The Error Is the Clue:
Breakdown In Human-Machine Interaction

Bilyana Martinovsky

Institute for Creative Technologies
University of Southern California
Marina del Rey, CA 90292
martinovski@ict.usc.edu

David Traum

Institute for Creative Technologies
University of Southern California
Marina del Rey, CA 90292
traum@ict.usc.edu

Abstract

This paper focuses not on the detection and correction of specific errors in the interaction between machines and humans, but rather cases of massive deviation from the user's conversational expectations and desires. This can be the result of too many or too unusual errors, but also from dialogue strategies designed to minimize error, which make the interaction unnatural in other ways. We study causes of irritation such as over-fragmentation, over-clarity, over-coordination, over-directedness, and repetitiveness of verbal action, syntax, and intonation. Human reactions to these irritating features typically appear in the following order: tiredness, tolerance, anger, confusion, irony, humor, exhaustion, uncertainty, lack of desire to communicate. The studied features of human expressions of irritation in non-face-to-face interaction are: intonation, emphatic speech, elliptic speech, speed of speech, extra-linguistic signs, speed of verbal action, and overlap.

1. Introduction

One model of spoken dialogue systems is that of conversational partners, able to use the modality of speech and conventions of natural dialogue to communicate. This model relies on the spoken dialogue competence that dialogue system users have built over a lifetime of interaction with other humans, and hopes for a willing suspension of disbelief on the part of the user to treat the computer system as a fully capable conversational partner, as well. Despite obvious differences between language processing abilities of humans and machines, this approach seems quite promising, given findings such as Reeves' and Nass' Media Equation that people respond to computers as if they were humans [1]. While current spoken language technology is quite error-prone, this is not necessarily a problem, since human dialogue also contains errors. What we are concerned with in this paper is not the detection and correction of specific errors, but rather cases of massive deviation from the user's conversational expectations and desires, such that the user is "thrown" [2] out of the suspension of disbelief and feels she is interacting with a

"stupid machine" rather than a competent conversational partner. Note that this is a very different measure than task completion or efficiency. It may be possible to complete a task (at least under some definitions) once the dialogue has become "unnatural". However, we feel that these dialogues are still sub-optimal, given extra stress on the user (having to actively "psycho-analyze" the actual capabilities of the system, rather than being able to effortlessly conform to familiar dialogue conventions. These problems are especially acute in cases where the user refuses to go on with what she perceives to be a farcical or impossible situation. Such breakdowns may also lead to increased reluctance to interact with these systems in the future. By breakdown we mean a specific point in a conversation when the interaction is interrupted with or without completion of the performed task because one or both parties give up the conversation. Before this point is reached there are breakdown symptoms, small incidents of dissatisfaction which lead to the final breakdown. By tracking those smaller incidents and studying the causes we hope to improve the co-operation in human-machine interaction.

In contrast with the established dichotomy of extremes, where communication is either smooth or erratic, we believe that 1. communication doesn't have to be smooth; 2. communication is **not** smooth; 3. errors are not always to be avoided; they can be used as indicators of state-of-mind changes which improve the cooperativeness of the system, and, moreover, can also serve educational purposes, forcing the speaker to think constructively about the topic of conversation. In this paper we will give examples of situations where the communication gets out of control and we will examine what causes the breakdown and how it can be avoided. We will also show examples where the non-fluency of the human-machine interaction can be overcome without breakdown. The idea is to build systems with communicative skills that inspire the human users to desire to cooperate rather then force them to adapt to the 'machine talk'.

The paper starts with two examples of miscommunication, which illustrate that fluency of communication has different aspects and that sometimes even non-fluent interaction may be positive and immersive.

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE **The Error Is the Cue: Breakdown In Human-Machine Interaction**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California,Institute for Creative Technologies,13274 Fiji Way,Marina del Rey,CA,90292**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The next three sections study causes, signals and remedies of breakdowns of communication, examples of breakdowns. The paper concludes with a more broad discussion on the character of human-machine interaction and a summary of the study.

