How Different Views of Communication Influence HCI Design: The Example of Shaping

Kerstin Fischer
Institute for Design and Communication
Alsion 2
DK-6400 Sonderborg
+45 6550 1220
kerstin@sdu.dk

ABSTRACT
In this paper, I argue that implicit views of communication can influence human-computer interaction design such that they may inspire radically different design approaches, with very different outcomes. Using the concept of shaping as an example, I show how the assumption of automatic processes in communication may lead to a restricted view of shaping by interactive alignment, whereas a collaborative view of communication allows a much broader range of strategies to be taken, which can enrich the designer's possibilities to shape users' behavior significantly.

Categories and Subject Descriptors
D.2.2 [**Design Tools and Techniques**]: User Interfaces

General Terms
Design, Human Factors

Keywords
Communication theory, HCI design, Alignment, Shaping

1. INTRODUCTION
The notion of 'shaping' was introduced into human-computer interaction (HCI) by Zoltan-Ford (1991) and describes users' convergence with the linguistic material presented to them. If people's behavior in HCI could be shaped in this way, this would facilitate automatic speech processing considerably and make the HCI designer's life much easier. The question is thus how shaping can be employed in HCI.

Shaping has been addressed from various points of view. While the term 'shaping' implicitly encodes the designer's perspective where human users are subtly guided into particular behaviors, the mechanisms underlying shaping are also studied from psycholinguistic, sociological and psychological perspectives. In psycholinguistics, the phenomenon that speakers adjust to their communication partners is investigated under the label of 'interactive alignment' (Pickering & Garrod 2004), whereas in sociology, especially in ethno-methodological conversation analysis, as well as in cognitive and social psychology, the phenomenon is rather viewed as an instance of coordination.

12th Danish HCI Research Symposium (DHRS 2012), November 21, 2012, Sonderborg, Denmark.

The ways in which the two views on shaping differ is related to the amount of automaticity assumed to be involved in communication; while in the first tradition adjustment to the partner is taken to be based on automatic, involuntary processes, in the other position it is taken to be based on strategic choice. These two schools of thought thus hold radically different views of what shaping, i.e. the adaptation to the partner's behavior, is caused by. In this paper, I argue that which perspective is taken has considerable consequences for HCI design.

2. TWO VIEWS OF COMMUNICATION
In the interactive alignment model, communication happens in part based on automatic responses to the communication partner's behaviors. In particular, people are taken to pick up linguistic material from their communication partners' utterances as a result of automatic priming. Alignment in this view is assumed to be based on automatic and subconscious responses to the partner's utterances. Because alignment is taken to be automatic and subconscious, it does not require a model of the listener (Garrod & Pickering 2007: 444). Speakers thus do not take their listeners into consideration unless there are problems or unless "the discrepancy between their knowledge and that of the speaker is made especially salient" (Garrod & Pickering 2007: 445). Explicit, as well as implicit, non-alignment is thus also possible, for instance when speakers try to conceal information, when they wish to disalign deliberately or when the previous representation was not understandable. Such strategies are however taken to be cognitively demanding and therefore to be rather exceptional.

The view that communication rests at least partly on automatic, involuntary processes concerns also other areas of communication, such as considering the communication partner's perspective and access to information in spatial perspective taking (e.g. von Stutterheim & Kohlmann 1998) or when referring to objects (e.g. Horton & Keysar 1996). In these approaches, speakers are assumed to make egocentric choices unless they have extra time or a particular reason to consider the partner's knowledge and perspective. Another approach that assumes automatic mechanisms is the computers-are-social-agents paradigm, which suggests that speakers transfer mindlessly from human communication to interactions with non-social communication partners, such as computers or robots (see Reeves & Nass 1996, Nass & Moon 2000, Nass 2004, for example).

In contrast, the collaborative view of communication holds interlocutors to be involved in constant implicit negotiation. In this view, communication is collaborative from the beginning (and orderly at all points, cf. Sacks 1984); thus, people will consider their communication partner already during early utterance planning (see Brown-Smith 2009). Correspondingly, in this view, people's mental models of their partners play a crucial role. Furthermore, linguistic labels are taken to be implicitly negotiated and collaboratively achieved (Clark & Wilkes-Gibbs 1981, Brennan & Clark 1996), perspective taking is carried out with addressees in mind (Schober 1995), and the addressees' knowledge and access to information are taken into account from the first moment of planning onward (Hannah et al. 2003).

Alignment in the collaborative view is accordingly a partner-oriented strategy rather than an automatic response. For instance, in the maze game studies reported on in Mills & Healey (2008: 49), the authors argue that alignment may be strategically employed in order to create a background against which an element to be corrected may be identifiable. Their findings suggest that alignment is used strategically for particular communicative purposes and is thus not due to automatic responses (see also Schegloff 2004).

