Dialog Behaviors across Culture and Group Size

David Herrera[1], David Novick[1], Dusan Jan[2], David Traum[2]

[1] Department of Computer Science, The University of Texas at El Paso,
500 West University Avenue, El Paso, TX 79968-0518 USA
[2] USC Institute for Creative Technologies,
12015 Waterfront Drive, Playa Vista, CA 90094-2536
herrera78@gmail.com, novick@utep.edu, jan@ict.usc.edu, traum@ict.usc.edu

Abstract. This study analyzes joint interaction behaviors of two-person and four-person standing conversations from three different cultures, American, Arab, and Mexican. To determine whether people use joint interaction behaviors differently in multiparty versus dyadic conversation, and how differences in culture affect this relationship, we examine differences in proxemics, speaker and listener gaze behaviors, and overlap and pause at turn transitions. Our analysis suggests that proxemics, gaze, and mutual gaze to coordinate turns change with group size and with culture. However, these changes do not always agree with predictions from the research literature. These unanticipated outcomes demonstrate the importance of collecting and analyzing joint interaction behaviors.

Keywords: Dialog, proxemics, gaze, turn-taking, multicultural, dyadic, multiparty

1 Introduction

When people converse with others, they participate in joint interaction behaviors, such as proxemics (interpersonal distance and orientation), mutual gaze, and turn-taking, which they may not consciously negotiate. How these behaviors manifest depends on many factors, such as gender, age, personality, culture, and number of participating conversants. Understanding these differences is important for situations where intercultural joint interaction behaviors are necessary for mission success, such as for military personnel in foreign countries. Currently, the United States has soldiers in war zones where they find themselves interacting with people of other cultures. Being able to decode or interpret these local behaviors correctly helps keep soldiers unharmed. Virtual reality training systems have been developed with this in mind (e.g., [1, 2]). In these applications, human trainees interact with embodied conversational agents (ECAs), intelligent virtual characters that possess conversational capabilities. ECAs need models of joint interaction behaviors to behave according to culture and group dynamic. Models based on dyadic or Anglo-American studies may not be appropriate for ECAs representing different cultures and interacting with multiple trainees or ECAs. At the same time, the research literature of

interaction behaviors has largely focused on dyadic conversations. The joint interaction behaviors for dyadic conversation may differ from those for multiparty conversation.

We have previously reported on the collection of a multimodal multicultural corpus of dialogs comprising four-person conversations on a range of five different activities [3]. A parallel corpus of dyadic conversations by people from the same culture groups on the same tasks was also collected [4]. In the present paper, we examine these corpora for differences in proxemics, speaker and listener gaze behaviors, overlap and pause at turn-transitions, and mutual gaze to coordinate turn. Our central question is how people use joint interaction behaviors differently in multiparty versus dyadic conversation and how this relationship is affected by differences in culture.

2 Review of the Literature

Our study of joint interaction behaviors begins with a brief survey of the sociolinguistic and anthropological literature of differences in conversation behaviors across cultures. We rely especially on the notion of high-context and low-context societies. We then briefly review the interaction behaviors on which our analysis centers.

2.1 Cultural Dimensions in Conversation

Non-verbal behaviors in cultures can be modeled through a structure of six cultural dimensions [5], based on previous work by Hofstede [6] and Hall [7, 8]. Table 1 summarizes these dimensions.

The first dimension is attributed to Hall, who contrasted two conversational styles, high-and low-context. In high-context societies, many things are left unsaid, allowing non-verbal behaviors to play a bigger role. This is typical of cultures that share similar experiences and expectations. Arab and Mexican cultures are considered higher-context cultures [9, 10]. In low-context societies, communication needs to be relatively more explicit, and the value of a single word is not as strong. The American culture is considered a lower-context culture ([9, 10].

The next four dimensions are attributed to Hofstede, who used them to describe cultural variability of people in organizations. Hofstede's individualism-collectivism dimension tends to track Hall's distinction between high- and low-context cultures. Cultures with a high individualism index prioritize individual goals, prefer autonomy and self-assertion while, at the other end, low index cultures emphasize group goals, harmony and avoiding confrontation. Hofstede also defined power distance, uncertainty avoidance, and masculinity dimensions. Power distance can be seen in terms of hierarchism versus egalitarianism and through factors of hierarchism are gender, age or family background.