2. Non fluency and cooperation

Given the current state of natural language processing technology, it is impossible to have normal conversations without errors. There are several approaches that can be taken to deal with this reality. First, one might carefully engineer the dialogue to minimize errors as much as possible, e.g., by allowing a user to only answer very simple yes/no or alternative set questions, with plenty of confirmations. A second strategy is to use very strong expectations and essentially ignore the input, and hoping that the context is strong enough to make likely interactions fluent. Yet a third option is to allow errors to crop up and deal with them in human-like manners when they do. The problem here is that there may be so many that a non-fluent interaction results. Each of these approaches has merit, but each also has potential pitfalls. We give examples of each, taken from two human-computer activities: TOOT, which consists of telephone conversations between clients and an automated train booking system [3], and MRE, which consist of video/audio interactions between a person and multiple virtual humans [4].

In this section we present two examples of non-fluent communication from the MRE domain. In this domain, a human trainee plays the part of an Army Lieutenant in charge of a peace-keeping operation. We have built several systems in this domain, including pure scripted demo systems, wizard of Oz systems, finite-state driven systems, and full multi-agent interaction systems. The first example is from a finite-state dialogue model, meant to keep a user as close as possible to a given script. The transitions control dialogue from all the artificial characters as well as recognitions and defaults from the trainee Lieutenant. Here is an excerpt from one interaction, after which the Sergeant character has already failed to carry out an order to secure a landing zone for a medevac helicopter.

Example 1: MREbox12
1. LT: sergeant. **send one squad to recon forward** on the route toward eagle one six while we wait for the medevac. <silence>
2. Sgt: sir, should we set up a landing zone now to prepare for the medevac?
3. LT: sergeant, I already told you to set up a landing zone for the medevac.
4. Sgt: yes sir. sergeant duran, get your team up to the square and **secure an lz.** …
5. E1-6: two six, it's urgent you get here right now. situation's getting critical. we're taking fire. over.
6. LT: eagle one six, this is eagle two six, we will be at your position as soon as possible. over.
7. E1-6: two six, this is one six. did you read me? you're breaking up. we need your help put here. over.
8. LT: **sergeant, send a squad to recon forward the route to eagle one six.**
9. E1-6: roger two six. I know you'll get here as soon as you can. we'll be waiting. one six out.
10. LT: sometimes it doesn't matter what I say.

11. Sgt: what should we do about one six if we need to wait here for the medevac?
12. LT: sergeant, I am assuming that is a rhetorical question.
13. Sgt: sir, we should keep our mass here to maintain a secure area for the medevac. **I suggest we send one squad to recon forward**, though, to give us a secure route to one six.
14. LT: sergeant, send one squad forward to recon for one six.
15. Sgt: fourth squad. mount up.
16. Mom: what happen? you are going?
17. LT: yes, we're leaving and your boy will probably die.
18. E1-6: two six how much longer before we can expect assistance? over.
19. LT: eagle one six, this is eagle two six. We'll be there once we finish our lattes. over.

Although the interaction is robust in the sense that the virtual humans move forward until a conclusion is reached the interaction is not successful. On line 2, the system asks a question rather than reacting directly to a given order, and moreover asks about an order that had been previously given and as a result the user reacts on line 3 with a reminder. This is the first indication of irritation, which at this point is cooperative and doesn't harm immersion. The missing order is now performed on line 4, however notice that the order on line 1 is still not recognized. On line 8, the Lieutenant repeats that order, which indicates his continued cooperation and immersion since it indicates that he has noticed that the order is not performed, he is not blaming the system, but simply repeats the order. The order is again not recognized, in this case the wrong agent answering, breaking the illusion of conversation. At this point the user feels his efforts to communicate have been futile and he displays his frustration on line 10. This line is interesting because it indicates a point in the conversation when the user feels alone, feels that there is no communication, no results of his efforts to interact in the given environment. This is a point of a breakdown, a display of lost sense of immersion. If this were said to a human, the person would most likely take up the conversation at a meta-level, perhaps apologizing, or dealing with the perceived failure of communication. Instead the system simply carries on as normal. When the system asks for instructions that have already been given twice, the user responds on line 12 rather cooperatively. Now the system suggests what has previously been given as orders and the user repeats the order, seemingly cooperatively. However, when faced with a question by the mother of the injured boy, who he has been trying to save, her response on line 17 is sarcastic rather than serious, because despite the miscommunication at that point all orders have been performed and the boy is obviously going to be saved. On line 19 the response is clearly ironic, although it is framed in the required radio call sign format. The user has given up trying to have a serious interaction, and is trying to amuse himself rather than participate in a serious problem-solving dialogue.

