MULTIMODALITY IN A THREE-DIMENSIONAL VOICE CHAT

Therese Örnberg Berglund
The Department of Modern Languages and HUMlab
Umeå University, Sweden

Abstract

In recent years we have seen an increase in computer applications that support multimodal online communication. This paper centers on communication in one such multimodal application, Traveler, a graphical three-dimensional environment that allows for voice communication. Based on preliminary results from a study investigating communication patterns and negotiation strategies in this environment this paper gives some examples of how modal density is created here, building on the methodological and theoretical framework proposed by Sigrid Norris (2004). Through qualitative analysis of the material, both actual and self-perceived behavior of beginner participants is compared to that of more accustomed users. The paper also includes a discussion on the relationship between modal density and notions such as Common Ground and Presence.

Keywords

Computer-mediated Communication, Modal density, Conversational negotiation, Common Ground, Presence

Multimodal interactions

Face-to-face interaction is always multimodal. As human interlocutors involved in communication we do not only have verbal language at our disposal, but by combining gestures, facial expressions, intonation, positioning and, admittedly, in most cases also language, we can transmit complex messages on several levels simultaneously. In addition, also appearance and spatial configuration influences our interactions. The interrelation between different modes is something which an increasing number of linguistic scholars are paying attention to when analyzing interactional meaning. If only language is taken into consideration, the full complexity of meaning construction cannot be accounted for, and the analyst misses out on important information which the interaction partner has access to when decoding the messages.

In recent years, many of our daily interactions have moved online, a transition which has consequences for communicative practices. Relevant in this context is the way in which technological mediation affects the modalities which can be drawn on in interaction and communication. In this paper, it is argued that these effects are not necessarily negative, but rather different media have different affordances (Gibson 1977, Norman 1988, Hutchby 2001). Nevertheless, if we want to learn how to identify the platforms that are best suited for our purposes, it is important to investigate how the different modes available are being employed, and how language and communication are adapted to fit these modalities by both beginners and more accustomed users in different online environments.

In this paper, I refer to preliminary results from a qualitative study of interaction and communication in a graphical three-dimensional voice communication environment, Traveler. Based on an analysis of the communicative interaction and on answers to questionnaires I give examples of how modal density is created in this environment, how multimodal behavior differs between beginner and non-beginner users, as well as how modality interrelates with concepts such as Common Ground and Presence. Before we turn to the preliminary findings, I will give a brief overview of previous research on multimodality and Computer-Mediated Communication (CMC), as well as an introduction to the theoretical and methodological framework on which this analysis builds.

Multimodality and Computer-Mediated Communication

The fact that the verbal is not the only means of communication is something which has been acknowledged by researchers of computer-mediated communication over the years. However, theories such as that of *media richness* (Daft and Lengel 1984) and the *cues filtered out* approach (see Walther 2002 for a summary) have seen face-to-face communication as the ideal speech situation, and argued that the modes should imitate those of face-to-face communication as closely as possible. Some CMC researchers, for instance Joseph B. Walther (2002), have heavily criticized these approaches and shown that just because some cues are missing this does not necessarily mean that communication will break down or that relationships will not thrive. On the contrary, Walther (1996) argues that interpersonal relationships that develop in, for instance, written modes may instead become *hyperpersonal*, a term which indicates that the specificities of the media used for communication may facilitate even more intimate relationships than face-to-face interactions. A similar pattern in change of emphasis can be identified when surveying the developments within presence research. Whereas Short et al.'s (1976) notion of social presence as communicating under face-to-face like conditions is still prevalent today, for instance in virtual reality research, there are also other movements within presence research where focus is turned to the larger social context, including attitudes and social equality (c.f. McIsaac & Gunawardena 1996), and where also the role of imagination is taken into consideration (c.f. McLellan 1996).

In this paper I would like to suggest an alternative approach to multimodality and CMC, which in line with the reasoning of J.B. Walther and recent research on presence does not claim that less modes equals less efficient communication, but instead centers on the different affordances of the different media and on how modal density is created.

Theoretical and methodological framework

Modal density

According to multimodality researcher Sigrid Norris (2004), it is never possible to count the modes available in a communicative situation, since they are merely heuristic units of analysis. Instead she advocates an approach to multimodality where focus is on the creation of modal density. Modal density relates to levels of attention, and there is no inherent hierarchy among modes, but it all depends on the situation. In Norris' view, modal density can be achieved either by *intensity*, which means that one mode is best suited to deliver a message under present circumstances. As an example of this, Norris points out how the verbal language is given prominence when speaking on the phone. *Complexity* may also result in modal density, in cases when several different modes are used simultaneously to deliver the same message and none of the channels is given higher prominence than the others.

