# Expressive Messaging on Mobile Platforms

Per Persson<sup>1</sup>, Jussi Karlgren<sup>1</sup>, Panu Korhonen<sup>2</sup>, Janet Galore<sup>3</sup>, Mark Tierney<sup>4</sup>, Chad Redmon<sup>3</sup>, Juha Hemánus<sup>2</sup>, Peter Lönnqvist<sup>1,5</sup> & Jarmo Laaksolahti<sup>1</sup>

 HUMLE laboratory, Swedish Institute of Computer Science (SICS), Box 1263, 164 29 Kista, Sweden
 Nokia Research Center, Helsinki, Finland
 Honkworm, Seattle, U.S.
 room33, Stockholm, Sweden
 Currently at Department of Computer and Systems Sciences, Stockholm University/KTH, Sweden

Contact: perp@sics.se

#### Abstract: (60-150 words)

This paper presents design requirements for expressive, avatar-based multi-modal messaging on mobile platforms. It is argued that expressive messaging needs to exploit context of peers, embodied appearance and behavior, in combination with text. Our approach allows strong expressiveness and yet simple, on-the-fly message compositions required in a mobile, noisy setting.

#### **Keywords:**

Expressive messaging, Mobile devices, SMS, Multi-modality, Avatar

#### **Project URL:**

www.sics.se/humle/projects/expressmess

#### Expressive Messaging

People use simple text messages (SMS) on mobile phones in a volume that exceeds all expectations - in Finland, over a hundred messages per inhabitant and year (ministry of transportation). The key to the popularity of the medium is certainly not the complex and non-intuitive interface. Neither is SMS 'useful' in any task-oriented sense of the term. In contrast to e-mail and ordinary telephone calls, for instance, SMS is not really helpful in making work and professional life more efficient. Instead it is the social, expressive and emotional functions of SMS that seems to be the

primary objective (Koskinen, 2000). 'What's up?' messages, humor, flirts, gags, and all kinds of play are central objectives of textual messaging, as well as last-minute coordination/changes of already scheduled appointments and rendezvous.

In all of these cases there is a rich and shared awareness of situation, preferences, sense of humor, and social context between sender and recipient. SMS messages are typically not sent to strangers, but used in peer-to-peer communication between friends, lovers, or family members (Koskinen, 2000). Since the SMS information bandwidth is extremely low - maximum 160 characters, in low-resolution monochrome - there is reason to believe that expressivity exploits and rely upon that shared context.

The purpose of the Expressive SMS project – or ExMS for short – is to utilize the shared context, but enhance it even further through combining text with embodied and avatarbased communication. By allowing the user to easily compose text with animated and expressive gestures of an animated character representing the user, ExMS investigates the nature of expressivity and how users will utilize such tools in instant messaging on small mobile devices.

This paper describes the design requirements for such a system from a user's

perspective, developed by the initial phase in the project as a mock-up demonstrator.<sup>1</sup>

# Related Research

Our approach draws on quite a few research themes within agent technology, collaborative software and social computing. ExMS constitutes a mixture of different parameters and related research.

## 1.1Synchronous vs. Asynchronous

First there is distinction hetween synchronous and asynchronous messaging tools (Sakamoto & Kuwena, 1993). Chat, telephone calls and hot lines, involve real-time conversations with participants present at the same time, requiring direct response in dialogue-based fashion. Asynchronous а systems such as newsgroups, post-it notes, e-mail and voice mail, on the other hand, the recipient to attend to the allow incoming message when he has the time (Nardi, 2000). In these cases the sender does not expect swift response. More importantly, the sender can spend time composing the message, editing, undoing and redoing. In asynchronous communication the sender is allowed to think about the message, ornament it, and make it artistic or expressive in ways he may not have time or competence to do in the rapid synchronous conversation. By writing letters instead of meeting our lovers, we can 'play it safe' expressing quite precisely what we want to convey.

In-between the synchronous and asynchronous we find message systems that notify the recipient and possibly also interrupts her which present activities, but in participants can attend to other things inbetween messages. SMS and instant messaging can be responded to directly in a chat-like manner, but senders do not expect, and receivers are not obliged to engage in, such synchronous conversation (semi-) (Nardi, 2000; Koskinen, 2000). In this way, these

systems constitute a direct channel to a given peer, but still allow time for message composing. An expressive messaging system such as *ExMS* should be able to balance between synchronous and asynchronous in these ways. For instance, while many visual embodied chats support gestures and movements of the chatter's avatar via keyboard and mouse, such real-time systems can be experienced as stressful since they require immediate interaction with a system that high-level requires and fast interaction skills. With a semi-synchronous approach, ExMS eliminates such 'conversation stress' by allowing the user to spend time on message composing.

