/exms/images.html). Although some of the expressions were similar across avatars, many differed due to the style and surface appearance of the avatar. (The Cat for instance had a very broad mouth, making his smile particularly strong.) Also, the bizarre clips were unique for each avatar (e.g. Mr. Y. pulling out a hat and a wand, and then pulling out a woman from the hat). Through trial and error, clips ended up being between 1 and 4 seconds.

It is important to point out that the expressions in the clip were animated, not only static images. With animation we could capture expressions that would have been difficult or even impossible with still imagery. With gestures and behavior, ExMS animations provided another layer of expressiveness than systems with only still imagery (such as *Comic Chat*).

The author of this paper labeled the clips and inserted them into a clip library for each avatar (see below). In order to avoid that the labels influenced the users' understanding of the behavior shown in the clip, we tried to keep labels as 'behavioristic' as possible. For instance, rather than 'sad' and 'happy' labels referred to 'tears' and 'laughing'.

2.1 The ExMS Composition Tool and Player

In terms of emotional depth and expressive richness, each ExMS avatar was now equipped with an animation library unprecedented by any messaging application avatar on the market. The next step was to design a composition tool (and a player) that allowed users to verbal and embodied messages in a simple way. This was the basic idea: since we had made sure that all clips within each avatar started and finished at the same frame layout, user could potentially create small micro-movies with continuous movement by sequencing those clips one after another.

We had a number of requirements on the composition tool. First, the user should be able to review and edit the message before sending (see discussion above). Second, the tool must give the support to the user to synchronize verbal and non-verbal behavior. Identified by Salem & Earle [[16]] as one of the most critical issues in the design of avatar systems, gluing an embodied

expression to its correct word or paragraph had to be easy both for sender and receiver (cf. the exact placement of smileys in e-mails). Since sound output was technically challenging, verbal behavior instead was represented in text form, contained in speech balloons above the avatar's head (Fig. 4). This layout allowed for synchronized *juxtaposition of text and behavior*, in contrast to web tools like *Movie cards* (http://vykort.passagen.se/ -> Filmkort). We hoped that users could exploit this interplay between graphics and text, for, e.g., conveying irony or stance towards the verbal content.

While emphasizing the expressiveness, another requirement was to minimize effort in the composition process. The purpose of ExMS was to minimize message composition time to 1 to 4 minutes and *still allowing the user to express herself in rich and unexpected ways*. Of course, the clip library provided the backbone for expressiveness, but it was also important to ensure

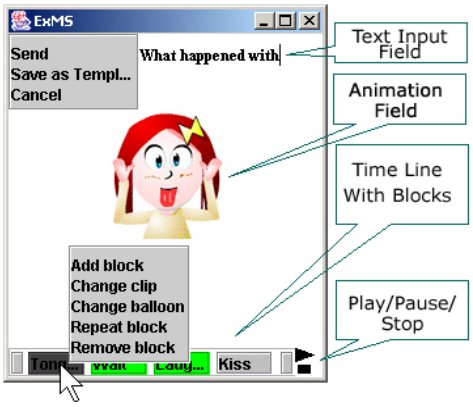

Figure 2. Composition interface with pop-up menu.

simple interaction with the composition tool itself (in contrast to tools such as http://www.3dme.com/, Director or Flash). In addition, we wanted it to be a "walk-up-and-use system" with no complex set-ups or configuration.

Based on these requirements, the composing interface was kept simple (Figure 2). It consisted of a time line at the button, an animation field in the middle and text input field at the top. When starting to create a new message, the time line would be empty with the exception of an entry and exit clip, displaying the avatar appearing and disappearing behind a stage curtain to give the impression of a staged performance. In between these clips the user could then add as many clips as he or she wanted, determining the length of the whole micro movie. In figure 2, the user has already added four clips (tongue, wait, laughter and kiss).

Clips were contained within "blocks". A block could contain one clip and a snippet of text. Although the size of the blocks in the time line of Figure 2 gives the impression that they are of equal length, their temporal duration were actually determined by the clip inside it. By containing both text and clip into the same block, there was only need for one "track" in the timeline (in contrast to multi-track editing tools such as *Adobe Premiere*).