By applying the theory of modal density to computer-mediated communication this would indicate that in fact only one mode is needed in order for modal density to occur – only more attention will be devoted to this one mode, as in the case of written CMC. Thus, here it is argued that an approach which focuses on modal density rather than on perceptual realism offers a worthwhile possibility when analyzing multimodality in CMC.

Conversational negotiation

Building on the work of interactional sociolinguists (Goffman, Gumperz, Goodwin, Duranti) this paper argues that interaction is key to an understanding of how meaning is construed. Here, the notion of conversational negotiation is used to refer to the recurring subconscious strategies through which reciprocal co-construction of content, structure and context, and thus discursive coherence, is enabled. Negotiation in communication can take on many different forms, depending on both the level of negotiation and the strategies employed.

Participants in conversation subconsciously negotiate in order to make sense of what the other expresses through an utterance as well as how to continue the conversation structurally, in line with Common Ground theory (Clark and Brennan 1991) Furthermore context is co-constructed in conversation (c.f. Duranti & Goodwin 1992), which for example can be seen in the negotiation for face (c.f. Goffman 1967), identity (c.f. Weedon 1997), and solidarity and support (c.f. Aston 1986, 1993). Table 1 shows some examples of negotiation strategies on the different levels of negotiation.

Table 1: The three levels of negotiation

Negotiation level	***Content***	***Structure***	***Context***
Negotiation type	Negotiation of meaning; shared conceptualizations	Negotiation of process	Negotiation of solidarity, support, face, identity, roles etc.
Representation in interaction	Back-channelling, clarifications, repairs, repetitions, elaborations, shared points of references	Communication management: turn-taking, structuring etc.	Attitudes explicitly expressed or implicitly suggested, face-saving strategies etc.

Material

The Traveler environment

Traveler is a three-dimensional voice-enabled virtual environment, which is mainly used for socializing, and where there is a strong sense of community among the regular users. The program, which is free to download to your computer, was created in the mid 90's and despite some improvements it still does not demand very much of your computer or your internet connection. Upon entering Traveler you choose and customize your avatar. The avatars are mainly big heads, which in itself has some interesting implications for multimodality. The graphics are on a quite basic level, as are the different non-verbal expressions that you can make your avatar express by clicking on certain buttons. In Traveler you communicate via voice, and there is lip sync between the sound and the avatar – however, not with any phonological detail. The sound is distance-attenuated, which means that the further away from someone you move the less well you hear what that person is saying and vice versa.

Modalities in Traveler

The modes available in Traveler can be grouped under three headings: Audible, Visual and Spatial modes. Table 2 lists the modes that have been identified as central in Traveler, to be further discussed in the section on preliminary findings.

Table 2: Modalities in Traveler

Audible modes	Visual modes	Spatial modes
Language, prosody, pause, extralinguistic audile markers	Facial expressions, push-to-talk, appearance	Layout, proxemics, movements

The filmed gatherings

My material consists of five filmed gatherings in Traveler, where different groups of people have met to discuss different issues. Four of the gatherings have been filmed from two different perspectives. The participants in the discussions are academic language teachers, a group of researchers, a student group participating in a course of English at a distance, and parts of the Traveler community. Apart from community members, most participants are beginner users of Traveler. The number of participants has varied between 5-15. The majority of participants have not had any contact prior to the meetings, but some have known each other from before, either through virtual encounters or through face-to-face meetings. In three of the gatherings I have participated myself and in the other two I was present as a passive camera doing the filming.

Preliminary findings

Generally speaking, the modes connected to the medium of sound have the highest modal density in Traveler. On the content level, verbal language is used to generate topics and meaning, to clarify, to repair, to repeat, to elaborate, as well as to emphasize and to give feedback on the content of interaction. Intonation and pauses are also used to indicate emphasis, and prosody and extralinguistic makers are sometimes used as back-channeling devices. On the structure level, the different modes connected with the sound all are used as cohesive devices. On the level of context, these modes indicate conversational style, which of course is an important part of a person’s identity. Pauses are sometimes also employed in face-saving strategies.