Alignment can indeed be related to strategic purposes; the studies carried out in the framework of Communication Accommodation Theory (Giles, Coupland & Coupland 1991) show that speakers' interactional goals and identity needs play a considerable role in speakers' decisions to re-use linguistic material from their partners. Furthermore, Fischer & Wilde (2005) found that speakers only aligned to linguistic material presented to them if it fit their concept of their addressee; in particular, they aligned with the robot's spatial descriptions because they expected the robot to be competent in this area, but not with a nonce lexical item, because they considered themselves more competent concerning natural language terms. This finding is in line with findings by Kraljic, Samuel & Brennan (2008) who find participants to align only with phonetic peculiarities of the communication partner when these constitute a characteristic trait of their communication partner, and not a contingent, accidental effect of the speaker's pronunciation. Thus, speakers may select to which linguistic features of the communication partner they align their utterances depending on their model of the communication partner, which speaks against automatic priming as the main causal factor.

To sum up, in the collaboration view, alignment is just one out of many strategies for cooperation, and shaping consists of guiding users subtly into appropriate representations of their artificial communication partner that help them choose behaviors that are adequate for the particular situation and its affordances. One way to do so may then be to present users with vocabulary or linguistic structures to pick up themselves, but in fact the scope of shaping is much broader in the collaborative view of communication.

3. CONSEQUENCES FOR HCI DESIGN

In the two views of communication, shaping would be approached quite differently.

3.1 Shaping in the automatic alignment view

In the interactive alignment model, shaping would be done by presenting linguistic features as clues that the communication partner is then intended to pick up automatically and subconsciously. Evaluation of successful shaping from the priming perspective then consists in counting the number of words and structures in which output by the system and input from the user are coordinated.

Several researchers have tried to shape users' linguistic behaviors by presenting them with linguistic material that the system can understand. The first study in this respect is Zoltan-Ford (1991); however, she finds alignment only for at most 51% of the system utterances, and in the conversational condition, in which the simulated computer produced complete natural language utterances, the amount of alignment found was only 35%.

Similarly, many of the studies on alignment show that not all speakers align. Assertions like "72-94% of the children showed positive accommodation on the different linguistic features examined"' (Oviatt, Darves & Coulston 2004: 16) do not only show that alignment occurs, but also that 6-28% of the children did not align with their (artificial) communication partner. If alignment occurs as an automatic priming effect, it remains open why only some and not all speakers align with their partners.

Furthermore, recent studies have identified a number of factors that influence the amount of alignment occurring; for instance, Branigan et al. (2007) show that people align to different degrees with their communication partners depending on their speaking roles (addressees versus overhearers), and Branigan et al. (2011) demonstrate that the amount of alignment depends on people's understanding of the capabilities of their partner: people align more if they believe their partner to be a non-native speaker and they align more with a basic than with an elaborate computer.

Thus, shaping from the perspective of automatic, priming-based alignment has to content itself with presenting linguistic material to the user, hoping that users will pick this material up and re-use it. Since Zoltan-Ford's (1991) initial study, however, no study has been able to report better numbers than hers (see Baber et al. 1997, Tomko & Rosenfeld 2006). Thus, there seem to be limits to the effectiveness of shaping from an automatic priming perspective.

3.2 Shaping in the collaborative view

Shaping from the collaborative perspective offers further possibilities than priming users with vocabulary and linguistic structures. In the collaborative view, participants will build up a partner model and take this into consideration throughout. From this perspective, the users' behavior can be shaped not only with respect to selected linguistic features, but with respect to their understanding of the task and their partner's strengths and weaknesses. Shaping thus concerns not only a careful selection of words for the user to pick up, but also choosing its behaviors, appearance and utterances in a way to allow the user to build up an appropriate mental model of the system. The methods available for shaping in the collaborative view are thus:

- presenting the user with linguistic material to make use of;
- presenting the user with linguistic material that matches the general competences of the system;
- presenting the user with other behavioral or visual cues that allow him or her to build up a coherent mental model of the system.

Thus, the collaborative view takes speakers' general need to build up a coherent model of their communication partner into account, which is particularly important in HCI since here people do not have a very accurate view of their communication partner (see Fig. 1).

Figure 1: In HCI, the communication partner is literally a black box!

A measure for successful shaping in the collaborative view concerns, for instance, users' understanding of the task as measured by the number of out-of-domain vocabulary, the size and appropriateness of the vocabulary used, as well as users' judgments of the naturalness and fluency of the interactions.

4. A CASE STUDY

In this section, we explore shaping from a collaborative perspective. In order to study the impact of robot utterances on users' behavior, we compare interactions with a robotic wheelchair (Lankenau & Roefer 2001) in two conditions that differ regarding whether or not the robot produces verbal behavior; in both conditions, participants (nine native speakers of English in condition 1 and eleven in condition 2) had to carry out the same four tasks. The first task, which is the one reported on here, was to steer the robot around in order to 'train' it on the environment and to provide it with verbal explanations, in particular to familiarize the robot with locations in a room for handicapped people by driving it to particularly interesting locations and labeling them (see Fig. 2).