Most interactions would be conducted via a popup menu, which appeared when a block was marked (Figure 2). Via this menu, users could add a new block to the time line (ending up to the right of the block marked), but also change the clip inside the block. Choosing "change clip" from the popup menu, the user would be brought to the clip library (Figure 3). Here, each clip was represented by a thumbnail and a label. When a clip had been chosen, the user would return to the composing interface. The clip had now been inserted in the block, and its label is written on the block (e.g., "tongue"). When a block was marked, text could be added right into the text input field.

By adding blocks, texts and clips the user was allowed to create a sequence of clips. Reviewing the message (Figure 4) meant that the application consecutively played the contents of each block (displaying the text meanwhile each clip is played). Since this was also the way in which the recipient would see the ExMS message, sender had full control of message appearance, tempo and text-image synchronization. When composing was completed, the composer was asked to enter text into a subject field (like e-mail). Then a recipient would be chosen from a list. For future reference, an inbox and outbox stored all traffic.

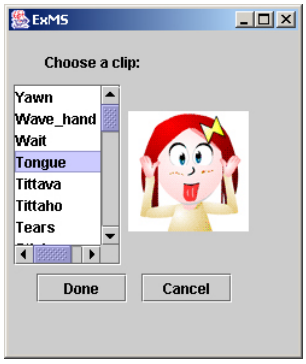

Figure 3. Clip Library.

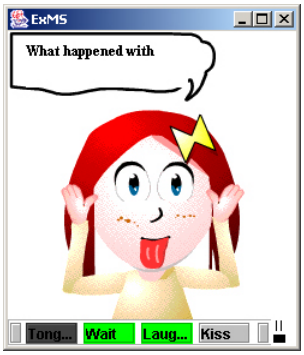

Figure 4. Review/play.

3. THE STUDY - METHOD AND PROCEDURE

Since ExMS was aimed at peer messaging, the requirements for choosing user groups were relatively strict. Each group had to be socially close and already engaged in in-group messaging. Whereas much research in embodied communication has been devoted to non-peers (e.g. [[17]]), no study to this date has investigated socially tight groups such as this one.

2 groups of 5 and 6 17-year olds from a school in Stockholm were enrolled by a teacher acquainted with the author of this paper. Since the system only had 6 unique avatars, each group used ExMS separately and independently, allowing users to send messages only within the group. Group 1 (G1) consisted of four females and two males. They attended the same class and most had been friends for some years back. G2 consisted of 3 females and 2 males. There was a love couple in the group. In addition, a male and a female, who were childhood friends, had just broken up a 3-months relationship (still being good friends).

First, the author met separately with the groups informing them about the test procedure. Each participant was given a pre-questionnaire about existing messaging habits. It encompassed both mobile and PC-based messaging tools and asked about their messaging habits, the most common purpose for messaging, message content categories and to whom they send most messages. At the end of the meeting, one clip from each avatar was presented. On the basis of this, each participant, individually

and hidden from view, wrote down first-hand and second-hand choice of avatar. The turnout was fortunate. All G1 members received their first-hand option, except one subject who got her second-hand choice. In G2, all were provided with their first option, except a female who could not attend the voting process. She was assigned an avatar. Since G2 only had five members, the Monster avatar was not used here. Before this balloting, users were informed that shifting avatar during the study was not possible.

Then, the ExMS client was installed on the central computer system of the school. In this way, the client was connected with their login name and could thus be accessed from any school terminal. We also provided encouragement and clear instructions of how to install the client at home. In addition, we provided a SMS gateway service to their normal mobile phones: whenever there was a new ExMS message waiting for them, they received a notification on their mobile phone. All participants owned and used a mobile phone.

Without providing any further directions or help, we asked subjects to use the system for three weeks in May 2002. At the end of this period, subjects were summoned to an individual 45-minute videotaped interview, together with the client and the messages they had composed. At the end of this meeting, subjects were given 3 movie tickets in return for their participation.