## 1.2 Automation vs. Direct Manipulation

solution to the Another conversation stress problem is to allow the system to automatically generate gestures, movements, expressions and even discourse of the avatar body. In this way, the user need not directly manipulate the behavior of the avatar, but can focus on the conversation itself. In the Comic Chat system (Kurlander, Skelly & Salesin, 1996), for instance, the system automatically creates comic panel compositions, what avatars to include in such panels, the placing of word balloons emotional expression. In and Swamped! (Blumberg, et al., 1999) the system seeks to infer the intentions of the user/avatar in a simple animal farm environment, and on the basis of this then generates behavior and camera angles. And AgentSalon (Sumi & Mase, 2001), collects behavioral data, information ratings, and touring records from museum or conference visitors and then automatically generates matchings and supposedly conversations between avatars interesting while the users observe their behavior on a large screen. (see also Kubota. et al. 2000)

In contrast to animation software, which permits the user to determine every nittygritty detail of the avatar's behavior, automation relieves the user of such time consuming work. At the same time, however, this also means less control, and the avatar might perform actions, express emotions and

<sup>&</sup>lt;sup>1</sup> Initially the project was called *MobiPal*. Implementation work is being conducted during 2001.

say things that were not intended by the user. As the artificial 'intelligence' in automation systems seldom is able to handle situated and often widely shifting goals and communication needs, automation will often be experienced as frustrating.

A mobile expressive messaging system needs to find a balance between automation and direct manipulation, supporting appropriate embodied expressivity in many kinds of situations, and yet allow compositions of such expressions in a time-efficient, onthe-fly manner. Such a trade-off system must permit user control as well as automation, allowing the user to easily shift between these two modes across communication situations.

## 1.3 Avatars vs. Characters

Focusing on the embodied expressivity (in combination with verbal discourse) we also need to distinguish between avatars and characters. A message can be expressive and embodied without any connection between the represented body and the sender of the message. Characters in films, on post-cards and paintings can perform expressive functions but disconnected from the director of the film or the sender of the post-card. Systems such as Miko<sup>2</sup> and Nescafé's NesCard<sup>3</sup> allow the user to create more or less electronic animated postcard with backgrounds, fictional characters and sounds, and sending them to acquaintances.

avatar-based communication In contrast, relies on a tight coupling between digital body and user. When the avatar speaks or behaves, the user behind it takes more or less responsibility of the contents. The avatar becomes a stand-in or body double of the speaker. This does not mean that the avatar must necessarily be photorealistically similar to the user as in video conferencing systems and web cameras; only that the behavior and discourse of the avatar can be interpreted as if they would come from the speaker if he or she were to be physically present. Smileys in text

messages, for instance, can be said to perform precisely this function.

In many visual chats, where the chatters enjoy a high degree of anonymity, the border avatars is between character and not distinct. Since there are few possibilities to find out the real identity of the user, speakers can choose and play with body representations, social roles and gender identities at their will (Turkle, 1995). Although the user direct the avatar and put the words in its mouth, the anonymity enables him or her to enter games of makebelieve, play fictional roles, harass other avatars and defy moral responsibility that would otherwise be present.

Since the focus of ExMS is on peer-to-peer messaging, each avatar will be recognized as belonging to a friend or acquaintance. In this way, our approach is widely different from anonymous chat rooms, even though both emplov avatars. It also differs from matching systems in which non-acquainted users and their avatar are brought together on the basis of some supposedly shared hobby, preference, taste or interest (e.g., Sumi & Mase, 2001). Instead, users of ExMS richer share much and more situated knowledge of each other.

## 1.4 Photo-realism vs. abstract graphics

mentioned, expressivity requires no As specific similarity between user and avatar. Moreover, in order to be expressive the avatar does not necessarily need to be represented in a photo-realistic style. A cartoon-like, abstract and sketchy animation of a character may be just as expressive as an animation that renders detailed features of skin, faces, gestures, shadows, limb configuration, motion patterns and gesture synchronization. Of course, such realistic styles of representation may have other effects. for instance believability or terms spectacle. In of humor, social bonding, emotion, attitudes, irony and other expressive communicatory functions, however, abstract animation styles may work just as well. Again, it is not in the animation itself that expressivity lies, but rather in the cocktail of image, verbal discourse and

<sup>&</sup>lt;sup>2</sup> www.fjallfil.com

<sup>&</sup>lt;sup>3</sup> www.nescafe.nu

users' shared context. The simplified, vet expressive appearance "emoticons" of in written text is a proof of this. By allowing the user to juxtapose blocks of simple and abstracted animation with blocks of text, ExMS exploits the same expressive mechanism as emoticons. Combining a textual message with a semi-imagery representation of a face adds new layers of meaning, e.g. irony, guides the recipients' interpretation of the message, and expresses the sender's state or character. emotional Creating animated emoticons that move and transform over the temporal flow of the message will merge modalities in similar ways to cartoon and animated film.