Figure 2: The robotic wheelchair 'Rolland'

Participants were free to move to as many locations as they considered relevant. There were no behavioral instructions. The robot was supposed to move autonomously only at the end of the instructions when it was meant to take the user to one of the locations it was previously trained on.

All verbal robot output in condition 2 was scripted and manipulated by a human 'wizard' hidden behind a flexible wall. Thus, for each location the respective participant steered the robot to, there was a set of robot utterances to be played in a particular order. For some utterances, the wizard had different choices depending on the label the participant had used, for instance, *sofa* versus *couch*, *fridge* versus *refrigerator*, *stove* versus *hot plate*. While this procedure may seem unnatural, the resulting dialogs are in fact quite fluent, and participants were found to find the interactions to be very enjoyable (cf. Andonova 2006). Scripting the robot output does not only render all robot output identical and thus the dialogs comparable across participants, it is also computationally the cheapest method possible. It should thus be impossible to discard the results of this study on the basis of the assumption that the dialogs used necessitate unrealistically sophisticated speech technology.

The robot utterances were designed in order to subtly guide users into appropriate understandings of the task and the capabilities of the robot. We applied the following four criteria:

1) We used everyday vocabulary, in order to prevent people from thinking that they have to talk in extra-ordinary ways to the robot.

2) We made sure that the robot used the terms consistently (cf. Zoltan-Ford 1991).

3) We provided implicit cues to the task; in particular, we had the robot announce its 'readiness' after the greeting by saying 'you can take us now to a place you want to name'. Furthermore, if the participant was driving the wheelchair without talking, the robot would ask 'where are we going to'. These two utterances serve as implicit clues to the task to label relevant locations in the room for the robot.

4) Since in conversation between humans explicit signs of understanding are very rare (see Heritage 1984), we provided only implicit feedback in the form of 'relevant next contributions' (Sacks et al. 1974, Clark & Schaefer 1989), in particular in the form of clarification questions designed to elicit further information relevant for the task given.

All interactions were recorded and transcribed. The transcripts were analyzed semi-automatically (using simple shell scripts).

Table 1: Results for the two conditions

	participant	total number of words	different words	out-of-domain words	aligned words
1	R024	443	160	100	32
	R026	223	85	27	17
	R030	1019	216	147	46
	R031	1499	301	209	49
	R033	67	33	11	2
	R035	359	140	54	28
	R036	503	149	67	27
	R037	873	262	162	52
	R039	1154	335	205	54
2	RD004	804	139	23	27
	RD013	644	148	30	33
	RD016	638	124	22	30
	RD017	545	143	31	28
	RD022	870	185	64	45
	RD025	740	175	54	30
	RD041	1763	416	234	64
	RD043	1193	233	91	59
	RD047	454	122	24	33
	RD048	642	172	53	31
	RD052	1114	212	68	48

The results show considerable linguistic variability in condition 1, the baseline condition, for instance with respect to the labels or the instructional strategy employed. The high variability observed is reflected quantitatively in the number of different words used in this condition (see Table 1). Concerning alignment, we can observe that participants in condition 2 make use of 27-64 items the robot had previously used.[1]

[1] The overlap in condition 1 is however due to the fact that the vocabulary used there was used for the design of the robot utterances in condition 2.

As Table 1 shows, condition 1 is furthermore characterized by considerable amounts of out-of domain vocabulary. In contrast, in condition 2, people use far fewer such terms, with the exception of participant RD041; in the collaboration view of communication, such interpersonal differences are expected since people's own agendas and partner models play a crucial role in linguistic choice.

To sum up, our findings suggest that users in the baseline condition were not very focused on what the robot could possibly understand, while in condition 2, participants generally exhibited a much better understanding of the task.

5. DESIGN IMPLICATIONS

The current study has shown that the view of communication implicitly embraced may have a considerable impact on the HCI design approach taken; depending on the model of communication assumed, the designer has more or fewer strategies for shaping the users' behavior at his or her disposal. In the current study, we have seen that the collaborative view of communication provides the designer with a bundle of possibilities, several of which turned out to be successful for shaping users' linguistic behavior. Especially providing users with implicit clues to task understanding turned out to be highly effective. In contrast, alignment based on automatic priming was found to be as limited in effect as reported on in previous studies (e.g. Zoltan-Ford 1991, Baber et al. 1997, Tomko & Rosenfeld 2006). Thus, with respect to shaping, the collaborative view of communication produces better results.