The last dimension, high/low-contact [11], describes accessibility-inaccessibility in relationships. This dimension deals with immediacy, such as closeness or distance and

behaviors expressing approach or avoidance. Examples of highly immediate behaviors include smiling, eye contact, open postures, closer distances and more vocal animation. Cultures with these behaviors are considered as high-contact cultures, because of their preference for close distances and touch [7]; Arabs and Mexican are members of high-contact cultures. On the other end of the spectrum are low-contact cultures, such as Americans, who prefer more distance and less touch [11].

While studies have not verified Hall's space patterns for other cultures, some studies have found significant differences in proxemics between cultures. One study of Anglo-, Black-, and Mexican-Americans in natural settings found that Mexican-American adults stood significantly closer than their Anglo-American counterparts as listed in Table 2.5 [12].

Table 1. Cultural variation along dimensions

	High/Low - Context	Individualism - Collectivism	Power Distance	Uncertainty Avoidance	Masculinity - Femininity	High/Low - Contact
Arab cultures	High	38	80	68	53	High
Mexico	High	30	81	82	69	High
USA	Low	91	40	46	62	Low

2.2 Interaction Behaviors

The six dimensions of cultural variation may help explain how conversation interaction behaviors, such as turn-taking, gaze and proxemic behaviors, are used in different cultures. The scores rank cultures along those dimensions. Of course, they describe cultural tendencies rather than what individuals of those cultures will necessarily do. While conversational interaction behaviors have been the subject of extensive study, here we briefly review the literature to illuminate our specific hypotheses.

Proxemics refers to the spatial distance between persons interacting with each other and their orientation toward each other. One could argue that proxemics is a relationship rather than a non-verbal behavior, although it may communicate things like a person's intention or emotion. A more elaborate definition of proxemics encompasses eight behaviors, including touch, amount of eye contact, voice loudness, and body-contact distance [13]. Like other joint interaction behaviors, the proxemics between interacting persons can be interpreted differently across cultures. In some societies close distances are reserved for personal relationships and may not be comfortable for interacting otherwise; in other cultures, close distances are not so exclusive and not interacting closely is interpreted as aloofness [14, 7, 8]. While proxemics are culturally defined, there are also variations based on gender, social status, environmental constraints and type of interaction. For a review of this literature, see [15].

With respect to turn-taking, speakers of English signal transition relevance points through use of cues, such as intonation-marked phonemic clauses, sociocentric sequences such as "you know", completion of grammatical clauses, paralinguistic

drawl, termination of hand gesticulation or decrease of paralinguistic pitch or loudness of sociocentric sequences [16].

Gaze plays an important role in coordinating turn-taking. A speaker can yield the floor or signal the next speaker by his or her gaze behaviors. Kendon [17] attributed at least four functions to gaze behaviors in a conversation: 1) to provide visual feedback, 2) to regulate the flow of conversation, 3) to communicate emotion and relationships, and 4) to improve concentration by restriction of visual input. He also showed that speakers tend to look away at the beginning of an utterance and look at the listener at the end of an utterance. In a later study, gaze played a role in coordinating turn-taking, where 42% and 29% of turn exchanges involved a mutual-break and a mutual-hold pattern, respectively [18]. Mutual-break is a term that describes a pattern where both conversants momentarily gaze at each other at a turn exchange followed by the turn-taker breaking gaze. Mutual-hold is a similar pattern, except that the turn-taker does not break gaze immediately, but later on in the turn.

2.3 Hypotheses

Based on our review of the literature research with respect to cultural differences and dialog interaction behaviors, we proposed a set of hypotheses that related changes in turn-taking, gaze and proxemics as a function of culture and group size. Table 2 summarizes the hypotheses.