An important cohesive device in Traveler is the push-to-speak function. In order to be heard you have to push the ctrl button. This is both audible and visible to the other participants, and if someone has indicated that he/she wants to take the floor by pushing down ctrl, this person will often also gain access to the floor. Apart from the push-to-speak function there are not many visual indicators that someone wants to speak, something which at times makes turn-taking difficult. It should also be noted here that an additional reason why turn-taking sometimes is difficult in Traveler is the short time lag on the server. People who know about this might choose to have short pauses between utterances to avoid overlaps, without this making them uncomfortable.

As far as the visual modes are concerned, my material indicates two quite interesting phenomena. For one, the ability to use facial expressions is employed only to a limited extent, and in most instances by accustomed users, either as face-saving strategies or to display level of attention. This indicates that in this environment emotional expressions to a great extent depend on the sound capabilities. In addition, the role of the avatar in identity construction is debatable. In my material, I have found indications that the visual representations, that is, the avatars chosen, start to matter less once you get to know the person behind the avatar. However, if someone changes avatar, this is often commented on as problematic, which indicates that even though these visual representations are not the focal point of attention they are still important. The anonymity of the avatars is another factor which influences roles and relationships, in that, at least in an initial stage, people are more equal – then these contextual factors are negotiated via other modes.

In Traveler, spatiality seems to be of greater importance than visual qualities. People tend to form circles and those who feel comfortable maneuvering their avatars usually turn toward the person who is speaking. Spatial cues indicating that one is paying attention are important strategies on all levels of negotiation. This form of gaze is also apparent when accustomed users want to show whether they are addressing someone in particular or the whole group. Apart from movements where the avatars turn to face one another, also head-movements, most often nods, are used for back-channelling – mainly by accustomed users or those who have been specifically instructed on how to use this feature. These cues mainly occur on their own rather than in combination with verbal contributions. Another spatial cohesive device that some accustomed users employ to indicate level of participation is by moving their avatars back and forth to show that they want to take the floor and then open it up again.

One further example of how the spatial mode has effects on interaction is the fact that the spatial layout itself influences the contributions of the participants, in that it sets the expectations on level of formality, etc. This is illustrated in the following excerpt from a transcript of a student session in Traveler:

Table 3: Example from a session in Traveler

((*M and O move around in the space and then return to S; O turns to face M*))
M Do you think that we should stay here at the eh at the gate or if we should gather around a table to be more structured I don know
O I don't think it matters since eh it's only us here
M Okay I was just eh thinking about ehm being able to like to to stay to stay in line of eh of everybody's view so that so that eh everybody sees everybody cause it's it's I don't know it probably doesn't matter
O But perhaps it's more professional if we sit at the table
M Shall we try it?
((*M moves toward the table*))
O Yeah, okay
((*O turns to face the table and moves in that direction*))
S Okay
((*S follows*))

Participants' attitudes

In order to get some insights into participants' own views of interacting in this environment, those taking part in the online discussions have been asked to fill out questionnaires. In the following, some of the answers received will be exemplified by relating them to two analytical concepts that are central to my thesis, namely Common Ground and Presence. I have chosen to include answers from three different people, to ensure that both attitudes representative of beginners and more accustomed users will be presented. Two of these participants are newbies, but with quite different experiences, and one is an extremely accustomed user of Traveler.

Common Ground

Common Ground theory deals with the ways in which people negotiate shared understanding, as regards both process and content of communication. It is established through *grounding*, a process by which participants in communication validate that they share a common understanding. This verification will take on varying forms depending on both the purpose of conversation and the media used. (Clark and Brennan 1991)

As we have seen in the previous section, negotiation for content and structure can be accomplished through several different modes. However, in my material, language is the most common mode used for this purpose. One possible reason for this is that most participants in these gatherings in fact are newbies, and whereas initially the employment of these strategies take an effort from the participants, after a while they appear to become internalized. To illustrate this I have chosen some answers which indicate how three of the participants in a gathering for language teachers think that visual cues influence their conversation management. A and B are newbies, whereas C is an accustomed user.

Table 4: Examples of replies (Common Ground)

Question:	**Did you use many non-verbal cues, and did you notice if others did? If so, do you recall on what occasions these were used? Did the non-verbal cues add anything, and when they were not used, did you miss them?**
A (beginner user)	"D and I played around with the emotions –angry, happy etc but we couldn't see much of a change. Also tried nodding. I think it would be essential when using this in a class, to go through the ways that you express non-verbal clues. I noticed that E seemed to zoom in and out and nodded which made it clearer how she was feeling and how she wanted to participate."
B (beginner user)	"I tried to use as many of the non-verbal affordances as possible – this was one of the main interests I had in joining the meeting and using Traveler. I used movement of the avotar, nodding agreement, smiling, change of position to face speaker, location to be inclusive, and reversing to indicate 'resting'. The non-verbal behaviour adds a new dimension to online communication and made a significant difference for me – though not all positive."
C (accustomed user)	"I often nod my head in agreement or bow in response to a "thank you." I think it adds the yes or no responses without interrupting. I can't say that anything is missing if others don't. It's just handy to respond without stopping somebody's train of thought."