The second example, (Example 2 below) is from an early version of a system including autonomous agents for some of the characters> While this interaction is not smooth, it keeps the immersion alive and avoids a cascade of miscommunications, which could lead to complete

breakdown. From an educational point of view it is also preferable not to avoid all misunderstanding because such situations may be instructive with regard to the activity. For instance, it is preferable to be explicit than vague and by being forced to reformulate one learns (admittedly the hard way) to be more exact in giving orders. In this case the first few orders are mis-recognized by the speech recognizer. Later the vague instruction is understood as ambiguous, possibly referring to securing a landing zone rather than the assembly area (the Sgt did not take the context of his suggestion into account in disambiguating the order).

Example 2: MREmay1-3
1. Sgt: sir   we   should secure thee assembly area
2. LT: sergeant, let's set up security ? three sixty perimeter .
3. Sgt: say again sir
4. LT: sergeant. set security .
5. Sgt: say again sir
6. LT: set security. sergeant ,
7. Sgt: thee lz is not secure sir
8. LT: <set security sergeant, > **<higher pitch>**
9. Sgt: secure what
10. LT: **sergeant**. set security .
11. Sgt: say again sir
12. LT: **put your men on security. sergeant,**
13. Sgt: I don't know which act you want me to do
14. LT: secure the l-z- sergeant,
15. Sgt: sir first we should secure thee assembly area
16. LT: secure the assembly area.
17. Sgt: understood sir. squad leaders listen up!

 The utterances on line 1 and 15 are very similar, and what transpired between them is an example of continuous tedious miscommunication, which finds a resolution. This kind of miscommunication may happen also between human speakers under certain conditions. Although there are errors thanks to the cooperative updates of the system of its own 'state of mind' the sense of immersion is not completely lost and the user has not lost desire to continue to cooperate.

These two examples show that fluency is not necessarily indication of cooperation and immersion and that it can be better to update the user with the state of mind of the system than to simulate communication and thus lead him/her far beyond the necessary level of believability and cooperation. Thus one way of improving human-machine interaction is not to eradicate all possible errors but to prepare the system for non-fluency and for state-of-mind updates. The other path of improvement is to study the actual causes and signals of breakdowns and prepare the systems to recognize them and in that way become more sensitive to the state-of-mind-and-emotion of the users. That is, both paths of improvement suggested here concern increase of grounding efforts by means of more and better opportunities for monitoring of the other.

3. Causes of Breakdown

The following text from the TOOT Corpus exhibits a clear example of growing irritation due to linguistic or rather pragmatic and dialogical inadequacy (A – automated Amtrak voice; C - human client on the phone; the seconds indicate pauses between utterances; <> - descriptions of the tone of voice of the last speaker):

Example 3: first query
1. A: alright this is the amtrak train schedule system, which city do you want to go to,
2. C: <boston>
<positive, breathing>
9 sec pause
3. A: what city do you want to go to boston from,
4. C: <washington dc>
<tired>
5 sec pause
5. A: on what day of the week do you want to go to from washington dc?
6. C: thursday through three thirty pm.
4 sec pause
7. A: what time do you want to leave on thursday?
8. C: <three thirty pm>
<falling disappointed intonation>
4 sec pause
9. A: do you want me to find the trains going to boston from washington dc on thursday around three thirty pm now?
10. C: <YES:>
<emphatic, impatient, tired>