4. RESULTS

One subject, Cat 1[3], did not make use of the system allegedly due to lack of interest in computers. She was removed from the study. The 10 remaining subjects sent 222 messages with G1 on 110 and G2 on 112. Manga 2 scored highest with 32 messages and Skate Girl 2 lowest with 10. Number of messages oscillated between 0 and 20 per day, with the exception of a boom around May 14-15 (with 81 and 45 messages respectively). G1 seemed to have centered their usage to these days, while G2's activity was more evenly distributed over the test period. Although ExMS allowed multi-recipient messaging, no subject took advantage of it. It is possible that users strongly connected ExMS to SMS, which is associated with 'one message–one recipient' practice. 3 subjects - all in G2 - managed to install at home. Only 17 messages were sent off school hours (after 6 PM and weekends) suggesting that most usage was connected to school.

4.1 Communicatory functions

What content was ExMS used for? What *communicatory functions* or *acts* were performed with this avatar-based messaging system?

Of the 222 messages, 15 contained only the entry and exit clip, by default provided by the system (see above). These messages seem to have been sent by mistake or for test purposes. The remaining 207 messages contained in total 703 blocks (excluding exits and entries), making the average to 3,4 blocks per message. Depending on clips, the average duration of a message could thus roughly be estimated to around 15-20 seconds (from start to end title). Mr. Y 1 composed the longest ExMS containing 17 blocks that played for nearly a minute (see video example 1 at http://www.nokia.com/exms/clips.html).[4]

In contrast to short text snippets, the length of ExMS messages made it problematic to specify communicatory acts since each message performed a number of them. Cat 2's message to Manga 2, for instance, asks how recipient is doing ('hey baby, how are you?'), it expresses feeling ('I am tired and have no energy to work anymore'), it coordinates future events ('if we go to McDonalds, why don't you come?'), it shows off Cat 2's avatar (licking the floor and laughs at her own strange behavior) and the final kiss clip expresses love for her best friend (see video example 2 at http://www.nokia.com/exms/clips.html). Thus, instead of categorizing ExMS messages, communicatory acts were categorized and counted. Each message could contain one or several communicatory acts, and one communicatory act could be performed across one or several blocks. In the whole message material 406 *communicatory acts* were identified (by the author) distributed over 6 basic *functions*:

Request. (1%) Dealt with asking the recipient a favor (e.g., borrowing money).

Coordination of everyday life. (3%) Related to appointments and planning activities in the nearby future (a couple of hours). One example was Calvin 1 asking Mr. Y 1 to accompany him to the hairdresser after school.

What's up. (6%) Finding out how a person is or what s/he does.

Test. (10%) Contained little or no expressive content. Only seemed to be motivated by checking the system, e.g. to test if the SMS notification service worked.

"Real" Expressions. (46%). The most common communicatory act was related to expressing emotion, attitude or opinion in relation to some phenomena. This included several subcategories. 'I-love-you' messages (8% of all communicatory acts) expressed friendship or love vis-à-vis the recipient. Some messages (2%) expressed support and comfort to the recipient, e.g., feelings of sympathy for a female friend in menstrual pains. Others questioned, scolded or argued with the recipient (3%). Manga 2, for instance, sent an angry message to her boyfriend - Mr. Y 2 - starting with a love-you kiss, but then rebuking him for taking the credit for school work she had done (see video example 3 at http://www.nokia.com/exms/clips.html). As she revealed in the interviews:

Manga 2: It was in English class and I wrote a whole project and he did not do anything and then he wanted to be in my group and we let him and then he took my work and handed it in and took all the credit for it.

Later that day, Mr. Y 2 sent an apology (attached to the video example 3). 'Real expressions' also included illustrations of drowsiness, hunger or general mood (5%). The largest portion, however, were attitudes and opinions about past, present and future events (28% of all communicatory acts). This included *excitement* for an upcoming football game, *enjoyment* of the fact that the school beat another school in a sport competition, *hope* that sender could date the recipient tonight, *fear* that he would be rejected, *opinions* about the ExMS application and its avatars,

[3] In the following, subjects are identified with their avatar name and group number.

[4] The expressiveness of moving imagery is notoriously difficult to capture in paper format. To judge the value of the system and this publication, the video examples at http://www.nokia.com/exms/clips.html is strongly recommended.

anger over an assault in school that occurred during the test period, *apologizing* for throwing a flower in the face of the (allergic) recipient, etc.

