Talk is silver, silence is golden: A cross cultural study on the usage of pauses in speech

Birgit Endrass
Matthias Rehm
Elisabeth André
University of Augsburg
Eichleitnerstr. 30
D-86135 Augsburg Germany
{endrass|rehm|andre}@informatik.uni-augsburg.de

Yukiko I. Nakano
Tokyo University of Agriculture and Technology
2-24-16 Nakacho, Koganei-shi,
Tokyo 184-8588, Japan
nakano@cc.tuat.ac.jp

ABSTRACT

In this paper we examine the usage of pauses in speech. Thereby we concentrate on cultural differences with the aim to build a computational model for virtual agents later. By adapting the agents' conversation management behavior to cultural background, we hope to get a better acceptance in a given culture. Therefore we have a closer look at the occurrence of pauses in speech with their features like length or emplacement. To ground our model in empirical data, we analyzed the occurrences of pauses in speech in the CUBE-G video corpus, recorded in the two participating cultures Germany and Japan. In a preliminary study we observed the number of pauses that occurred in videos of approximately five minutes duration. First we took into account pauses that lasted for more than 1 second and later only those out of them that lasted for over 2 seconds. By comparing the two cultures, we found out that Japanese subjects used significantly more pauses for both lengths than German subjects.

Author Keywords

Embodied conversational agent, Pauses in speech, cross-cultural communication

ACM Classification Keywords

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces— interaction styles, Natural language, Theory and methods;

INTRODUCTION

Knapp and Vangelisti [11] examine personal relationships and their impact on interpersonal communication. For describing the possibility of deepening a friendship between males by using silence, they cite Roger Rosenblatt, who wrote an article for the Time Magazine called "The Silent Friendship of Men":

(…) Older Story: Wordsworth goes to visit Coleridge at his cottage, walks in, sits down and does not utter a word for three hours. Neither does Coleridge. Wordsworth then arises and, as he leaves, thanks his friend for a perfect evening. (…)

Would the same "conversation" have taken place if Mrs. Wordsworth and Mrs. Coleridge would have met? Or, if Wordsworth and Coleridge never met before? There are differences in the usage of silence in speech. But where do they come from? Some are evoked by gender or age, others by personal relationships. The utilization of pauses also varies across cultures.

We want to use tendencies about the frequency of pauses in speech, described in literature and confirmed by our corpus study, to adapt the dialogue model for Embodied conversational agents (ECAs) to a specific cultural model. ECAs can be regarded as a special case of multimodal dynamic interaction systems. They support the idea that humans prefer to interact with an artefact that possesses some human-like qualities. In the media equation [15] the authors state that people respond to computers as if they were humans. Thus people might also build up social relationships with virtual agents. To enhance the believability of those agents they could be extended with cultural background. Following Hofstede [8] human behavior is dependent from human nature, culture and personality. Although cultural background plays an important role in human interaction and virtual agents communicate with the user in a natural way, so far little effort has been made to integrate cultural differences into technical systems.

We believe that by realising culture specific dialogue management styles for ECAs, their believability could be enhanced. As the usage of pauses in speech is one important aspect in dialogue management we want to have a closer look at their occurrences to answer the following questions. How often and when do pauses take place? How long do they last? Who breaks the silence? What kind of speech acts are followed by pauses and which utterances are used for start ups? As a starting point we concentrate on the number of pauses in a conversation, namely pauses that last for more than 1 second and 2 seconds respectively.

This paper is organized as follows: First we describe some related work where ECAs already use silence in speech explicitly, although they do not use cultural differences. In the next section we give an overview of the usage of pauses in speech and their cultural differences. We then explain feasible enhancement for virtual agents, which should also serve as a basis for further research. Then a preliminary study is described, where the frequency of pauses in conversation is analysed in the two cultures Japanese and German. In the end of this paper, we discuss our results and give a foresight to our future work.

RELATED WORK

Although ECAs communicate in a more and more human manner, so far little effort has been made to integrate cultural context, as for example the different usage of silence in speech. Pauses in speech do occur in dialog simulations, but they often arise due to a lack of celerity in the speech components and thus appear to be distracting for the user. Nevertheless, pauses are used successfully to handle turn taking in some systems. So far a cultural aspect in the usage of silence has not been taken into account.