This extract is only the first query out of three unsuccessful queries by the same user. A simple question about 'from where to where' is divided into two different utterances, which have identical, repeated structure with a long pause in between. Although such explicit formulation of questions into two different turns and clear repetitive interrogative structures avoids misunderstandings it adds an unnatural feature to the interaction, namely lack of pragmatic adaptation. Adaptation is the process in which one speaker adapts to the situation, to the other speaker(s), to the flow of the talk. Adaptation is expressed in many different ways, such as use of pronouns and other indexicals instead of full names, none-repetitiveness of sentence structure (such repetition is usually marked in the sense that it expresses additional attitude), in certain activities such as radio talk, telegraphic highly elliptic speech, etc. So, from line 1 to 4 above we have fragmentation of a common question, non-adaptive over-explicit formulation, repetitive intonation and tone of voice, and a long pause. Even if the pause is reduced, the other three features will contribute to a sense of unnaturalness and cause irritation. The same features are repeated on lines 5 to 8 and the frustration of the human user is now close to complete lack of desire to interact. At this point the user has no illusions that a 'real communication' is possible, that s/he has a communicator on the other side. On line 9 the system formulates a fully coordinated repetitive structurally and intonation-wise summary of the otherwise carefully fragmented request and on top of that asks a 'stupid' question, which completely breaks the pragmatic assumptions of the purpose of the activity. The question is 'stupid' because the whole purpose of the activity is to find a train so asking if one wants to do that after a long fragmented interrogation is overwhelmingly unnecessary, even if it is a nice safeguard for making sure the previously understood information is correct. The gradation of irritation has now reached a higher level and accordingly the human user starts rising voice and displaying obvious dissatisfaction. However, the system doesn't pick up on this, it just gets an answer but is completely insensitive to the state of mind of the user.

The causes for communication breakdown (besides speech recognition issues) we noticed in both corpora consist of features such as long pauses, over-fragmentation, over-clarity, repetitiveness of verbal action, syntax, and intonation, over-coordination, and over-directedness, general lack of pragmatic adaptation, lack of, insufficient or exaggerated state-of-mind updates and repair requests.

4. Signals of breakdown

The human user expresses disappointment and/or simply tries to cope with the small or more noticeable breakdowns before giving up. Although the users are aware that they communicate with a machine they are more optimistic at the beginning than in the end of the interaction. Characteristically the user notices that although what s/he says affects the talk, s/he is alone, there is no real communication partner because there are common sense discursive habits and regulations, which are violated and because the interactant is not sensitive to the state-of-mind clues given by the user (see example 1 above). As a result the main purpose of communication, namely exchange, but not only of concrete information, is lacking. Many of these clues are typical of human-human dialogue as well, however, people are good at recognizing interactive problems and adjust behavior accordingly (if cooperative). Reactions to failure in communication may appear in stages: tiredness, tolerance, anger, confusion, irony, humor, exhaustion, uncertainty, lack of desire for communication. The studied features of human expressions of non-face-to-face interaction irritation are: intonation, emphatic speech, elliptic speech, hyper-articulation, extra-linguistic signs, and vocatives.

4.1. Intonation

The intonation of each utterance can be a clue for the state of mind of the human user. By studying typical expressions of irritation, impatience, anger, frustration, irony one may prepare the dialogue system to react to such expressions, rather than attending only to the content information. In example 3 above line 8 illustrates falling disappointed intonation as a reaction to over-fragmentation and line 10 carries emphatic, impatient, tired intonation as reaction to violation of Grice's quantity maxim. The rising tone of voice contributing to the emphatic effect is also a result of the preceding disappointment on line 8, i.e. the irritation is escalating quickly especially when the inadequacies are one after the other. That is, the architecture of the tasks, responses and actions should have in mind the effect of repeated over-fragmentation and avoid it.

