Avoiding and Resolving Initiative Conflicts in Dialogue*

Fan Yang and **Peter A. Heeman**
Center for Spoken Language Understanding
OGI School of Science & Engineering
Oregon Health & Science University
{fly,heeman}@cslu.ogi.edu

Abstract

In this paper, we report on an empirical study on initiative conflicts in human-human conversation. We examined these conflicts in two corpora of task-oriented dialogues. The results show that conversants try to avoid initiative conflicts, but when these conflicts occur, they are efficiently resolved by linguistic devices, such as volume.

1 Introduction

Current computer dialogue systems tend to be system-initiative. Although there are some mixed-initiative systems that allow the user to make a request or state a goal, such systems are limited in how they follow natural initiative behavior. An example is where the system always releases the turn whenever the user barges in. However, in a complex domain where the computer system and human user are collaborating on a task, the computer system might need to interrupt the human user, or might even need to fight with the human user over the turn. Thus the next generation of computer dialogue systems need a better model of initiative (Horvitz, 1999). In what situations can the system try to take initiative from the user? What devices can the system use to fight for initiative? We propose examining human-human conversation to answer these questions. Once we understand the conventions people adopt in negotiating initiative, we can implement them in a computer dialogue system to create natural interactivity.

In this research work, we examined two corpora of human-human conversation: the Trains corpus (Heeman and Allen, 1995) and the MTD corpus (Heeman et al., 2005). The research purpose is to understand conversants' behavior with initiative conflicts, which we define a situation where both conversants try to direct the conversation at the same time, but one of them fails. We found that (1) conversants try to avoid initiative conflicts; and (2) initiative conflicts, when they occur, are efficiently resolved by linguistic devices, such as volume.

In Section 2, we review related research work on modeling initiative and turn-taking. Dialogue initiative and turn-taking are two intertwined research topics. When conversants fight to show initiative, they are also fighting for the turn to speak. In Section 3, we describe the two corpora and their annotations. In Section 4, we define initiative conflict and give an example. In Section 5, we present the evidence that conversants try to avoid initiative conflicts. In Section 6, we present evidence that initiative conflicts are efficiently resolved by linguistic devices. We discuss our findings in Section 7 and future work in Section 8.

2 Related Research

2.1 Initiative Models

Researchers have been investigating how people manage dialogue initiative in their conversation. Whittaker and Stenton (1988) proposed rules for tracking initiative based on utterance types; for example, statements, proposals, and questions show initiative, while answers and acknowledgements do not. Smith (1993) proposed four different initiative strategies with differing amounts of control by the system. Chu-Carrol and Brown (1998) distinguished dialogue initiative from task initiative, and proposed an evidential model of tracking both of them. Cohen et al. (1998) proposed presenting initiative in different strengths. Some researchers related initiative to discourse structure. Walker and Whittaker (1990) found a correlation between initiative switches and discourse segments. Strayer et al. (2003) proposed the *restricted initiative model* in which the initiator of a discourse segment, who introduces the discourse segment purpose, is in control of the segment and shows most of the initiative. These models allowed the possibility that multiple conversants will want to show initiative at the same time; however, none of them addressed initiative conflicts.

Guinn (1998) studied another type of initiative, task initiative, which is about directing the problem-solving

*This work was funded by the National Science Foundation under IIS-0326496.

Proceedings of NAACL HLT 2007, pages 17–24,
Rochester, NY, April 2007. ©2007 Association for Computational Linguistics

of a domain goal. Guinn proposed that the person who is more capable of coordinating the current goal is the person who should be leading the dialogue. Initiative switches between conversants as goals get pushed and popped from the problem-solving stack. However, because conversants only have incomplete information, initiative conflicts might occur when conversants overestimate their own capability or underestimate the other's. Guinn proposed a negotiation model to resolve these conflicts of task initiative. Conversants negotiate by informing each other of positive and negative information of their plans to achieve the goal. By comparing each other's plan, the conversant whose plan has the higher probability of success takes initiative. Guinn's research on conflicts of task initiative, however, has little bearing on conflicts of dialogue initiative. For dialogue initiative, very often, one of the conversants just gives up the attempt very quickly, without giving a justification. As stated by Haller and Fossum (1999):"... conflicts are often simple clashes that result from both participants trying to take the initiative at the same time. Such conflicts do not necessarily require complex negotiation to resolve. Often, unwritten rules based on factors like social roles, personal assertiveness, and the current locus of control play a part in determining who will give away." However, Haller and Fossum did not further investigate how conversants efficiently resolve conflicts of dialogue initiative.