With this theoretical framework in mind, we identified the following design challenges: How to allow users to compose highly situated and expressive avatar-based messages in a time-efficient manner? What is the relation between gestures and movements and the surface skin/appearance of the avatar? The interaction descriptions given been realized in a mock-up below have concept demonstrator.

# Design Choices

## 2.1 Composing Messages: Moods and Events

In order support simple expressive to messaging, we decided to center the composing around two axes, which we called moods and events. Moods relate to the basic emotional undertone in the message. Composition starts by selecting from a palette of moods, e.g., happy, distressed, angry and busy. The animation showed, supposedly reflects this mood in a simple cartoon-like fashion. Depending on the length of the total ExMS - decided by the sender - the animation would make continuous loops until the end. In the preview mode the user can, at any time in the composition phase, inspect the message composed so far (Figure 1).



Figure 1 Example Screen

In addition to mood animations, avatars should be able to perform simple events, e.g. jump, dance, smile, laugh, and weep (possibly connected to sound effects). Events are thought to mark and enforce certain text snippets, or break the basic mood undertone of the message. Events can be added at specific points in the message. This is done through previewing the message and then. at the appropriate moment. interrupting the message where the event should be inserted. From the list of events, one is chosen, which brings the user back into preview mode again, this time with the event inserted in the loop break closest to the point where the user interrupted the message.

Since all mood and event animations start and stop at a neutral position frame, continuity of movement can be ensured without involving the user.

Combining text and animations in appropriate temporal constellations will be central in conveying punch lines or other emotional content. In the mock-up, however, such support is not well articulated. It allows only simple text input, which distributes itself over the length of the message. In the final implementation this has to be further investigated.

On an algorithmic level such temporal juxtapositions also have to be supported. To this end, we have sketched a standard for encoding expressively enhanced messages: the Mood Markup Language or MoodML (Figure 2).

Composition is executed though loops of inserting events/texts and reviewing the message. When the user is satisfied with the work, he sends it away via his telephone book.

```
<MoodML skin="88F2A0F" id="janet"
to="0405047123" mood="happy;8">
I'm excited to see you!
<event="jump">
Be there at ten.
<event="dance">
</MoodML>
```

#### Figure 2 Sample MoodML – an internal format for expressively enhanced multimodal messages

## 2.2 Quickies and Recycled Messages

In some cases, standard scripted responses to messages can be useful. Our design gives users the possibility to save the script from both incoming and outgoing ExMSs as 'quickies'. Such quickies can be reused later, exchanging the balloon text to fit the new situation. Quickies enable users to share particularly spectacular or useful sequences of moods and events. Reusing other friends' messages as quickies is possible through the saving of the message MoodML script. When I want to reuse the quickie for my own avatar, whose skin might be totally different than my friend's, the system identifies the name of the animation in the MoodML script and then re-instantiate the sequence in the skin of my avatar. This takes us to the issue of skins.

## 2.3 Skins - avatar appearance

Although the temporal juxtaposition of text. moods and events constitutes one channel of expressivity, naturally the animations, or the surface appearance of the different animations add, transform and personalize that expressivity. The purely graphical design of the avatars we call event can thus be skins. Any mood or type of skin represented in any that reflects a personality feature, identity or trait preferred by the user. Again, skins may not reflect features that the user necessarily possesses, but rather one which support the kind of expressivity the user favors.

Before initiating the ExMS service, users should be able to choose a prefabricated from a skin library. Each skin will represent moods and events through varying styles, in addition to attributes such as age and gender. In order to ensure personalization and individualization of the avatars, it should also be possible to complement a skin with a set of add-ons such as clothing style, colors, T-shirt texts, hats and even tools and instruments such as walkmans, sticks and roses.

Most users will probably make use of standard skins included in the service package. Other skins may be produced professionally for marketing purposes as a product placement platform: users can get new skins for free, if they allow their avatar to wear a brand emblazoned T-shirt.

Some artistic users, however, will want to manufacture their own skins. To this end, open standards and APIs are absolutely essential. No single organization will have the stamina to produce and uphold the interest of the potential user community: the creativity of the users must be tapped into. Thus, it is important to create a community in which sharing scripts and skins are encouraged and awarded.

However, in order to ensure compatibility between skins, scripts, and users, there have to be some minimal requirements for any given skin. Each skin has to be able to perform a minimum list of moods and events.