Sidner and colleagues [17], developed a model of engagement for a conversational robot, based on an analysis of human-human conversation. Engagement “is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake”. The appropriate use and correct interpretation of engagement signals are necessary prerequisites for the success of an interaction. In particular, pauses are used to recognize inattentiveness of the user, which encourages the robot to show engagement behavior.

Pauses in speech are often used for grounding behavior for ECAs. Cassel and colleagues [4] present a Real Estate Agent (REA) that acts in the function of a virtual realtor. In Smalltalk situations she gains information about the users preferences in buying a house. In [5] Cassel states that short pauses in speech lead to feedback behavior. Thus, the REA agent nods her head or emits a paraverbal (such as “Mmhmm”) or a short statement (such as “Okay”) as reaction to short pauses in the user’s speech.

Nakano and colleagues [14] developed a grounding model for the kiosk agent Mack that provides route descriptions for a paper map. The agent uses verbal and nonverbal grounding acts to update the state of the dialogue. They state, that pauses influence the choice of following actions.

Traum and Heeman [19] also consider grounding behavior in dialogues. They examine the co-occurrence between turn-initial grounding acts and utterance unit signals, e.g. prosodic boundary tones and pauses. Silence was divided into two groups: short silence (less than half a second) and long silence (longer than half a second). Then correlations with boundary tones and relatedness markings were analysed. They found out, that long pauses are positively related with the previous utterance being grounded and that they seem to be an indicator of utterance unit completion.

Nakanishi and colleagues [13] describe a helper agent that plays the role of a party host in a virtual meeting space where different cultures meet. In this system silence is used to detect conversations that are going badly. When the helper agent locates a pause in speech, it directs a series of yes/no questions to both conversation partners in order to find a topic that is interesting for both. Although the agent is developed to help in intercultural encounters, the length of silence that initiates the agent is not adapted to culture. After analysing their results, the authors state that an adoption to the user´s cultural background would make the agent more efficient.

PAUSES IN SPEECH

According to Clark [3] pauses are powerful cues for what is happening in a conversation. To use them as a basis for analyzing culture specific behavior, we first have to check carefully what purposes pauses may serve in conversations and how the usage differs across cultures. As we want to build a computational model for Germany and Japan, those two cultures are of special interest.

In [6] Goodwin describes his research on gaze behavior and manipulation. According to him gaze is used to manage turn taking and to signal understanding or attentiveness. If attention signals of the listener are missing, pauses are used by the speaker to regain attention. In this case the duration of the silence is dependent from the nonverbal signals of the hearer.

Pauses in speech can be used for the following purposes:

- cognitive processing
- control mechanism
- acceptance / refusal
- turn taking

Rochester [16] gives a brief history of studies dealing with filled and silence pauses. During a filled pause, sounds like *uhmm* and *ahhm* might occur as well as nonverbal behaviors like head nods or gestures. In comparison a silent pause is, as the name predicts, silent. Rochester summarizes the history of researches dealing with pauses in speech according to three models of the speaker.

In the first model pauses are supposed to reflect the strength or weakness of verbal habits; the second model enhances the first and constitutes pauses as signalling cognitive decisions about both immediate and later speech. Here pauses are assumed to stand in a temporally proximal relationship to the choices to be made. According to that, two particular functions are supported: (a) pauses signal some word choices, and (b) may reflect decisions at major constituent boundaries. A third function is the semantic decision-making. The matter of content and the function of pauses for the speaker are examined here.

Until that point, the speaker is simply a language generator which pauses either in the course of normal decision-making operations or because of disruptions in those operations. However the speaker can be seen as a participant in the social act of speech. According to Rochester, "*pauses and other phenomena of spontaneous speech should be functionally related to changes in the interpersonal situation and/or to changes in the responsiveness of the speaker, given a constant interpersonal situation*". In his work he examines the theoretical implications of pause location. In addition, the functional significance of pauses is considered in terms of cognitive, affective-state, and social interaction variables. He found out that two sorts of social interaction variables influence pausing in spontaneous speech:

- Mediating variables: e.g. changes in the audience situation and predispositional responsiveness to listeners, and
- Control variables: e.g. the number of potential speakers and the individual desire to speak.