Although the human user is aware of the limitations of the communication with the system at the third query, she is still expressing emotions and additional attitudes:

Example 4: Last query
56. A: there are two trains available around one pm. the first train leaves at twelve pm every day and it take ten hours ten minutes, the second train leaves at one ten pm from sunday to friday and it takes eight hours ten minutes <do> you want to continue and find a new set of trains?
<goes on directly, no pause at all>
2 sec pause
57. C: no:
<uncertain, tired>
58. A: I didn't hear anything from you:! say yes to find a new set of trains, or no to end the dialogue,
59. C: no // I am done here
<moody, tired>
<hangs up>
2 sec pause
60. A: thank you for using the ntnt amtrak system. see you next time.

At line 57 after four failures the user is confused, lost energy and at line 58 we have reached the final breakdown. The user could very well just hung up, not follow the instructions to end the dialogue with words, but the obligative function [5] [6] of communication still works, despite the breakdown. In fact, the user continues to add expressions of emotional attitudes, despite the fact that she is aware that they have no effect on the functioning of the system nor on the result of the search.
The changes of the mood of the user are indicated in below. They don't develop linearly. The first query consists of utterance 1 to 11, the second query 12-16, the third attempt for booking is between 17-28.

The first mentions are always much more cooperative, have more stable pitch and intensity, observe utterance 1 versus utterance 13 and 18, when the user repeats again and again her request of destination, namely Washington DC and the pitch decreases. Utterance 18 and in fact almost all following utterances by the user are in complete overlap with the speech of the system, which is another indication of failing communication and lack of immersion. Similar decrease in pitch is observed between utterances 2 and 3, where 3 is an immediate repetition due to over-fragmentation. Since the expression of the time of return twice caused failure in task accomplishment in utterance number 20 and 24 the user prevents further problems by hyper-articulation and empathic speech, indicated by the higher and more varying pitch.

Emphatic speech is the first expression of impatience in the TOOT example as well as in the MRE
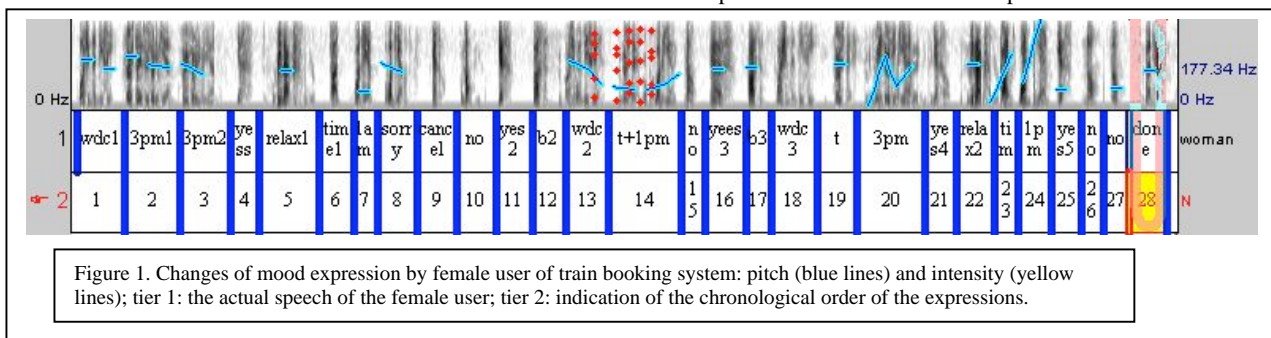

Figure 1. Changes of mood expression by female user of train booking system: pitch (blue lines) and intensity (yellow lines); tier 1: the actual speech of the female user; tier 2: indication of the chronological order of the expressions.

Example 2. The first 'yes' in TOOT example has prolonged final consonant, which contributes to the impression of impatience, line 9 in Example 5 below, utterance 4 in Figure1.