Table 2. Hypotheses

Joint Interaction Behavior(s)		Changes observed as group size increases (dyadic to multiparty)		
		American (non-contact)	Mexican (contact)	Arab (contact)
Turn-Taking	Overlap	Decrease	Increase	Increase
	Pause	Increase	Decrease	Decrease
Gaze	By Speaker	Increase	No change	No change
	By Non-speaker	No change	No change	No change
Turn-taking x Gaze: Mutual gaze at turns		No change	No change	Increase
Proxemics		Decrease	Decrease	Decrease

3 Methodology

To address our central question of how people use the joint interaction behaviors of proxemics, turn-taking and gaze differently in multiparty versus dyadic conversation, and how this relationship is affected by differences in culture, we analyzed the conversations collected in the UTEP-ICT Cross-Cultural Multiparty Multimodal Dialog Corpus [3, 4], extended to include dyadic as well as multiparty conversations (Herrera 2010).

The extended corpus comprises approximately 20 hours of audiovisual multiparty interactions in three different cultures and languages. Groups of two or four native

speakers of Arabic, American English and Mexican Spanish completed five tasks and were recorded from six angles. The subjects were recruited from local churches, restaurants, on campus, and through networks of known members of each cultural group in the El Paso area, which borders Mexico and has, in part because of the university, many representatives of other nations and cultures. Tasks 1, 4, and 5 were mainly narrative tasks, where the participants can take turns relating stories or reacting to the narratives of others. Tasks 2 and 3 were constructive tasks, in which the participants must pool their knowledge and work together to reach a group consensus. Tasks 3 and 4 were designed to have a toy provide a possible gaze focus other than the subjects themselves, so that gaze patterns with a copresent referent could be contrasted with gaze patterns without this referent. Task 5 was meant to elicit subjective experiences of intercultural interaction. The interactions were recorded with six Apple iMac computers, placed around the periphery of a large open room that serves as a computer lab. We thus recorded six simultaneous views of the subjects as they conversed, making it possible, with rare exceptions, to code the subjects' proxemics, gaze and turn-state.

From the recordings, we produced time-aligned partial codings of each of the 24 conversations. Specifically, we coded two 30-second excerpts of each of the conversations for tasks 1 through 4 for proxemics, turn-taking, and gaze. For proxemics, we measured the distance (in inches) between subjects; we avoided inflated numbers (due to distances in quads of conversants across from each other or standing shoulder-to-shoulder and not interacting) by calculating the minimum spanning forest of the positions of the conversants. For gaze, measurements were calculated respective to the conversant's role as speaker or listener; that is, for an annotation of look-away, the talk state of the subject would be considered such that, if the conversant was talking, it was taken as speaker look-away, and if listening, then listener look-away. For turn-taking, we calculated the average of pause (in seconds) and overlap (in negative seconds) at turn-transitions.

The coding was performed by three students trained in UTEP's Interactive Systems Group. The coders followed written rubrics for each of the behavior types, and entered the data using ANVIL[19]. Coded data were assessed for interrater reliability. For the three behaviors, Kappa was at least 0.80. (For proxemics, positions were considered equivalently coded if they differed by less than 6 inches.) If outliers were found in this cross-check, the videos were revisited and recoded, if needed.

From these data, we calculated summary statistics and assessed each of the hypotheses. For each dependent variable, we conducted a 3 x 2 x (4) mixed factorial ANOVA, controlling for relevant covariates, including gender, age, familiarity and acculturation. Follow-up t-tests were computed to assess differences between conditions that demonstrated significant main effects or interactions. Additionally, within-subject analysis was conducted for the repeated task measure, and its interactions. Finally, the interaction between joint interaction behaviors was examined to find any interesting correlations.

4 Results

For turn-taking, the analysis confirmed our hypothesis that Americans quads pause more at turns than American dyads. The other hypotheses with respect to turn-taking were not confirmed, probably because the effect size was small relative to the sample size. For gaze, most of our results surprised us: contrary to our hypotheses, Americans and Mexicans (speakers and non-speakers) gazed at each other more in quads than in dyads, while Arab (non-speakers) gazed less in quads than in dyads. Again contrary to our hypotheses, mutual gaze at turns declined from dyads to quads across all three cultures. For proxemics, the analysis confirmed our hypotheses that conversants in all three cultures would stand closer to each other in quads than in dyads. Table 3 presents the complete set of results.