In his work pauses in speech can either be used as control mechanism to control the flow of the conversation, as well as for cognitive processes, as decision making.

Another usage of pauses is described in [2], where politeness strategies are constituted as an aspect of social interaction. The authors describe some parallelisms in the linguistic construction of utterances with which people express themselves in different languages and cultures. One motive of these parallels is isolated – politeness. They claim the existence of conversational structure sequences and with it the intentional usage of pauses for politeness purposes. Note that a carefully located pause can on the one hand mean acceptance and on the other hand refusal. In their example (where A is a man, and W his friend's new bride) the silence conveys acceptance:

A: Do you sing?

W: (silence)

A: Hooray! Give us a song!

Whereas silence can also be a polite refusal like in a situation, where A writes to B for a favour and B does not reply.

Thus, pauses can be used to express refusal or acceptance in a polite way. But the interpretation of the pause remains a challenge to the interlocutor

Another common use of pauses in conversations is to initiate turn taking behavior. Louis ten Bosch [1] states that turn-taking is one of the basic mechanisms in all types of dialogues and that it is also a crucial mechanism in human-system interaction. They analysed the turn-taking mechanism in 93 telephone dialogues recorded in the Netherlands. Temporal phenomena of turn taking, such as the duration of pauses and overlaps of turns in dialogues were investigated. Pauses were divided into pauses between turns and pauses between utterances within turns and the average pause duration per dialogue was calculated. Their analysis shows that speakers adapt their turn-taking behaviour according to the average pause duration in the given conversation.

These results illustrate that people belonging to the same culture adapt their pause behavior in turn taking to each other. But the usage of silence in speech is also a well known difference between cultures [18]. This might lead to problems and misunderstandings in intercultural encounters.

CULTURE SPECIFIC DIFFERENCES

Hall, cited in [18] describes high- and low context cultures. In High context (HC) communication little explicitly is encoded and the conversation relies mainly on physical context. We find HCs in long lasting friendships, where conversations are difficult to understand for outsiders. Besides verbal utterances, meaning is transported through context (e.g social roles or positions), situation, nonverbal clues (e.g. pauses, silence, tone) and cultural information. In contrast low context communication (LC) explicitly code information. Therefore clear descriptions, unambiguous communication and a high degree of specificity are required.

The degree of context used in communication is dependent on culture. Germany is explicitly named in [7] as one of the probably lowest context cultures. However Japan like most Asian cultures belongs to the high context cultures, where communication partners are expected to be able to encode the implicit intent of the verbal message. Hall (1983), cited in [7] claims that silence serves as a critical communication device in Japanese communication patterns. Pauses reflect the thoughts of the speaker and can contain strong contextual meaning. In European conversations pauses are often sensed as unpleasant. Thus we expect people belonging to the Japanese culture to use pauses more frequently than Germans.

As culture is a rather abstract concept, there are several attempts building a concrete model. Hofstede [8] explains culture as a dimensional concept. His theory is based on a broad empirical survey in which over 20 different cultures were categorized into a five dimensional model. Each dimension contains two extreme sides, for which he clearly defines stereotypical behavior norms. He defines a given culture as a point in a five-dimensional space, according to the dimensions.

One of these dimensions is the so called identity dimension with the two extreme sides individualism and collectivism. It defines the degree to which individuals are integrated into a group. On the individualist side ties between individuals are loose, and people are expected to take care for themselves. On the collectivist side, people are integrated into strong, cohesive in-groups, often extended families which continue protecting them in exchange for unquestioning loyalty.

According to Hofstede Germany lies on the individualistic side of this dimension, whereas Japan is a collectivistic culture. In [9] he states, that in collectivistic cultures silence may occur in conversations without creating tension. Thus we expect to find more pauses in the Japanese conversations than in German ones, as the later should try to avoid embarrassing situations like silence, whereas the Japanese should not feel uncomfortable.