Here we can see how these three participants have all given this aspect of the environment some thought, and how the two newbies have tried to use the non-verbal cues in an experimental way, whereas the accustomed user has started using the ones that he finds most functional in his interactions. By combining sound and those visual and spatial cues that can be used for feedback without interrupting, modal density is created. It should be added that this specific accustomed user also has been observed using emotes on a number of occasions where they have had a social/contextual function rather than a structural.

Presence

Closely related to Common Ground is the notion of Presence. As illustrated in the section on previous research, presence is a complex notion with many different definitions. In this paper, presence refers to a sense of sharing a space, which can either be accomplished by perceptual stimuli or by less technologically advanced techniques that nonetheless can cause a feeling of immersion, such as personal and intimate language use. In this environment, the sense of shared space has implications for the quality of interaction. The feeling of being present can for instance be seen in the use of deictic expressions and reference. Address is another indicator of presence – sharing a space like this makes it possible to address the whole group or parts of it with personal pronouns or even with spatial cues only, rather than with proper names. The type of multimodal representation which Traveler allows for might both make communication smoother and create presence through relating to a basic spatial communicative situation. The answers to the following question illustrate participants' experiences of presence when communicating in Traveler.

Table 5: Examples of replies (Presence)

Question:	**Did you feel as if you were transported to a place which you shared with the other participants, or did you think about how this all took place on your computer screen?**
A (beginner user)	"Yes and no-I was involved in the place but started to realize that I could be quite rude as well and check my mail or surf the net while people were talking."
B (beginner user)	"It took me to another world and was a real adrenaline buzz. It was on my screen and I was conscious of it always, but I was definitely virtually gone from my usual habitat. It took me a little while to come down again...."
C (accustomed user)	"I am always immersed. I throw my mind into the environment easily. It doesn't matter that the environment is artificial. My house is man-made too but I prefer it to a cave. I think of the place as real, even though I understand better than most the mechanics of how Traveler worlds are constructed."

The answer of B is representative of most answers to this question. Interesting to note is that both one of the beginners (B) and the accustomed user (C) express similar attitudes, since this indicates that the habituation effect has not made the experience less immersive for the accustomed user. Of course, there are probable side explanations to why the accustomed user expresses these attitudes, for instance the social aspect – C has made friends in Traveler and regularly meets with them here. The fact that A has been paying attention to other things while participating in the meeting might explain in part her low level of immersion. A has also expressed a dislike for the surreal avatars in answers to previous questions, which most likely will have had an effect on her experiences.

Most users in my material seem to be comfortable with the surrealistic avatars and environment, and instead the realism of the situation – people meeting in a shared space and discussing via voice – appears to have a great influence on their experiences. This indicates that personal relationships and involvement in combination with a realistic sense of shared space are the greatest generators of a sense of presence in this environment.

Conclusion

In my material, modal density is most often created through intensity. Especially beginner users, but also more accustomed ones mainly depend on the sound resources, and thus, sound carries the greatest modal density in this environment.

Interesting to note is what appears to be a lack of complexity between the different modes. Only on a few occasions is modal density achieved through complexity. I believe that this can be explained by the newbies' relative unfamiliarity with using the visual and spatial cues, and the more accustomed users' ability to select the most functional cues from the different modes. However, the visual mode is nevertheless important, since especially with larger groups involved, turn-taking, feedback on the content level as well as emotional support and face saving strategies will be smoother when not having to rely on sound only. Those who have learnt how to make use of the visual and spatial modes have a great advantage here.

One possible explanation to why participants sometimes express frustration when visual cues appear to be missing is that the situation in itself so closely resembles face-to-face interaction that beginner participants expect the same conventions to apply here as in face-to-face. Accustomed users, on the other hand, learn new conventions specific for this type of environment in the social context of communicating here, and thus learn to make the most of the modes available. One example of this is the way in which accustomed users are starting to employ the spatial modes from a functional perspective, for instance by maneuvering the avatars to show either general or specific address, or by moving them back and forth as part of communication management.