Table 3. Results

Joint Interaction Behavior(s)		Changes observed as group size increases (dyadic to multiparty)		
		American (non-contact)	Mexican (contact)	Arab (contact)
Turn-Taking	Overlap	Not confirmed	Not confirmed	Not confirmed
	Pause	Confirmed: Significantly more	Not confirmed	Not confirmed
Gaze	By Speaker	Disconfirmed: Significantly more	Disconfirmed: Significantly more	Not confirmed
	By Non-speaker	Disconfirmed: Significantly more	Confirmed: Significantly more	Disconfirmed: Significantly less
Turn-Taking x Gaze: Mutual Gaze at Turns		Disconfirmed: Significantly less	Confirmed: Significantly less	Confirmed: Significantly less
Proxemics		Confirmed: Significantly less	Confirmed: Significantly less	Confirmed: Significantly less

To assess the possible interactions between joint interaction behaviors, we looked at correlations among speaker and listener gaze, proxemics, turn-transition overlap and pause, and mutual gaze to coordinate turn-transition. Our results suggest that that speaker gaze and listener gaze are significantly correlated ($r = .815$, $p < 0.01$) suggesting that conversants reciprocate gaze behaviors. Proxemics correlates negatively with speaker ($r = -.268$, $p < 0.05$) and listener gaze ($r = -.309$, $p < 0.01$), an unexpected result contradicting the Equilibrium Model [20]; this result may result from the increased gaze and the reduced proxemics in quads.

5 Conclusion

This work was motivated in large part by the need for more realistic models of joint interaction behaviors for digital simulations conversations in, for example, immersive cross-cultural training environments (see, e.g., [21]). A key problem faced by the builders of such systems is how to set the parameters for joint interaction behaviors so that these behaviors would provide realistic training for people who would be expected to interact with people in cultures other than their own. While our results cannot completely determine these parameters, the results do move forward with respect to the way in which the parameters should be set.

In terms of the overall question, it seems that having more conversants has a slightly bigger impact on joint interaction behaviors than do cultural differences for gaze, turn-taking and proxemics. However, culture helps make more accurate predictions. For example, for proxemics, although all quads stood closer, some cultures did not do so as much as others.

Unfortunately, not all of the statistical tests were conclusive, which may be attributable to the small sample size. Even so, and beyond the main hypotheses, the data led to additional insights about the relationship of culture and group size to interaction behaviors.

For turn, Americans were thought to use high-considerateness style, keeping overlap to a minimum and allowing sufficient pause. But in our data, differences between quads and dyads showed marginally significant differences, suggesting American quad conversants had a high-involvement style. Arabs' mean measures for pause/overlap increased marginally in quads, suggesting they use high-considerateness style with more conversants. Mexican mean pause/overlap behaved as Americans, decreasing with more conversants, although the difference was not as large.

With respect to gaze, it appears that the overwhelming factor to consider for quad gaze is the number of conversants, as an increase in the number of conversants provides more persons to look at, thus increasing gaze. Mexicans did not seem to follow Americans and Arabs in this trend, though. Mexican gaze seemed to remain steady across group size.

For mutual gaze to coordinate turn exchange, an odd result was that for Mexicans mutual gaze significantly coordinated a smaller percentage of turn-transitions. These results may arise because their gaze did not increase in quads. It may be that Mexicans, rather than relying on mutual gaze to coordinate turn, simply used the timing of pause/overlap. Perhaps turn-taking was not competitive, and their high tolerance for overlap permitted such an arrangement.

For proxemics, differences did occur for Americans, but not in the direction we had predicted: our data indicated that dyads maintained more distance than quads. In the dyadic case, the conversants seem to prefer a distance comparable to the distances of conversants diagonally across from each other in quad conversation. For Arabs and Mexicans, the results were confirmed, although their differences were not as pronounced as American differences. Quads stood slightly closer than in dyads, but this may be the product of the minimum spanning forest measurement.