2.2 Turn-Taking and Initiative

Turn-taking in conversation is highly related to initiative. Conversants have to possess the turn in order to show initiative. When conversants are fighting for initiative, they are also fighting for the turn to speak. Thus the mechanisms of turn-taking might share some similarity with initiative. On the other hand, turn-taking is different from initiative; for example, an answer takes a turn, but answering does not show initiative.

Turn-taking in conversation has been discussed in linguistics literature. Duncan (1974) examined cues (gesture, acoustic, and linguistic) that conversants use to signal turn-taking or turn-releasing. A model based on these signals was created to account for conversants' turn-taking behavior. In this model, miscues are the cause of overlapping speech: for example, the hearer misrecognizes the speaker's cue to keep the turn, or the speaker fails to properly signal.

Sacks et al. (1974) proposed a set of rules for turn-taking: the current speaker can select somebody else to speak; otherwise, hearers can self-select to speak; otherwise, the speaker can self-select to speak. This model suggested that overlapping speech results from either the hearer waiting too long to speak, or the speaker not waiting long enough.

Schegloff (2000) examined overlapping speech in detail in human conversation. He concluded that (1) fights for turn are often accompanied with sudden acoustic alteration, such as louder volume, higher pitch, and faster or slower speaking rate; (2) the vast majority of fights for turn are resolved very quickly; (3) fights for turn are resolved through an interactive procedure, e.g. syllable by syllable negotiation, using devices such as volume, pitch, and speaking rate. However, his analysis only consisted of a few examples; no statistical evidence was given. It is thus unclear whether his conclusions represent human conventions of initiative conflict, or are occasional behavior that would only occur under special circumstances.

3 Corpora and Annotations

To understand human behavior in initiative conflicts, we examined two corpora, the Trains corpus and the MTD corpus. These two corpora have very different domain setups. The distinct behavior seen in each corpus will help inform us how domain settings affect initiative, while the common behavior will help inform us the cross-domain human conventions.

3.1 The Trains Corpus

The Trains corpus is a collection of human-human task-oriented dialogues, in which two participants work together to formulate a plan involving the manufacture and transportation of goods. One participant, the user, has a goal to solve; and the other participant, the system, knows the detailed domain information including how long it takes to ship and manufacture goods.

We annotated eight Trains dialogues totaling about 45 minutes using the tool DialogueView (Yang et al., 2007). We tagged each utterance with a simplified DAMSL scheme (Core and Allen, 1997). Utterances were tagged as forward or backward functions, stalls, or non-contributions. Forward functions include statements, questions, checks and suggestions. Backward functions include agreements, answers, acknowledgments, repetitions and completions. Examples of stalls are "um" and "let's see", used by a conversant to signal uncertainty of what to say next or how to say it. Non-contributions include abandoned and ignored utterances. The flow of the dialog would not change if non-contributions were removed.

Hierarchical discourse structure was annotated following Strayer et al. (2003). To determine whether a group of utterances form a discourse segment, we took into account whether there exists a shared goal introduced by one of the conversants (cf. Grosz and Sidner, 1986).

3.2 The MTD Corpus

The MTD corpus contains dialogues in which a pair of participants play two games via conversation: an ongoing game that takes a relatively long time to finish and an interruption game that can be done in a couple turns but has a time constraint. Both games are done on computers. Players are separated so that they cannot see each other.

In the ongoing game, the two players work together to assemble a poker hand of a full house, flush, straight, or four of a kind. Each player has three cards in hand, which the other cannot see. Players take turns drawing an extra card and then discarding one until they find a poker hand, for which they earn 50 points. To discourage players from simply rifling through the cards to look for a specific card without talking, one point is deducted for each picked-up card, and ten points for a missed or incorrect poker hand. To complete this game, players converse to share card information, and explore and establish strategies based on the combined cards in their hands.

From time to time, the computer generates a prompt for one player to start an interruption game to find out whether the other player has a certain picture on the screen. The interruption game has a time constraint of 10, 25, or 40 seconds, which is (pseudo) randomly determined. Players get five points for the interruption game if the correct answer is given in time. Players are told to earn as many points as possible.

We annotated six MTD dialogues totaling about 90 minutes. Utterances were segmented based on player's intention so that each utterance has only one *dialogue act* that is to share information, explore strategies, suggest strategies, or maintain an established strategy (Toh et al., 2006). We applied the same simplified DAMSL scheme on utterance tag annotations. Figure 1 shows an annotated excerpt of an MTD dialogue. We grouped utterances into blocks. Block *b21* is a game block in which conversants completed a poker hand. Blocks *b22* and *b23* are two card blocks in which conversants picked up a new card, discussed what they had in hand, and chose a card to discard. Block *b24* is an interruption segment in which conversants switched their conversation to the interruption game. No claim is made that the game and card blocks are discourse segments according to Grosz and Sidner's definition (1986).

