Sensation: a Presence Enabler for Long-Distance Relationships using Skype and Visual Presence Representation

Alexander Ljung

MSc. student of Mediatechnology, Royal Institute of Technology (KTH)

Högalidsgatan 40 D S-117 30 Stockholm

aljung@kth.se

Eric Wahlforss

MSc. student of Industrial Economics, Royal Institute of Technology (KTH)

Bergsundsgatan 15 S-117 37 Stockholm

ericwa@kth.se

ABSTRACT

Media spaces have existed in the labs for some time, but they have yet to enter the homes of ordinary people. Most research on media spaces has been focusing on *mediating physical spaces*, *aiding communication in work*, or *providing for multi-part conversations*. There's little research on media spaces for private use, and even fewer working systems.

In this paper we describe *Sensation*, a system that enables two persons to establish a non-intrusive audiovisual link between one another, utilizing Skype for audio transmission and subtle pulsating lights for presence indication. Sensation is *not* a phone, *not* a conference system, and – most importantly – it is *not* about connecting just spaces. The Sensation is a *presence enabler* in long-distance relationships, connecting people who happen to live far apart. It is intended as a realistic near-future consumer product, as such affordable and easy to use.

Author Keywords

Skype, media spaces, audio spaces, presence over distance, mediated conversation, tangible interface, design, HCI

ACM Classification Keywords

H.5.2. User-centered design

INTRODUCTION

Emerging technology is bridging the distance between people who wish to feel closer to each other than their geographical location allows. However, even though technology for mediating audiovisual presence between two or several participants exists today, widespread use is limited. In audio communication, the telephone – a 130-year-old invention – still remains by far the most popular device in use. We believe this is due to several reasons. Firstly, dedicated alternative communication hardware remains relatively expensive for the average consumer. Secondly, even though many people are already in possession of technology such as computers and broadband connections, capable of creating joint media spaces, these technologies have had limited impact for audio communication. We see this as partly due to the users'

mental image of what a computer is and does, but also due to the fact that computers aren't *designed* for audio communication; additionally, easy to use software solutions have not been available. In the recent past however, Skype [1], a software application for Internet telephony, has become widespread; today many people are complementing their mobile and landline telephone with Skype, using their computers to make phone calls.

We believe that the power of software like Skype – that is essentially providing worldwide audio communication for free – is enormous and clearly capable of more than is indicated by today's usage patterns. Compared to the audio quality of a regular telephone, Skype is vastly superior, even on a limited Internet connection.

We believe that the use of software like Skype, utilizing the Internet for audio communication, has great potential in enabling private long lasting *media spaces* between people. This means that usage will not be limited to telephone-like conversations; users will also be able to establish long-lasting high quality "audio links" that enables them to interact in a more face-to-face manner. As Ackerman et al. have noted, audio links are attractive for consumers as a media space "base" since the cost of bandwidth is low in relation to video links [2]. By creating an audio space between two party's in a distance relationship, a sense of increased *closeness* and *togetherness* is possible. Furthermore, several studies [e.g. 2, 3] have showed that audio-only, as opposed to audio *and* video is a plausible option when creating presence over distance.

Good interfaces are crucial for this emerging technology to reach widespread adoption. Our approach has been to create a tangible interface and dedicated artifact for using Skype in a shared media space fashion. We wish to connect users over a distance and enable them to discover the presence creating possibilities that informal mediated conversation and high quality ambience sounds afford.

BACKGROUND

The term *media space*, originally coined by researchers at Xerox PARC, is referring to environments created using video, audio, and networked computers to support

interaction between distributed groups of people. Much has been written about media spaces, especially those involving video. Karahalios et al. have created *Telemurals*, connecting two university dormitories with interesting results [4]. However, most studies in the field are focused around shared workspaces and collaboration over distance. As noted by Ackerman et al., research on audio only as an enabler for media spaces is scarce [2]. Moreover, research on how this can affect the users in the private sphere who simply wish to *feel each other's presence* over distance and have no explicit work related collaboration or efficiency desires is virtually non-existent. However, there is some research that has provided us with fruitful insights and guidance.