5.1 Summary

Our principal result is that joint conversation control behaviors in digital simulations of conversation should reflect the number of conversants. The results suggest that as conversations go from dyads to quads:

- Turn-taking: For Americans and Mexicans, the amount of pause/overlap should decrease; for Arabs, the amount of pause/overlap should increase.
- Amount of gaze: For Americans, the amount of time that speakers and listeners gaze at each other should increase; for Arabs, the amount of time that listeners gaze at the speaker should decrease.
- Mutual gaze at turn transitions: For Americans and Arabs, the amount of mutual gaze at turn transitions should increase; for Mexicans, the amount of mutual gaze at turn transitions should decrease.
- Proxemics: For all groups, the mean distance among conversants should decrease. A reasonable guide would be that the longest distances among conversants in quads should be similar to the direct distance between conversants in dyads.

A second result is that it is probable that joint conversation behaviors do reflect differences between high-contact and low-contact cultures.

- The amount of time gazing at the other participant in dyads should be lower for Americans than for Arabs and Mexicans.
- Interpersonal distances in dyads (significant) and quads (suggestive) should be greater for Americans than for Arabs and Mexicans.

5.2 Limitations and future work

The first limitation of our study that ought to be addressed is the small sample size. Although it is a huge undertaking to annotate video excerpts, unfortunately several of the hypotheses produced inconclusive results.

Another issue is how best to select the excerpts to analyze. To the extent possible, 30-second excerpts were selected at the same time into task for all groups, but perhaps it would have been better to select excerpts based on conversational situation, such as many turn exchanges, or specific interactions such as adjacency-pairs, grounding and repair, or using speech acts as a factor. Joint interaction behaviors differed significantly across tasks, and sometimes differed across excerpts. This is not surprising, as some tasks encouraged more turn-taking, and some required closer proxemics, such as the toy-naming task. Similarly, in some tasks, mutual gaze coordinated turns more than in others. To better understand the process that may govern these joint interaction behaviors, it would be useful to consider the context. This would provide more insight into these mechanisms and ease the efforts to annotate the videos.

Our study led to questions both of the cognition involved in interaction behaviors and in the methodologies for understanding these behaviors.

Substantively, it appears that in some tasks mutual gaze played a role in a larger percentage of the turn exchanges. How did conversants negotiate the next turn in other tasks? Did they mainly rely on detecting transition-relevant places? What

behaviors can be used to model ECA behaviors to improve turn-taking in group situations?

Likewise, mutual gaze to coordinate turn-transition was different for each culture. For Americans and Arabs, this significantly increased, as did the gaze for speaker and listener, but for Mexicans, it did not. This could be used to modify the turn-taking model in [22], where gaze plays a bigger role in American multiparty conversation than in dyadic, a big role in Arabs, though not much more than in dyads, and a smaller role in Mexican multiparty conversation.

Methodologically, our experience in this study suggests that it would be worthwhile to address the correlation measures for computational models. While speaker and listener gaze are correlated, these correlations are significantly different across cultures. Arabs seem to fall into one category, with high amounts of gaze, while Americans and Mexicans seem to fall into another. Lower values for proxemics do not seem to decrease gaze levels and increase mutual gaze at turn-transitions as well as reduce turn-transition times.

Timing poses another methodological issue. While 0.5 seconds is a good pause/overlap measure for American dyads, quads in all cultures dropped pause/overlap to half that amount. Models that run on half-second intervals may not be adequate for multiparty interaction. The model in [23] uses center of structure to calculate proxemics, while this study analyzed proxemics of the quad using a minimum spanning forest measure. Measures using center of structure may be a more fair measure and should be considered. Nevertheless, significant results were achieved across culture and group size, suggesting that looking at the proxemics differences of quads across culture in more detail could be fruitful.

These improved correlations hold the promise of improving the model of joint interaction behaviors across cultures and, correspondingly, improving both our understanding of the way people coordinate their conversations and our ability to reflect this understanding in digital environments.