Example 5: First query
8. A: do you want me to find the trains going to boston from washington dc on thursday around three thirty pm now?
9. C: <YES:>
<emphatic, impatient, tired>
Second query:
34. A: do you want to continue and find a new set of trains?
35. C: <ye:s>
 <irritated, low voice>
Last query:
44. A: do you want me to find the trains going to boston from washington on thursday round [three thirty] pm now?
45. C: <[yes]>
<quick>
3 sec
…
53. A: do you want me to find the trains going to boston from washington dc around pm now?
54. C: <[yes]>
<funny voice>

The frustration increases with the initiation of the next query indicated by the low voice, lower pitch and prolonged vowel on line 35 above, utterance 16 in Figure1. After two unsuccessful attempts to book a train the user tries one more time. At that point she has passed the stages of disappointment, irritation, anger, and has now come to a point of irony, as we noticed also in the MRE Example 1 in Section 2 above. The irony being a result of the increased irritation of the inadequately repetitive structure of the dialogue but also of the realization of the 'aloneness' in this communicative situation, which is signaled by the increased overlap, i.e. the user's attempts to shorten the procedure by creating short cuts expressed in overlaps, which are not sanctioned as they would be in human-human interaction, see line 45 and line 54 above.

4.2. Elliptic speech

Another symptom of upcoming breakdown or failure in human-machine verbal communication is elliptic speech. In this category we include not only syntactic ellipsis but also ellipsis on turn-taking level and on speed of speech level. The more the human user notices that his/her interaction partner does not read his/her attitude clues and/or violates adaptation rules or Grice's maxims of cooperative conversation the more elliptic his/her speech becomes on all levels, as we noticed in Example 5 above and as in Example 6 below.

Example 6: Second query
20. A: I can't understand you [please repeat]
21. C: [no]
…
26. A: what city do [you want to go to boston] from
27. C: [washington dc]
4 sec
28. A: on what day of the week [do you want] to boston…
29. C: [thursday]

3 sec
30 A: what time do you want to leave [on thursday] ?
31. C: [one pm]

4.3. Hyper Articulation and Extra-Linguistic Signals

Hyper-articulation is a typical measure used by the human trainees in the MRE. The testers of this system are very cooperative and hyper-articulation is expression of cooperation and adaptation to the needs of the machine talk. Utterances 14 and 20 in Figure 1 are also examples of exaggerated articulation. Such attempts for adaptation are to be expected in the middle of the interaction or not far from eventual breakdown, which is a result also of the futility of repeated efforts.

Sighs are a typical expression of exhaustion as seen in the TOOT dialogues in line 25 below and utterance 12 in Figure 1.

second query
24. A: <what city do you want to go to?>
<the same intonation as first time>
25. C: <boston>
<sigh>

It can be contrasted with more positive regular breathing in the first query, example 7.

Example 7: first query
1. A: alright this is the amtrak train schedule system, which city do you want to go to,
2. C: <boston>
<positive, breathing>
…

4.4. Attention Calls

Attracting the attention of the Virtual agents in MRE is another way of dealing with delays and misunderstandings used by the human trainees. Thus vocative use of name or title may be a signal for the user's sense of lost contact. In the MRE data there are examples of how the human user utilizes utterance initial calling by name as attention and reaction elicitor:

Example 8: MREMAY1:3
User: where is the nearest hospital.
Virtual agents: (silence)
User: sergeant.
Virtual agents: (silence)
User: sergeant, where is the nearest hospital.

Also:

Example 9: MREMAY1:3
User: treat the victim.
Virtual agents: (silence)
User: tucci, treat the victim.

The same strategy can be used of course by the system when it detects decrease in the users attention.

5. Discussion and conclusion

Contemporary technology urges us to believe that it not only provides but also facilitates and improves 'communication'. As a result of such a belief there is increased negativity towards failure in 'communication' [7], which is cured only by more 'communication'. In this context, it is not surprising that one of the most aching problems in modern times is what does it mean to communicate [8]. The linguistic and philosophical view on this matter is divided in two camps:

> one, which defines communication as exchange of information and sees no other issues but eliminating the reasons for miscommunication and increasing communication for the benefit of the social communion [9]

> second, the phenomenological view which describes communication in ethical and pre-knowledge terms and sees breakdowns of communication as inherent properties of the activity and thus as opportunities for communication rather than problems of communication.