The fact that visual and spatial expressions in Traveler depend on deliberate actions on behalf of the user has consequences for grounding. Participants in interaction need to focus their attention and intentionally send out signals that will help in the grounding process. After having learnt the social conventions, also these expressions will become internalized, and only then can complex modal density help participants reach common ground.

Modal density is related to levels of attention, and it is through modal density that one shows situational and interactional focus. The ability to express level and focus of attention has implications for both negotiation of context, structure and content. Further, my findings indicate that by making it possible to detect the level of attention of the others, modal density influences the sense of presence that participants experience. Another strong generator of presence here is the way in which the three-dimensional environment creates an illusion of shared space.

In sum, these preliminary findings support the argument that when studying multimodality in CMC a focus on modal density may be more fruitful than one on perceptual realism, since this approach draws the attention to environment specific conventions and affordances rather than to face-to-face interaction. In order to compare the suitability of the modal density framework over other multimodal approaches more investigations need to be undertaken, especially so in different types of environments with different modal affordances.

References

Aston G. 1986. ‘Trouble-shooting in interaction with learners: the more the merrier’. *Applied Linguistics,* 7(2): 128-143.

Aston G. 1993. ‘Notes on the interlanguage of comity’. In Kasper G. & S. Blum-Kulka (eds.), *Interlanguage Pragmatics*. Oxford: Oxford University Press: 224-50.

Clark H.H. & S.E. Brennan. 1991. ‘Grounding in Communication’. In Resnick et al. (eds.), *Perspectives on Socially Shared Cognition*. The American Psychological Association: 127-149.

Daft R.L. & Lengel R.H. 1984. ‘Information Richness: a new approach to managerial behavior and organization design. *Research in Organizational Behavior*, 6: 191-233.

Duranti A. & C. Goodwin (eds). 1992. *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Gibson J. J. 1977. “The theory of affordances”. In R. E. Shaw & J. Bransford (eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goffman E. 1967. *Interaction Ritual: Essays in Face-to-Face Behavior*. Chicago: Aldine Publishing Company.

Gumperz J.J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Hutchby I. 2001. *Conversation and Technology. From the telephone to the internet.* Cambridge: Polity Press.

McIsaac M.S. & C.N. Gunawardena. 1996. ‘Distance Education’. In Jonassen D.H. (ed.), *Handbook of Research for Educational Communications and Technology*. New York: Macmillan Library Reference: 403-437.

McLellan H. 1996. ‘Virtual Realities’. In Jonassen D.H. (ed.), *Handbook of Research for Educational Communications and Technology*. New York: Macmillan Library Reference: 457-487.

Norman, D. A. 1988. *The Design of Everyday Things.* New York: Doubleday.

Norris S. 2004. *Analyzing Multimodal Interaction. A methodological framework.* New York & London: Routledge.

Short J., Williams E. & B. Christie. 1976. The Social Psychology of Telecommunications. New York: John Wiley & Sons.

Walther J.B. 1996. ‘Computer-mediated communication: impersonal, interpersonal and hyperpersonal interaction’. *Communication Research,* 23: 3-43.

Walther J.B. & M.R. Parks. 2002. ‘Cues filtered out, cues filtered in: computer-mediated communication and relationships’. In Knapp M.L. & J.A. Daly (eds.), *Handbook of interpersonal communication*. Thousand Oaks, CA: Sage: 529-563.

Weedon C. 1997. *Feminist Practice and Poststructuralist Theory.* 2nd edn. Oxford: Blackwell.

Biography

Therese Örnberg Berglund is a doctoral student at Umeå University, Sweden, affiliated with both the Department of Modern Languages/English (http://www.eng.umu.se) and the humanities computer lab HUMlab (http://www.humlab.umu.se). Her research project deals with emerging communication patterns, and she is also interested in questions to do with ICT and (language) education. Therese coordinates the online activities of a Swedish national network for ICT in academic language education, ITAS (http://www.humlab.umu.se/itas), and is the national representative for EUROCALL in Sweden (http://www.eng.umu.se/eurocall). She also has a teacher's degree in German and English for upper secondary school.

Visit Therese's blog, Emerging Communications, for more information: http://emergingcommunications.net.

Contact

Therese Örnberg Berglund
The Department of Modern Languages/English
Umeå University
901 87 Umeå
Sweden

+46 90 786 61 58

therese.ornberg@engelska.umu.se