Audio-Only Media Spaces

Ackerman et al. research on the Thunderwire project contains information about how a high quality long lasting audio only media space can be adopted in a work environment [2]. Interestingly, they find that the primary usage of the system (that was installed, with its usage observed for 3 months) was for informal, non-work related communication. In fact, less than 10% of the actual information exchange was related to work issues. Moreover, the users describe the system as something allowing them to 'hang out' together, creating a possibility for socializing like one would in e.g. a café or during a coffee break. Users describe their liking of the telepresence aspect as derived from them being able to overhear background noises such as papers moving and people tapping keyboards.

In the MIT Media Lab project *Visiphone* [5], both the audio-only media space and visual conversation representation was explored. The group mentions the importance of an audio only system to have a clear and salient visual feedback, showing *when* the system is in use (i.e. sending audio):

> An open long term audio connection alone between two spaces would make it possible to walk around a room and talk to a distant friend at will. Yet such a system would also have an unfortunate surveillance-like quality: it would be easy to forget that one's space was actually a portal to another space and that all of one's comments were being heard elsewhere. A visible interface that indicates the existence of the live connection alleviates this problem, serving as an ongoing reminder of the audio link. – Donath et al. Visiphone [5].

Issues of privacy become crucial for an audio system since audio itself lacks the affordance of visual feedback; in moments of silence, audio lacks *any* feedback whatsoever.

Interaction

Singer et al. have noted that a Graphical User Interface (GUI) is not a desirable interface solution for audio spaces [6]. Moreover, their research shows that users prefer not to have controls for such parameters as audio timbre, localization (i.e. *where* the voice should be placed in a multidimensional playback environment) and such. Furthermore, they found additional indications that coincide with observations by Ackerman et al. on that awareness of the other users' presence is highly desirable [2].

Singer et al. have also undertaken comparative studies of various ways of interacting with an audio space, and they have found that the users consider a tangible interface highly engaging [6]. One of the key benefits examined is the "casual communication", and they state it is much due to two factors: *lightweightness* and *persistency*.

> Persistency refers to communication being available continuously, in contrast to telephone calls which are explicitly started and stopped. Lightweightness refers to the lack of effort required to initiate or end communication; again, the telephone is not lightweight because of the need to pick up the handset, dial a number, and so on. – Singer et al. Hanging on the 'Wire [6]

Mediated Conversation

Levow states that less than 5% of the pauses in a regular telephone conversation are more than one second long and that this is substantially shorter than in a face-to-face conversation [7]. According to Levow this is due to the audio channel's lack of visual cues existing in a face-to-face conversation. As showed by Karahalios, full duplex audio, is best catered towards informal conversation, but technically this also introduces problems with feedback, audible as echoes [5]. Ackerman et al. have reinforced this by stating that low quality half duplex audio causes more formality and conversational awkwardness [2]. Furthermore, audio space communication seems to have a more face-to-face like quality than the telephone as that the conversation appears to have fuzzier beginning and ending points [2].

Additionally, we wish to emphasize the intrusive nature of the telephone as stated by Kanada among others [8]. The process of ringing is insensitive of the receiving party, thus intrusive by neither being context aware nor allowing the receiver to indicate if – or how – he or she wishes to be contacted.

SENSATION

With this background in mind we set out to create *Sensation*, a tangible presence application designed – in particular – for long-distance relationships. Sensation enables two persons to establish a non-intrusive audiovisual link between one another, utilizing Skype for audio transmission and a subtle pulsating light for presence indication. Sensation is *not* a phone, *not* a conference system, and – most importantly – it is *not* about just connecting spaces. Sensation is a *presence enabler* in long-distance relationships, connecting people who happen to live far apart. It is intended as a realistic near-future consumer product, as such affordable and very easy to use.

Description

Sensation is based on three elements; *the base station, the indicator* and *the sender*. The base station functions as the

"brain" of the system, containing most of its communication technology. It also acts as a charger for the wireless indicator and sender. The indicator functions as an omnidirectional speaker and furthermore contains the visual feedback – essentially a pulsating light – of the other party's activity. The sender is a small wearable device comprised of a microphone and controls for the base station and the indicator. Sensation is based on a representation of the relationship; the sender representing *you* and the indicator representing your partner.