References

1. Johnson, W. L., Beal, C., Fowles-Winkler, A., Narayanan, S., Papachristou, D., Marsella, S., et al.: Tactical Language Training System: An interim report. In: Proceedings of the International Conference on Intelligent Tutoring Sys., pp. 336-245 (2004)
2. Deaton, J., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., & McCollum, C.: Virtual environment cultural training for operational readiness (VECTOR). Virtual Reality 8 (3), pp. 156—167 (2005)
3. Herrera D., Novick D., Jan D., & Traum D. The UTEP-ICT Cross-Cultural Multiparty Multimodal Dialog Corpus. In: Multimodal Corpora Workshop: Advances in Capturing, Coding and Analyzing Multimodality (MMC 2010), Valletta, Malta (2010)
4. Herrera, D.A.: Gaze, turn-taking and proxemics in multiparty versus dyadic conversation across cultures, doctoral dissertation, The Unviersity of Texas at El Paso (2010)
5. Hecht, M. L., Andersen, P. A., & Ribeau, S. A.: The Cultural Dimensions of Nonverbal Communication. In: M. K. Asanti, & W. B. Gudykunst (Eds.), *Handbook of International and Intercultural Communication*, pp. 163—185. Sage Publications, Newbury Park, CA (1989)

6. Hofstede, G.: Culture and Organizations: Software of the Mind. McGraw-Hill, New York (1997).

7. Hall, E. T.: The Hidden Dimension. Doubleday, Garden City, NY (1966)

8. Hall, E. T.: Beyond Culture. Doubleday, Garden City, NY (1976)

9. Gudykunst, W. B., & Ting-Toomey, S.: Culture and Interpersonal Communication. Sage Publications, Newbury Park, CA (1988)

10. Hecht, M. L., Andersen, P. A., & Ribeau, S. A.: The Cultural Dimensions of Nonverbal Communication. In: M. K. Asanti, & W. B. Gudykunst (Eds.), Handbook of International and Intercultural Communication, pp. 163-185. Sage Publications, Newbury Park, CA (1989)

11. Altman, I., & Gauvain, M.: A cross-cultural dialective analysis of homes. In: L. Liben, A. Patterson, & N. Newcombe (Eds.), Spatial representation and behavior across the life span. Academic Press, New York (1981)

12. Baxter, J. C.: Interpersonal Spacing in Natural Settings. Sociometry 33 (4), pp. 444—456 (1970)

13. Hall, E. T. A System for the Notation of Proxemic Behavior. American Anthropologist 65, pp. 1003-1026 (1963).

14. Hall, E. T.: The Silent Language. Doubleday, Garden City, NY (1959)

15. Altman, I., & Vinsel, A.: Personal Space: An Analysis of E.T. Hall's Proxemics Framework. In I. Altman, & J. Wohlwill (Eds.), Human Behaviour and Environment: Advances in theory and research, Vol. 2, pp. 181—260). Plenum Press, New York (1977).

16. Duncan, S., & Fiske, D. W.: Face-to-Face Interaction: Research, Methods, and Theory. Lawrence Erlbaum Associates, Hillsdale, NJ (1977)

17. Kendon, A.: Some Functions of Gaze Direction in Social Interaction. Acta Psychologica 32, pp. 1—25 (1967)

18. Novick, D., Hansen, B., & Ward, K.: Coordinating turn-taking with gaze. In: Proc. of ICSLP- 96, pp. 1888—1891. Philadelphia (1996)

19. Kipp, M.: Spatiotemporal coding in ANVIL. In: Proceedings of the 6th international conference on Language Resources and Evaluation (LREC-08) (2008).

20. Argyle, M., & Dean, J. Eye-Contact, Distance and Affiliation. Sociometry , 28, pp. 289—304 (1965)

21. Jan, D., Herrera, D., Martinovski, B., Novick, D., & Traum, D. A Computational Model of Culture-Specific Conversational Behavior. Proceedings of the 7th international Conference on Intelligent Virtual Agents. Paris, France (2007).

22. Jan, D., & Traum, D.: Dialog Simulation for Background Characters. In: Proceedings of 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece, pp. 65—74 (2005).

23. Jan, D., & Traum, D.: Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversations. In: Proceedings of ACL 2007 Workshop on Embodied Language Processing, Honolulu, pp. 59—66 (2007).