In all 'real-expression' communicatory acts, the avatar's behavior could be said to convey or illustrate the sender's attitudes and feelings.

***Fictional* expressions and behaviors (34%).** Much usage of ExMS, however, engaged the avatars in *fictional* expressions and behaviors. Goofing around, acting out, pretending and playing with identities were central activities in these sequences. (The distinction between real and fictive expressions was difficult to make only by inspecting the messages. However, the interviews, with a few exceptions, clarified the context and its fictional status.)

This included showing off one's avatar and its capabilities to the others in the group. In Mr. Y 2's 17-block message, for instance (video example 1), the text balloon basically describes what the avatar does or feels. Even though the sender speaks in first person (now I got angry, now I got tired etc.) it is clearly just 'an act' and excuse to put the animations on display. It does not represent the real feelings of the sender. Another fictive usage was invented by Manga 1, whose avatar appeared on the stage 'singing' songs by Infinite Mass and other artists.

In the beginning of the evaluation period, staged performances like these were basically self-contained, but as time passed these expressions started to get incorporated into everyday life phenomena. For instance, a couple of messages related the avatars' appearance with real people. Skate Girl 2 displayed her most ugly face to Cat 2 and argued it looked like one of their teachers (video example 4 at http://www.nokia.com/exms/clips.html). Although still within the fictional realm, the messages were more closely tied to the users' everyday world.

Another prevalent practice involved the scheme of first insulting the recipient and then finish off with a 'just kidding', thereby fictionalizing the insult. In a message to Skate Girl 2 (female), Calvin 1 (male) first exclaims that he intends to 'slap her face' accompanied with an aggressive animation. In the next moment, however, the avatar says 'no, I love you my little fish' winking innocently (video example 5). In fact, fictional invectives like these were quite common among both male and female senders. These "just-kidding"-twists probably tapped into an already existing communication practice among these 17 year olds, moving between reality and fiction in order to check social boundaries and moral consequences of verbal and physical behavior. Thus, for some users, ExMS provided a fictional world in which they could enact things they fantasized of doing, but did not dare to in the real world. As expressed by Cat 2 (female):

Q: Did you get to know your Cat avatar?
A: A lot of things corresponded to my personality. I never break things when I get angry but I would like to be able to. If this cat had a clip displaying throwing a vase to the floor, I would use it.

As was disclosed in the interviews, ExMS was on some occasions used locally with sender and recipient sitting next each other on two different computers. To see the recipient's reaction seems to have been gratifying for the sender. Such practices surely influenced the communication content.

In conclusion, ExMS seemed to have been used for light communication and playful matters, sometimes related to the real attitude of the users but also purely fictional expressions and stories. In many of these functions, the image-based and temporal features of ExMS seem to have triggered new forms of expressions, impossible to accomplish in text only. One reason for the lack of "serious" and "useful" messaging may be related to the latency time for messages. Although the SMS notification service worked, senders complained that they could not be sure if the recipient was close to a computer to check the ExMS message immediately. For this reason, ExMS tended not to be used for important matters that needed prompt attention. Cat 2 put like this:

Cat 2: If you are sad for real, you want to show it directly. Now, not tomorrow. If I was angry with [Calvin 2] I called or sent a normal SMS.

4.2 Clips and avatars

How did users perceive of the clip library and the design of the avatars provided by ExMS? Were the clips expressive enough?

It is difficult to make general conclusions about the most popular type of expression or clip type, since clips were so intimately tied to the graphical design of the avatar. It is hazardous to equate a smile from the Cat with a smile from Calvin. In order to get some feeling of the usefulness of clips, however, each user's messages were analyzed. For each user, all clip labels were collected and sorted according to number of times it had been used. The number of clip labels depended on how many messages and blocks the user had sent. For each user, we then generated a top-5 list of the most popular clips (for that user). The top 5 lists of all users were then merged into one, listing in total 50 labels (for all 10 users). Although this list contained labels from different avatars, many of them had the same name since we had allocated similar terms to them when we incorporated the clips into the system. Wait, for instance, could be found across many avatars. In this long list, we simply counted the occurrences of labels (independent of their graphical instantiation). If some labels were different, but the underlying clip seemed to perform similar behavior, they were merged into the same clip type. For instance, while most avatars had flirt, the Monster contained a similar clip called blink.