Pauses are used as means of conversation in Japan. But does this not hold for every culture? In [12], Morsbach warns not to read too much into the Japanese way of using silence and not to mystify it. He refers to the so called "Rare-Zero Differential", which means that something is rare in one culture, but completely nonexistent in another and thus taken as typical for the former. He refers to phenomena like kimonos or geishas, which tourists visiting Japan tend to see more often than nationals. But still he states that in specific situations there are differences in the usage of silence, e.g. mother-child relationships or female behavior and hiding of feelings. Also he reveals that the Japanese are often regarded as "silent", whereas westerners tend to be revered by the Japanese as "verbose". He agrees, that the average Japanese will use more pauses in speech than the average American, but additionally he states that there will be overlaps.

EMPIRICAL DATA

According to the literature overview given above, we hypothesize that in Japanese conversations pauses in speech will occur more frequently than in German conversations. To ground our expectations about culture specific dialogue management in empirical data, we additionally analysed the video corpus of the Cube-G project (CUlture-adaptive BEhavior Generation for interactions with embodied conversational agents). Therefore around 20 hours of video material were collected in the two participating cultures Japanese and German, with the aim to analyse nonverbal behavior.

It is organized as follows. Subjects were told that they take part in a study by a well-known consulting company for the automobile industry. To attract their interest in the study, a monetary reward was granted depending on the outcome.

One of the recorded scenes was a first time meeting, which is a variation of the standard first chapter of every language textbook. This includes a short introduction and small talk. We told our subjects that they should know each other slightly to be able to solve a task together later. This scenario takes about five minutes for every subject. The same design was used in Germany as well as in Japan. 21 subjects (10 female, 11 male) joined the study in Germany and 26 (13 female, 13 male) in Japan. To ensure that they all meet the same conditions we hired actors as communication partners. To control for gender effects, we had a male and a female actor, interacting with the same number of male and female subjects.

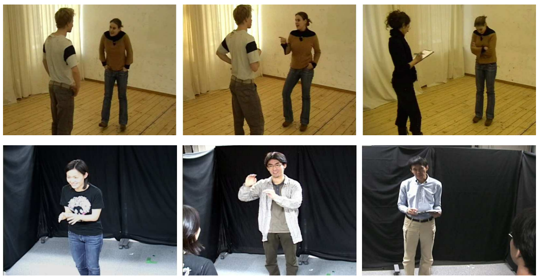

Figure 1. Figure 1 shows details from the corpus collection in both cultures, which serves as empirical data for the analysis of conversation management described in this paper.

ANNOTATION

In order to analyse the corpus an annotation using the Anvil Tool [10] was done. First, the video sequences had to be transliterated and translated into English language to allow analysis in both cultures. Figure 2 shows an example with a German subject.

With the annotation of speech, we were able to calculate Gaps between speech sequences. Time spans in which neither the subject nor the actor spoke were automatically computed and saved as pauses. Thus silent pauses and filled pauses that were filled by nonverbal clues were observed. To sort out short silences (like those while breathing or hesitating) we only observed pauses that last for more than one second. In a later analysis we restricted to pauses over 2 seconds. Please note that pauses over 2 seconds are also included in those that last for more than 1 second. As this paper only describes a preliminary study, we did not yet take into account the emplacement of pauses, but claim that further analysis of culture specific usage of pauses in speech seems to be a promising research field.

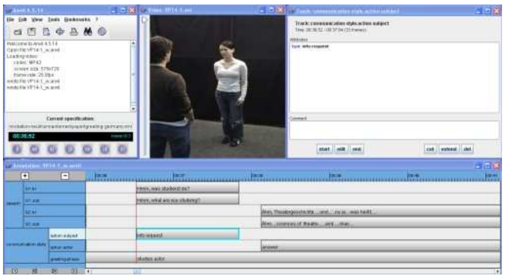

Figure 2. Example annotation with a German subject.

As a starting point to realize culture specific communication management behaviors for virtual agents, we first need to have a closer look at inner cultural communications, to answer the following question: Are the observed communication management behaviors typical for the given culture or do they simply show up because of personality, age or gender? Thus we started with an indepth analysis for the German samples, to compare the impact of gender combinations. As described above the corpus was recorded in all gender constellations. The situation in which the conversation takes place also influences the communication as well as the given interlocutor. Therefore we restricted our analysis to one typical scene out of our video corpus. As it was recorded with students, participants are all in the same age group.