When a joint media space is fully active the sender's microphone picks up audio, and via the base station sends it to the partner's indicator speaker. Additionally, the audio picked up by the microphone is processed by the base station and sent as information represented by the partner's indicator light pulse. The light pulse emitted is a representation of the activity of the other party's "audio sphere" over time. The light is an average intensity indicator over longer periods of time; i.e. if the audio intensity has been high for fifteen minutes the light will pulsate more rapidly than if it has been low. It is important to stress though that this is *not* a volume meter; the light intensity is averaged over time thus not rapidly pulsating due to sudden fluctuations in the audio.

The indicator is completely mobile and can be picked up from the base station and placed in a new location. For instance, the user might have Sensation placed in his or her living room, but wishes to keep the ongoing conversation active while he or she moves into the kitchen to prepare dinner. In this scenario, the user would simply pick up the indicator, bring it to the kitchen, and place it close by – without having to stall the conversation.

The interface is simple; it consists of three buttons for controlling what the user wishes to send to the other party (sound and/or light, or nothing at all). The controls are easily accessible by being situated on the wearable sender device. The sender also contains a volume knob for controlling the received audio level.

Design Process
The design process used for developing the Sensation has been highly user-centered. We have engaged potential users at as many stages of the process as possible, and through iteration we have tried to narrow down the feature set of the Sensation as much as possible, focusing on user requirements.

Since the intended use is for long-distance relationships, we have spent time trying to understand some of the social aspects of these relationships to gain insight into what desires these conditions create. We posed ourselves the question of what is it these people lack the most in their relationships, and how our vision of communicative technology could aid them.

A decision to create a high fidelity prototype was taken at an early stage in the process. This was motivated by the unconventional nature of media spaces and the belief that users would need to actually experience a media space in order to comprehend their potential drawbacks and flaws. Furthermore, prototyping is an excellent way to quickly find out what is functioning as the users expect, and also to understand what needs to be tweaked in the design.

User Testing
The most important user testing, observations and interviews can be divided into the following four categories: qualitative interviews concerning long-distance relationships, qualitative interviews with Skype users, interface testing and prototype testing.

Relationships Over Distance
Qualitative interviews with persons engaged in long-distance relationships were conducted to understand more about the nature of this type of relationship, and to understand what social elements these people miss due to the distance from their partner. By looking at the needs of the people in the relationship we wished to find hints of what Sensations design should accomplish. In total, thirteen relationships were examined. The subjects varied greatly in age span and demographics, their only common denominator being that they were engaged in a long-distance relationship.

Not surprisingly, most of the interviewees stated that the physical presence was the quality most sought after, followed by the desire to have an everyday "uncomplicated" experience of one another.

Two groups where distinguished among the interviewed couples. The first group saw the long-distance relationship as a temporary phase, something that would soon pass. The other group, on the other hand, saw the distance as something lasting and as *part* of their relationship. The second group appeared to have come to terms with living separately and thus was more interested in solutions for minimizing the perceived distance between them. Furthermore, the first group expressed uncertainty as to whether an increased sense of presence would necessarily be positive. They felt that an increase in sensed presence while still lacking physical quality could be frustrating and actually *amplify* feelings of longing.

Skype Users
Due to limited time (the entire project spanned over two months) only a limited set of qualitative interviews with current Skype users were conducted. The two interviews were undertaken with frequent users of Skype to learn about their perceived benefits, drawbacks and usage habits. One of the users described how he would sometimes let the connection stay open, even though there was "no reason" for doing so other than just "feeling the presence" of the person at the other end. Both users described a frustration of having to use Skype in front of a computer and not being more spatially free.

Interface Testing

A GUI based interactive model of the interface was programmed and tested on five potential users in ages 20 to 25. The tests consisted of observations of their behavior and interaction patterns with the interface model, and furthermore of a brief discussion on how they perceived the interface.

The tests showed a small initial learning barrier; the users could easily understand the symbols and a few controls. However, after some brief usage certain users would become slightly confused and mistake the sender device for being a remote control for their indicator (i.e. they would press the *send light* button and expect the light on their indicator to come on). This indicates that the conceptual model is not totally grasped by the user at first but takes some mistakes, and possibly frustration, to understand.

Once the controls were learned there was no evidence of users forgetting what function a certain control had or how to achieve a certain desired action. This would indicate that the system is lightweight and not cognitively cumbersome for the user.