The most popular label was smile with 6 occurrences, followed by kiss (5) giggle (4) and idle (4). Jump, laughter, tears, show teeth, anger/aggressive and flirt/blink all received 3 points (with other labels on two and one).

How to interpret this data? First, the high popularity for idle may be due to the fact that this clip was automatically inserted when the user added a block. For the other most popular clip types, warm and friendly expressions such as smile, giggle and kiss seem to dominate. Only tears, anger/aggressive - and to some extent also show teeth - have a colder and more negative tone. Although the occurrence of irony and joking make general conclusions unreliable, this suggests that ExMS messaging involved primarily positive atmosphere. This concurs with the study on V-chat in which the friendly and positive gestures (silly, waves, flirt and smile) by far outweighed (81%) the conflictual and non-committal gestures (shrug, sad, angry) [[17]].[5]

For the most *unpopular* clips, a similar analysis was performed. For each user, the labels of all *unused* clips were collected and inserted into a long list. This list had 68 entries, making the average number of unused clips 6,8 per user. (With libraries containing between 27 and 32 clip types, in average 25% of the

[5] Of the 7 gestures provided in V-Chat most popular was silly and waves, followed by flirts, smiles, angry, shrugs and sad.

clip types stayed unused.) The long list of non-used clips was analyzed in the same way as the popularity list. The least used clip types were looking left (8) and looking right (7). These were animations in which the avatars turned their bodies and looked off-frame, away from the camera/spectator. Apologize/regret received 5 points, followed by ehhh? and gazing (all having 4 points).

The small usage of clips looking off frame seems to indicate that ExMS was perceived as a face-to-face communication medium. Addressing the spectator/recipient was important. More interesting is the low usage of apologize/regret and ehhh? (The latter clip type was found in most avatars and displayed a shrugging and bewildered gesture with arms and shoulders). We can only speculate into the reasons for this. Given that ExMS was used for play and staged performances between peers, often with some fictive component, it is possible that timid and unsure behaviors were not needed to the same degree as extrovert and spectacular ones.

Beside the clip analysis, the interviews provided valuable feedback on the issue of clips. Although most subjects liked their libraries and found them varied enough for their messaging needs, a couple of users reported getting tired of some clips. Repeatedly deploying them in messages made some clips overused and predictable (e.g., smile). These users suggested the possibility of altering or personalizing them to update their freshness. Interestingly, however, this tediousness only related to users' *own* avatar, not to others'.

Cat 2: I had seen this clip on my own [avatar], but when [Manga 2] sent messages it never struck me that I had seen her clips before. Clips were only tedious for myself, not when I received messages from someone else.

The (non-) popularity analysis performed above presupposes that clips were similar across avatars. Each avatar, however, had several clips that were unique, often performing some kind of surreal or bizarre behavior (see above). Thus, although these clips did not show up in the statistics, their importance was revealed in the interviews. Cat 2, for instance, claimed that she always tried to find a reason to insert her dancing clip in ExMS messages, since the Cat was the only avatar with this kind of behavior. Taking pride in one's avatar seems to have been prevalent among all users. This is reflected in the messages content, often promoting/showing off the expressive abilities of the avatar of the sender at the expense of the recipient's avatar.

The relative success of clip libraries and avatar design was also reflected in the fact that no user expressed an urge to swap avatar during the study. Although there was general consensus on which avatar was "the best" (Monster in G1 and Cat in G2), all users explained their satisfaction with their avatar choice. In addition, when asked if they during the test period felt the need to have several avatars (e.g., for different recipients or situations), nearly all subjects gave negative replies. In terms of expressiveness, the variety provided by the clip libraries seemed to be enough. In addition, such practices would, according to the interviewees, confuse the reception process and destroy the clear and often creative associations people made between the sender and his/her avatar. It seems like the idea of unique avatars, proved to be appropriate for this user group.