## 2.4 Configuration

Configuring the interface and the characteristics of the skin, archiving message scripts, sharing scripts and skins, community building and other administrative aspects of the system are accessed by the user from a Web interface rather than over the mobile device: the mobile device is used solelv for composing, transmission and receipt of messages. We expect most users will not be too concerned with the niceties of configuration; but as in most consumeroriented communication applications we do expect some users to want access to all controls.

## Technical challenges

From reviewers comments:

Give an system overview according to your slides on your website, current. next. future terminal, web tools and server concept. Please state the technical challenges of your proposal, if possible give more information on MoodML.

Per: Here I think Jussi (or Panu?) needs to fill in some of their initial ideas we had last year and that was presented in the PowerPoint presentation.

Mark: You could highlight EMS and MMS here. We should also state somewhere that users are also downloading icons and ring tones to personalize their phones and avatar skins can be a natural extension of this.

# Conclusion

In terms of overall research questions in the ExMS project, we must address a number of relatively vaguely understood notions, and first and foremost gain some provisional understanding of expressivity in a social context. How can embodied interface avatars enhance it? How does expressivity relate to role-playing and an individual's sense of self and personal and social identity? On a expressive level, we want the social messaging functionality to be a shared notion in a community of users. How can we support the growth and development of such a community? And on another level of expressivity, regarding the message content, we need to understand the relation between discourse flow, gestures and situation.

From the interaction design proposed in this paper, the ExMS project has now entered implementation phase, and expects to initiate a longitudinal user study in late 2001. Surely, the design proposed here will transform during the way, due to new insights as well as limitations of present wireless technology. It is however our intention to conduct a realistic, nonlaboratory user study, including about 20 subjects over a time of one month. ExMS is typically an application whose effects, ease of use and social functions will emerge first after long-term usage.

Moreover, evaluating a messaging tool such as this cannot be done using utility measures. From a task-oriented perspective, this design will make communication slower and more expensive - each message will take more time to compose and require more transmission capacity for the same informational content. But the driver for this design is that of enjoyment and expressiveness, not efficiency or cost.

Composing a message on a mobile device is that of doing it different to on а stationary terminal at work: using "dead" time on buses, in bars, in cafés and on school hour breaks will give ample time to compose and perfect avatar-based messages. In contrast to work where deadlines and concentration are crucial to perform well, mobile users in the city, countryside and on public transportation can be expected to be more leisure oriented. They are probably more curious about social space, and more be entertained. Thus willing to the notification signals of instant messaging systems such as SMS or ExMS be may experienced as happy surprises rather than disturbances. It seems to us that mobile application development and studies need to take the separation between work related and leisure time applications seriously. (cf. Espinoza, et al., 2001).

## References

- 1. Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore, E. & Bylund, M (2001) GeoNotes: Social and Navigational Aspects of Location-Based Information Systems, In Proceedings of UbiComp'01, Atlanta, September.
- Johnson, M.P., Wilson, A., Blumberg, B., Kline, C. and Bobick, A. (1999) Sympathethic Interfaces: Using a Plush Toy to Direct Synthetic Characters. In *Proceedings of CHI'99*, 152-58.

- Hudson, S. & Smith, I. (1996) Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems, in *Proceedings of CSCW'96*, 248-57.
- Koskinen, Topi (2000) Mobile Asynchronous Communication: Exploring the Potential for Converged Applications, *Personal Technologies*, 4, 45-53.
- Kubota, H., Nishida, T. & Koda, T. (2000) Exchanging Tacit Community Knowledge by Talking-virtualized-egos, Agents'2000, Barcelona, Spain, 285-92.
- Kurlander, D., Skelly, T. & Salesin, D. (1996) Comic Chat, in *Proceedings of SIGGRAPH'96*, 225-236.
- 7. Ministry of Transport and Communications, Finland (1999) Suomen Tekstiviestimarkkinat. Helsinki: Edita.
- Nardi, B., Whittaker, S. & Bradner, E. (2000) Interaction and Outeraction: Instant Messaging in Action, *Proceedings* of CSCW'00, 79-88.
- 9. Sakamoto, Y. and Kuwana, E. (1993) Toward integrated support of synchronous and asynchronous communication in cooperative work: an empirical study of real group communication, *Proceedings of Organizational computing systems*, 90-97.
- 10. Sumi, Y. & Mase, K. (2001) AgentSalon: Facilitating Face-to-Face Knowledge Exchange through Conversations Among Personal Agents, Agents'01, Montréal, 393-400.
- 11.Turkle, Sherry (1995) *Life on the Screen*. New York: Touchstone.