Prototype Testing

Whereas two Plexiglass mock-ups of the Sensation where created, they were not fully functional prototypes. Consequently, conventional loudspeakers, wireless near-field microphones and laptops in combination with the mock-ups where used to create an approximate real life test of the Sensation. We installed equipment in the apartments of two people in their mid twenties engaged in a distance relationship, and let them use the system over the course of four days. We then conducted qualitative interviews with them.

The experiences of our test couple seem to verify earlier research [2] in that, after about 20 minutes time, their mediated conversation would take on a different character than a face-to-face- or phone counterpart. Our test couple experienced "less talk and to-the-point conversation", but rather a sort of comfortable "presence exchange" after being online for some time. They pointed out that near-field microphones weren't good enough for providing a sense of room ambience – a solution to the problem could be to integrate two microphones in the system; one for ambiance and one for voice. Both of our interviewees said they "sometimes stared at the loudspeaker" when listening to the voice of the partner; something that point to the importance of a well-designed indicator artifact that is able to convincingly mediate the presence of the other. They also raised the surveillance issue, saying they sometimes felt like being monitored by the other party. This could possibly be due to the fact that no visual presence indicators were functional in the tests.

In general, our interviewees were quite enthusiastic about using the system; in fact, they both expressed a will to install the Sensation permanently in their homes.

FUTURE

Sensation is the result of a research and design project extending over two months. We are now at the point where we could build a functional prototype, conduct further user testing, and continue development. There are several areas that need to be researched in detail. How, for example, should the visual presence indicator work? What audio-to-light algorithm is optimal for conveying presence without intruding too much? Could the physical design evolve? Do users want the artifact to be strikingly present or should it rather be "disappearing", a non-design? On the technical side, how do we integrate Skype with the base station? Can the interface be made yet easier to use than today? There is a multitude of problems to be discussed and solved.

CONCLUSION

Media spaces have existed in the labs and in high profile business environments for quite some time, but they have yet to enter the homes of ordinary people. As stated earlier, most research that has been made on media spaces has had scopes quite different from ours; it has been focusing on *mediating physical spaces*, *aiding communication in work*, or *providing for multi-part conversations*. We believe – and our research seems to show – that there's room for a presence application aimed for private and intimate relations, in particular for long-distance relationships. Judging by the qualitative interviews, there seem to be two different subgroups within our main target group; couples with temporary- and couples with permanent long-distance relationships. In the latter group, were the long-distance relationship has become part of everyday life, there seems to be a strong interest in a product like the Sensation.

While projects like the already mentioned Visiphone [5], have been created with "casual communication" in mind, we argue that a system like the Sensation has more realistic chances of getting widespread adoption. The Visiphone for example, is bulky, rather expensive, and has not been properly tested on its potential target users. At the time of its creation, it was very much seen as an unconventional experiment. The Sensation, on the other hand, is a product that could be manufactured at a reasonable price *today*.

REFERENCES

1. http://www.skype.com/

2. Ackerman, M.S., Hindus. D., Mainwaring, S.D., Starr, B. Hanging on the 'Wire: A Field Study of an Audio-Only Media Space. ACM Transactions on Computer-Human Interaction, Vol. 4, No. 1, March 1997, 39-66.

3. Isaacs, E.A., Tang, J.C. What Video Can and Can't Do for Collaboration: A Case Study. Multimedia '93.

4. Karahalios, K., Donath, J. Telemurals: Linking Remote Spaces with Social Catalysts. In Proc. CHI 2004, ACM Press (2004).

5. Karahalios, K.G. Social Catalysts: Enhancing Communication in Mediated Spaces. Massachusetts Institute of Technology, MA, Cambridge. 2004.

6. Singer, A., Hindus, D., Stifelman, L., White, S. Tangible Progress: Less Is More in Somewire Audio Spaces. In Proc. CHI 1999, ACM Press (1999), 104-111.

7. Levow, G. Making Sense of Silence in Speech User Interfaces. CHI '97 Workshop: Speech User Interface Design Challenges.

8. Kanada, Y. Multi-Context Voice Communication in a SIP/SIMPLE-Based Shared Virtual Sound Room With Early Reflections. In proc. NOSSDAV '05. ACM Press (2005), 45-50.

9. Golder, S. The Keep-In-Touch Phone: A Persuasive Telephone for Maintaining Relationships. In Proc. CHI 2004, ACM Press (2004), 1551.