In the first tradition, the success of communication is described as part of the definition of what communication is. Thus if not successful the communication is no longer communication. Lack of or breakdown in communication is defined as no transmission of information, no signal in the wire, as misunderstanding, as a call for information therapy, and even as a disease (autism) [7]. Incommunicability is seen as mental and social abnormality. Other limits of communication i.e. points of expected breakdown are the four MAAD boundaries: Machines, Aliens, Animals, Dead. Mead's assumption of Reciprocity of Perspectives (taking the position/attitude of the other), which today comes in the form of the Theory of Mind, implies that communication transpires only among those who have a priori something in common. But when such communion does not succeed, what remains of the sublime ideal is a bitter disappointment of a promise that failed to arrive [8]. "But if communication bears the mark of failure or inauthenticity in this way, it is because it is sought in fusion" writes Levinas in his essay "The Other in Proust" [10]. Levinas meant fusion of humans, of views, of perspectives between humans. In this tradition the breakdown is part of communication and it is even the essence of communication because it is in the breakdown that the otherness transpires and thus calls for ethics. Human-machine technology aims at masking the obvious otherness of the machines by tracing and simulating human communication features. The users, even when aware of the machine, approach the interaction with expectations typical for human-human interaction. At a certain point they realize that the assumptions they carry are not satisfied and they have two choices: to adapt to the 'machine styles' or to get their hats and leave. Thus on one hand, we don't need to work for fusion between humans and machines by frenetically trying to eliminate any possible misunderstanding first, because misunderstanding is part of communication, no matter who the interlocutors are, second, because misunderstanding teaches the participants a sense of otherness and thus enhances attention, opens the interlocutors to surprise which is one of the "highest reaches of apperception in conception, judgment, and thought" [11], and third, because the desire for fusion leaves the participants unsatisfied, simply because the fusion is not possible nor meaningful, no matter who the interlocutors are. On the other hand, since the dialogue technology is still error-prone even on a speech recognition level, there is still space for improvement. Thus one way of improving human-machine interaction is not to eradicate all possible errors but to prepare the system for non-fluency and for state-of-mind updates. The other path of improvement is to study the actual causes and signals of breakdowns and prepare the systems to recognize them and in that way become more sensitive to the state-of-mind-and-emotion of the users. That is, both paths of improvement suggested here concern increase of grounding efforts by the means of better and more opportunities for monitoring of the other.

6. Acknowledgements

7. Reference

[1] Reeves, B. & Nass, C., *The Media Equation: How people treat computers, television and new media like real people and places.* Cambridge University Press, Cambridge, Mass., 1996.

[2] Heidegger, M., *Being and Time*, Harper Row, New York, 1962.

[3] Diane J. Litman, Shimei Pan, and Marilyn A. Walker. Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent., in *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL'98)*, pp. 780-786, Montreal, Canada, August 1998.

[4] Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., & Swartout, W. Toward a New Generation of Virtual Humans for Interactive Experiences, in *Proceedings of IEEE Intelligent Systems*, July/August, 32-38, 2002.

[5] Austin, J. L., *How To Do Things With Words.* Clarendon: Oxford, 1975.

[6] Allwood, J., "Obligations And Options In Dialogue", *Think, Vol. 3, 1994.*

[7] Pinchevski, A., *Interruption and Alterity*: *Dislocating the Ethical Possibilities of Communication,* PhD Thesis, McGill University, Montreal, 2003.

[8] Peters, J.D., *Speaking into the Air*, The University of Chicago Press, Chicago and London, 1999.

[9] Shanon, C., *The Mathematical Theory of Communication*, University of Illinois, Urbana, 1949.

[10] Levinas, E., "The Other in Proust", in *The Levinas Reader*, ed. Sean Hand. Basil Blackwell, Oxford, 1989.

[11] Charlesworth , W. R., "The Role of Surprise in Cognitive Development", in *Studies in Cognitive Development*, Oxford University Press, New York, 1969.