Later we compare the German samples with the Japanese video recordings.

ANALYSIS

As a preliminary study we analysed eight German videos with four male and four female subjects. To fix as many conditions as possible, we chose to examine only videos from the first time meeting scenario. All gender combinations were observed, in order to analyse differences in the occurrence of pauses in mixed and same gender combinations respectively.

Table 1 shows an overview of the pauses in speech in the German videos. We found 7,1 pauses on average that lasted for more than one second, and only 1.3 pauses on average that lasted for more than 2 seconds in the 8 videos that were all approximately 5 minutes long.

A comparison of female and male subjects showed no significant difference in the usage of pauses (t-test), for both pauses, those over 1 second (p=0,748) and 2 seconds (p=0,750). The same holds for pauses in videos with mixed gender combinations compared to those where both subjects had the same gender (p_{1sec}=0,795; p_{2sec}=0,578). An interesting point for further research is that pauses over 2 seconds, which occurred after an utterance spoken by the male conversation partner was never broken by a female. All other combinations of breaking silence took place.

These results have to be taken with care, as we only analysed eight video samples for the German culture.

Subject/ Pauses	m	m	m	s-f	m	s-m	m	m
> 1 sec	14	1	7	4	4	12	12	3
> 2 sec	2	0	2	1	0	2	2	0

Table 1. Overview of the pauses in speech in the German video samples (where m=mixed gender; s=same gender; f=female; m=male)

Subject/ Pauses	s-f	m	s-f	s-f	s-m	m	s-m	m
> 1 sec	40	20	27	34	26	36	35	30
> 2 sec	12	4	6	7	10	10	10	8

Table 2. Overview of the pauses in speech in the Japanese video samples (where m=mixed gender; s=same gender; f=female; m=male)

As for the German video samples, we analysed eight Japanese videos with 4 female and 4 male subjects, where all gender combinations took place. Like the videos analysed above, the Japanese samples are from the first time meeting scenario and lasted about five minutes.

Table 2 shows an overview of the pauses used in the Japanese video recordings. We found 31 pauses on average that lasted over 1 second and 8,4 pauses on average per video, that lasted for more than 2 seconds. As for the German videos, we found no significant difference in the usage of pauses between the genders (t-test with (p_{1sec}=0,770; p_{2sec}=0,252). Again, different gender combinations showed no significant results, compared with same gender constellations (p_{1sec}=0,473; p_{2sec}=0,425).

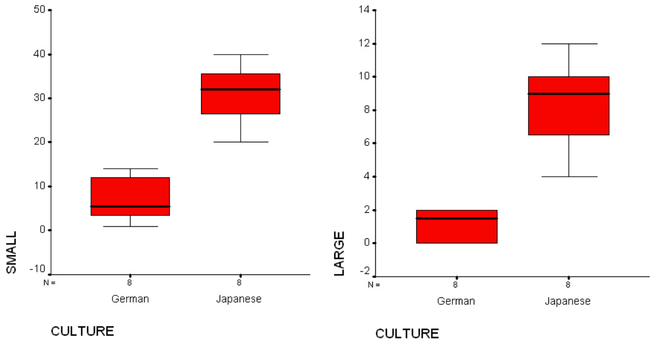

Figure 3. The usage of short (left) and long pauses (right) in speech in the two cultures Germany and Japan.

Interestingly, in the Japanese videos, too, no situation was found where the female conversation partner broke a silence that was longer than two seconds, when the male communication partner spoke the last utterance. All other combinations took place.

Comparing the flow of conversation between the two cultures, the results are promising. As provided in literature, the Japanese video samples comprise apparently more pauses. We found significant differences between the two cultures (t-test), for both pauses over 1 second (p<0,001) and 2 seconds (p<0,001) respectively. Figure 3 shows the Box plots for short (left) and long pauses (right), where the difference in the usage of pauses between the two cultures is shown graphically.

CONCLUSION AND FUTURE WORK

In this paper we gave a brief overview of the usage of pauses in speech and focused on differences caused by cultural background. By comparing the two cultures Germany and Japan in a preliminary study, we found promising results. Like predicted in literature, Japanese subjects showed significantly higher numbers of pauses between speech utterances than German subjects. Thus we emphasize this as a promising research field with the aim for integrating cultural differences in embodied conversational agents. Although the results are promising, we do not want to declare prototypes, but think we found interesting tendencies for further exploration. As future work, we need to analyse all videos recorded for the CUBE-G corpus, in order to strengthen our results.