4.3 Composing with time, space and text

With this broad picture of ExMS usage and clips we can now shift attention to the expressive techniques used by subjects. How did balloon texts and clips interplay to generate meaning? And how was the linearity and synchronization feature so central in ExMS made use of?

First, the subject field of ExMS messages provided little guidance in interpreting the content of the image. Surprisingly, few users wrote anything sensible here. Typical examples included "=)", "hahaha", "LALALALA" and "djkfgs". This suggests that users approached ExMS with metaphors of SMS and chat, systems in which subject lines are not available.

Instead, text in the talk balloons provided important guides to the messages. Over 95 % of all 703 blocks contained text, suggesting that text was an important support for both composers and recipients in creating meaningful communication. The most common way to use text balloons was to explain or clarify how the behavior or gesture of the clip should be interpreted, e.g., "I'm angry" in connection with an angry face, "I am really tired" juxtaposed with a yawning avatar (see video example 1). On several occasions, the text stretched the interpretation of a behavior in direction we had not expected. Cat 2, for instance, associated tears-behavior with both "fear" and "pride". Manga 2, on the other hand used tears to express "envy". Skate Girl 2 used the red face of an animation we labeled blush to indicate the summer heat that prevailed in school that day. Monster 1 used blush-behavior to suggest anger. In all of these cases, the text explained how the image "should be" interpreted. Sometimes this explanatory function of the talk balloon was employed for humorous purposes. For instance, in one message Cat 2 winked her eye in an animation called 'flirt' and added "I have something in my eye, hahaha".

Other texts were less explicit. The before-mentioned reprimand of Manga 2 to her boyfriend Mr. Y 2, placed a clip with a wagging finger called 'shame on you' with the text: "I did all the work and you got all the credit" (video example 3). Here is no explicit text label describing the emotion of avatar/sender. Instead, the recipient needs to make inferences on the basis of behavior, text and the shared context in which this message was sent. This was perhaps the most common usage of text balloons.

Although the text often acted as a specifier, "nailing down" the meaning of the open-ended image track, the opposite movement also took place. When the text was ambiguous and open, the image helped the recipient in interpreting the nuances of the message. A text-balloon from Mr. Y 2 saying, "are you not coming? Do you need help?" could be expressing benevolence and helpfulness. Juxtaposed with the somewhat impatient Mr. Y clip waiting, however, its tone became much more negative and almost accusing (video example 6).

In addition, subjects exploited the juxtaposition between text and image in creative ways. Sound effects were translated into text and inserted in the balloon, for instance "uhhuu" linked to a clip called tears. Verbal puns were used on several occasions. In an ExMS to his girlfriend (Manga 2), Mr. Y hoped that she would accept his invitation for a date. Since the Swedish word for jumping and hoping is basically the same (*hoppas*), Mr. Y 2 concluded his message with a jumping clip juxtaposed with this word.

In spite of the fact that ExMS provided a sophisticated image track, smileys were still used in the text balloons (31 smileys were found). In addition, a couple of users continued to use conventions of text-based chat rooms. Mr. Y 2, for instance, sent a message to his girlfriend Manga 2 including the text "come and hOOOld me" and then added "*waiting*" (video example 7). The

fact that the image track redundantly depicted the activity described in the balloon (the avatar standing inactively) was not troubling for the user. As he later explained in the interview, "it presents image and text at the same time.....the text describes what is happening". Since all subjects were heavy chatters and SMS users, it is not surprising to find smileys and text-chat practices in ExMS. As Skate Girl 2 explained, "writing that smiley was probably a pure reflex or something". With greater exposure to image-based messaging, such 'reflexes' may disappear, taking real advantage of the expressive capabilities of imagery.

Temporality

ExMS was not, however, only blocks of text and image, but also a temporal medium, sequencing blocks over time. How did users make use of this linearity of the medium?

The "just-kidding" messages described above rely to a great extent on the temporal capabilities of ExMS: *first* insult and *then* "just-kidding". This twist is accomplishable in a text format, but probably becomes more effective in a "player" medium that dictatorially determines the pace of reception.