6. ACKNOWLEDGMENTS

This research was partly carried out during research visits at Stanford University funded by a research award from the BHJ-Fond. The data reported on were elicited in the framework of the Special Research Area SFB/TR8 'Spatial Cognition,' funded by the German Research Foundation (DFG).

7. REFERENCES

[1] Andonova, E. 2006. On Changing Mental Models of a Wheelchair Robot. *Proceedings of the Workshop on 'How People Talk to Computers, Robots, and Other Artificial Communication Partners'*, Hansewissenschaftskolleg, Delmenhorst, April 21-23, 2006, pp. 131-139.

[2] Baber C., Johnson, G.I. and Cleaver, D. 1997. Factors Affecting Users' Choice of Words in Speech-based Interaction with Public Technology. *International Journal of Speech Technology* 2: 45-59.

[3] Branigan, H., Pickering, M., McLean, J. and Cleland, A. 2007. Syntactic Alignment and Participant Role in Dialogue. *Cognition* 104: 163-197.

[4] Branigan, H. Pickering, M., Pearson, J., McLean, J. and Brown, A. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121: 41–57.

[5] Brennan, S. and Clark, H.H. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22: 1482-1493.

[6] Brown-Schmidt, S. 2009. Partner-Specific Interpretation of Maintained Referential Precedents during Interactive Dialog. *Journal of Memory and Language* 61: 171-190.

[7] Clark, H.H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22: 1-39.

[8] Clark, H. H. and Schaefer, E. F. 1989. Contributing to discourse. *Cognitive Science* 13, 259-294.

[9] Fischer, K. 2011. How people talk with robots – Designing Dialog to Reduce User Uncertainty. *AI Magazine,* 32(4): 31-38.

[10] Fischer, K. and Wilde, R. 2005. Methoden zur Analyse interaktiver Bedeutungskonstitution. In: Solte-Gresser, C., Struwe, K. and Ueckmann, N. (eds.): Von der Wirklichkeit zur Wissenschaft. Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften, pp. 163-174. Hamburg: LIT-Verlag.

[11] Garrod, S. and Pickering, M. 2007. Alignment in Dialogue. In: *The Oxford Handbook of Psycholiguistics*, pp. 443-451. New York: Oxford University Press.

[12] Giles, H., Coupland, J. and Coupland, N. (1991), *Contexts of Accomodation. Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

[13] Hanna, J.E., Tanenhaus, M.K. and Trueswell, J.C. 2003. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 49, 43-61.

[14] Heritage, J. 1984. A change-of-state token and aspects of its sequential placement. In: Atkinson, J. & Heritage, J. (eds.): *Structure of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

[15] Horton, B. and Keysar, B. 1996. When Do Speakers Take into Account Common Ground? *Cognition* 59, 91-117.

[16] Kraljic, T., Samuel,A.G. and Brennan, S.E. 2008. First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science* 19: 332-338.

[17] Lankenau, A. and Roefer, Thomas 2001. A Safe and Versatile Mobility Assistant. *IEEE Robotics and Automation Magazine* 7: 29-37.

[18] Mills, G. and Healey, P. 2008. Negotiation in Dialogue: Mechanisms of Alignment. *Proceedings of the 8th SIGdial workshop on Discourse and Dialogue.*

[19] Nass, C. 2004. Etiquette Equality: Exhibitions and Expectations of Computer Politeness. *Communications of the ACM* 47, 4: 35-37.

[20] Nass, C. and Moon, Y. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1, 81-103.

[21] Oviatt, S.L., Darves, C. and Coulston, R. 2004. Toward Adaptive Conversational Interfaces: Modeling Speech Convergence with Animated Personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*.

[22] Pickering, M. & Garrod, S. 2004. Towards a Mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences* 27: 169-225.

[23] Reeves, B. and Nass, C. 1996. *The Media Equation.* Stanford: CSLI and Cambridge: Cambridge University Press.

[24] Sacks, H. 1984. Notes on Methodology. In: Atkinson, J. & Heritage, J. (eds.): *Structure of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

[25] Sacks, H., Schegloff, E.A. & Jefferson, G. 1974. A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language* 50: 696–735.

[26] Schegloff, E.A. 2004. Putting the Interaction Back into Dialogue. (Commentary on Pickering & Garrod). *Behavioral and Brain Sciences* 27: 207-208.

[27] Schober, M. 1995. Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes* 20, 2: 219-247.

[28] Tomko, S. and Rosenfeld, R. 2006. Shaping user input in Speech Graffiti: A first pass. *Proceedings of the ACM Conference on Human Factors in Computing Systems*.

[29] von Stutterheim, C. and Kohlmann, U. 1998. Selective Hearer Adaptation. *Linguistics* 36, 3: 517-549.

[30] Zoltan-Ford, E. 1991. How to Get People to Say and Type what Computers Can Understand. *International Journal of Man-Machine Studies* 34: 527-647.