Additionally we want to have a closer look at the positions where pauses take place, to answer the following question: Who breaks the silence? What kind of speech acts are followed by pauses and which utterances are used for start ups? Therefore we need to categorise the speech utterances, which also allows an analysis of sequences of speech utterances that evoke pauses.

ACKNOWLEDGMENTS

The work described in this article is funded by the German Research Foundation (DFG) under research grant RE 2619/2-1 and the Japan Society for the Promotion of Science (JSPS) under a Grant-in-Aid for Scientific Research (C) (19500104). The authors would like to thank Prof. Toyoaki Nishida and Hung-Hsuan Huang for their support collecting the Japanese corpus and Afia Akhter Lipi, Yuji Yamaoka and Franziska Grüneberg for annotating the video samples.

REFERENCES

1. Bosch, ten L., Oostdijk, N., Ruiter, de J. P., Turn-taking in social talk dialogues: temporal, formal and functional aspects. In *Proceedings SPECOM* 2004.
2. Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. New York: Cambridge University Press.
3. Clark, H. H., *Using Language*. Cambridge, England: Cambridge University Press. 1996.
4. Cassell, J., Embodied conversational interface agents. In *Communications of the ACM*. Vol. 43, No. 4, April 2000.
5. Cassell, J., Nakano, Y., Bickmore, T., Sidner, C. L., and Rich, C. Non-verbal cues for discourse structure. In *Proc. Of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2001.
6. Charles Goodwin. *Conversational Organisation - Interaction between Speakers and Hearers*. New York: Academic Press, 1981.
7. Hecht, M. L., Andersen, P. A., Ribeau, S. A., The Cultural Dimensions of Nonverbal Communication. In *Asante, M. K., Gudykunst, W. B., Handbook of International and Intercultural Communication.* (p. 163-185). London: Sage Publications. (1989).
8. Hofstede, G. Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Thousand Oaks, London: Sage Publications. (2001).
9. Hofstede, G. J., Pedersen, P. B., & Hofstede, G. *Exploring Culture: Exercises, Stories, and Synthetic Cultures.* Yarmouth: Intercultural Press. 2002.
10. Kipp, M. *Gesture Generation by Imitation – From Human Behavior to Computer Character Animation*. Universität des Saarlandes, PhD. Thesis. 2003
11. Knapp, M. L., Vangelisti, A. L., *Interpersonal Communication and Human Relationships.* – 5th ed. Pearson Education. 2005
12. Morsbach, H., The Importance of Silence and Stillness in Japanese Nonverbal Communication: A Cross-Cultural Approach. In *Fernando Poyatos(Edt.) Cross-Cultural Perspectives in Nonverbal Communication.* C.J. Hogrefe, 1988.
13. Nakanishi, H., Ishida, T., Isbister, K., and Nass, C., Designing a Social Agent for Virtual Meeting Space. In *S. Payr & R. Trappl (Eds.), Agent Culture: Human-Agent Interaction in a Multicultural World* (p. 245-266). London: Lawrence Erlbaum Associates. (2004).
14. Nakano, Y. I., Reinstein, G., Stocky, T., and Cassell, J., Towards a model of face-to-face grounding. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 553–561, 2003.
15. Reeves, B., & Nass, C. The Media Equation — How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge: Cambridge University Press. (1996).
16. Rochester, S. R., The Significance of Pauses in Spontaneous Speech. In *Journal of Psycholinguistic Research, Vol. 2, No. 1,* 1973.
17. Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N.. Where to look: a study of human-robot engagement. In *IUI ’04: Proceedings of the 9th international conference on Intelligent user interface*, pages 78–84, New York, NY, USA, 2004. ACM Press.
18. Ting-Toomey, S. *Communicating across cultures*. New York: The Guilford Press. 1999.
19. Traum, D. and P. Heeman. *Utterance Units and Grounding in Spoken Dialogue*. in *ICSLP*. 1996