Another strategy was to initiate a message with something puzzling or not totally clear and then allow upcoming images to explain what was just shown. Again the message of Mr. Y 2 to his girlfriend can serve as an example (video example 6). The first clip just shows an angry face without accompanying text. The recipient is in no position to determine the reason for this. Next block, however, shows a waiting Mr. Y saying, "are you not coming? Do you need help?" Now the recipient understands the anger of the avatar/sender. This kind of temporal technique is quite common in visual fictions in cinema and TV [[3]].

The linearity of ExMS, also allowed users to shift tone and atmosphere of the message. Replying to a request from Mr. Y 1 to send more ExMS messages, Calvin 1 tells him that "well, I am too tired to send anything" (yawning), "by the way, my nose is running..." (sour), "because you threw a dandelion in my face!!" (kicking towards the spectator). In this message (video example 8), the sender makes a turn from the relatively still and subdued small talk in the two initial blocks, to a violent accusation in the finale. Such twists are really only possible with a varied clip library in addition to a time-based, linear medium.

In addition, subjects marked beginnings and endings of messages. Although we provided titles of start and end as well as an entry/exit clip, users added "hellos" and "goodbyes" and "kiss". 37 messages (17%) started with some greeting in the balloon, while 57 messages (26%) contained some form of verbal goodbye, often expressed by "kiss" or "hug". (This contrasts with the results from V-Chat in which 23% of all posted messages had some form of greetings and 4% some form of goodbye in the dialogue text [[17]].) Of these messages, 18 contained both (8%).

These verbal start and endings were, however, unevenly distributed. Some users employed them extensively, other not at all. Most verbal goodbyes were accompanied with clips displaying the avatar kissing (a clip included in all avatars). Very seldom, however, was the avatar allowed to perform this goodbye kiss without the assistance of the text. Again, text was important, and text-image redundancy did not seem to pose a problem.

As to beginnings and endings, surprisingly enough some users added entry/exit clips in addition to the ones the system by default provided for each message. This practice doubled the clips, generating a jump-cut effect. This suggests two things. First, senders did not seem to review their messages before sending, otherwise they would have noticed these defects. All the same, the fact that entry/exit clips were utilized indicates that such markers of beginning and end are important. Just like social everyday situations are initiated with "Hi" and concluded with "See-ya", ExMS performances seem to need the same framework indicating the establishment and closure of a situation.

In conclusion, both the linearity of ExMS and the interplay between clips and balloons provided important ways of creating and understanding the meaning of messages. The synchronization of text and embodied behavior, enabled by the time-line and block-based structure of ExMS, provided good support for these activities.

4.4 Balance between expressiveness and workload

Although ExMS seemed to support expressiveness, how much work did user have to put down to create these messages?

When asked to show us how they typically created an ExMS message, a surprisingly coherent composing pattern emerged. 7 users started and finished with one block, filled it up with clip and text before they added a new block and filled that one etc. The rest of the subjects tended to start the composition process by adding three to four empty blocks directly to get the basic structure of the message, after which they started to edit the contents of the blocks. For each block, 9 out of 10 user picked the clip *before* the text. This means that the creation of ExMS messages seems to have been guided by the content of the image track and the clip library, rather than the verbal content.

Few users seem to have employed the other functions in the composition pop-up menu (change balloon, remove block and repeat block). Subjects also stated that little time was spent on reviewing and re-editing messages. When asked for time estimation for composition on average per message, interviewees reported between 1 and 3 minutes. A couple of users experienced this to be too long and involving too much work. No one stated that his or her composition technique had changed during the course of the test period.

In all, the composition process seemed to have been a quite straightforward process. The relative non-importance of reviewing seems to indicate, again, that ExMS was used for light communication, not serious matters in which the control of exact expressions may have been more pivotal. The composition process seems to have been a reasonable trade-off between effort and expressiveness. This was also confirmed in another section of the interview. Here we asked if users ever felt the need to personalize their avatars with photos of themselves, props, different clothes, skins and colors. This also included the ability to change the backdrop and environment in which the avatar's performance took place (e.g., school, home, city and café). All agreed these would be cool features, but few imagined having the energy and timing to fiddle with their avatar more than the present system allowed them to do.

The interface was reported to be easy to handle and learn (although graphically a bit "boring"). Thus, the basic structure of a simple time line sequencing blocks filled with clips and text seems to have balanced workload and expressivity.

5. CONCLUSIONS

The results from this study suggest that a rich and well-designed animation library in a simple time-line based structure can

provide a useful tool for expressive, yet "quick-and-dirty" messaging. The avatar behavior represented in the clips made some communicatory functions more prevalent than in a pure text-based medium (e.g., real and fictional expressions). Users were comfortable with being restricted to one avatar, provided it contained a wide variety of expression options. The most popular behaviors used were warm and positive rather than cold and aggressive. Both text-image juxtapositions as well as the temporality of ExMS, provided useful tools for creative content creation. In addition, the design of the composition interface provided a good trade-off between workload and expressiveness.

6. ACKNOWLEDGMENTS

Funding for ExMS was provided by Nokia Research Center and Swedish Research Institute for Information Technology. Fredrik Bromée and Vesna Cengic implemented the system. Vesna also assisted in the user study. Christer Engström and Claes Jurander created the animations. Jussi Karlgren, Juha Hemanus, Panu Korhonen and Jaakko Lehikoinen provided valuable input and support.

7. REFERENCES

[1] Argyle, M. (1990) *Bodily Communication*, International Universities Press, (2nd edition).

[2] Barrientos, F. & Canny, J. (2002) Cursive: controlling expressive avatar gesture using pen gesture, Proceedings of *The 4th international conference on Collaborative virtual environments*, Bonn, Germany, 113-119.

[3] Bordwell (1985) *Narration in the Fiction Film*, Routledge.

[4] Cassell, J. & Vilhjálmsson (1999) Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous, *Autonomous Agents and Multi-Agent Systems*, 2, 45-64.

[5] Cherny, L. (1999) *Conversation and community: chat in a virtual world*, Palo Alto, CA: CSLI Publications.

[6] Dourish, P. (ed.) (1998) Special issue of *Computer Supported Collaborative Work*. (Volume 7, Issue 1-2).

[7] Fahlén, L.E., Brown, C.G., Stahl, O. & Carlsson, C. (1993) A Space based model for user interaction in shared synthetic environments. Proceedings of the ACM conference on *Human Factors in Computing (InterCHI'93)*.

[8] Greenhalgh, C. & Benford, S. (1995) MASSIVE: A Virtual system for teleconferencing. *ACM Transactions on Computer Human Interfaces (TOCHI)*, 2 (3): 239-261.

[9] Isaacs, E. & Tang, J. (1993) What video can and can't do for collaboration: a case study, Proceedings of *The first ACM international conference on Multimedia*.

[10] Jeffrey, P. & Mark, G. (2003) Navigating the Virtual Landscape: Coordinating the Shared Use of Space, in Höök, K., Benyon, D. & Munro, A. (eds) *Designing Information Spaces: The Social Navigation Approach*, Springer, 105-124.

[11] Kurlander, Skelly & Salesin (1996) Comic Chat, *SIGGRAPH'96*.

[12] Lee, C., Ghyme, S., Park, C. & Wohn, K. (2000) The control of avatar motion using hand gesture, Proceedings of the *ACM symposium on Virtual reality software and technology* 1998, 59-66.

[13] Murray, Denise. (1988). The context of oral and written language: a framework of mode and medium switching, *Language in Society*, 17, 35l-373.

[14] Persson (2003) *Understanding Cinema: A Psychological Theory of Moving Imagery,* Cambridge University Press.

[15] Rivera, K., Cooke, N & Bauhs, J. (1996) The Effects of Emotional icons on Remote Communication, *CHI'96* (Interactive Poster).

[16] Salem, B. & Earle, N. (2000) Designing a Non-Verbal Language for Expressive Avatars, *CVE 2000*, 93-101.

[17] Smith, M., Farnham, S. & Drucker, S. (2000) The Social Life of Small Graphical Chat Spaces, Proceedings of *CHI'2000*, 462-469.

[18] Turkle, Sherry (1995) *Life on the Screen*, New York: Touchstone.