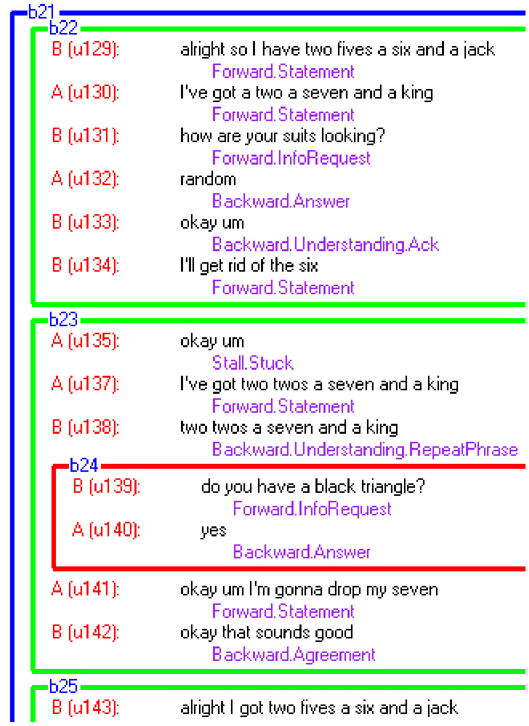

Figure 1: An excerpt of an MTD dialogue

4 Defining Initiative Conflicts

An initiative conflict occurs when a conversant's attempt to show initiative fails because someone else is showing initiative at the same time. Following Whittaker and Stenton (1988), we use utterance tags to determine whether an utterance shows initiative: forward functions show initiative while others do not. Non-contributions are viewed as failed attempt to show initiative. Thus we identify initiative conflicts as overlapping utterances that involve either a forward function and a non-contribution or two non-contributions.

Figure 2 gives an example of an initiative conflict from the MTD corpus. The top conversant says "that's pair of threes and pair of fours", which ends at time point A. After a short pause, at time B, the bottom conversant asks "how many threes do you have", which is overlapped by the top conversant's second utterance "I'll drop" at time C. The top conversant then abandons the attempt of showing initiative at time D. Hence the bottom speaker is the winner of this initiative conflict.

We use the term *preceding-pause* to refer to the time interval between the end of the previous utterance and the first utterance that is involved in the overlap (from A to B in Figure 2). *Offset* refers to the interval between the start times of the two overlapped utterances (from B to C). *Duration* refers to the time interval from the beginning of overlap till the end of overlap (from C to D).

In the Trains corpus, there are 142 cases of overlapping speech, 28 of which are initiative conflicts. Of the remaining, 96 cases involve a backward function (e.g. an acknowledgment overlapping the end of an inform), and 10 cases involve a stall. The remaining 8 cases are other types of overlap, such as a collaborative completion, or conversants talking about the same thing: for example, one saying "we are a bit early" and the other saying "we are a little better".

In the MTD corpus, there are 383 cases of overlapping speech, 103 of which are initiative conflicts. Of the remaining, 182 cases involve a backward function, 21 cases involve a stall, and 77 cases are others. Initiative conflicts

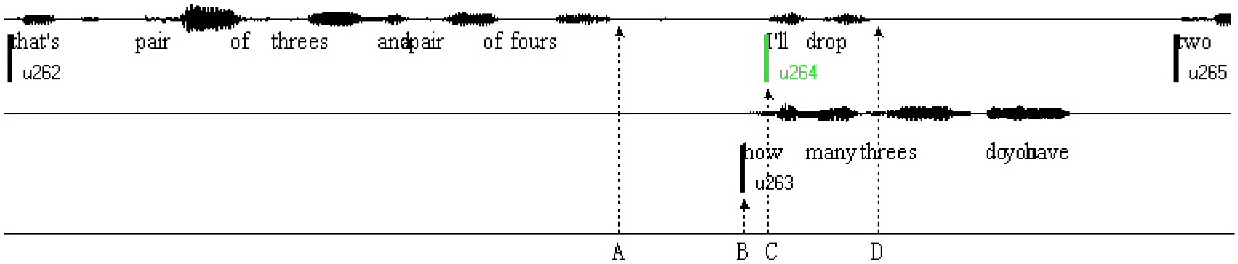

Figure 2: An illustration of an initiative conflict

are more frequent in the MTD corpus (103 cases in 90 min) than in the Trains corpus (28 cases in 45 min).

There are three cases in the Trains and thirteen cases in the MTD corpus where the preceding-pause is negative, i.e. the first overlapped utterance is started before the other conversant finishes the previous utterance. Sometimes the hearer starts a little bit early to take the turn. If the original speaker does not intend to release the turn, a conflict arises. Because these cases involve three utterances, we exclude them from our current analysis and save them for future research.[1] This leaves 25 cases in the Trains corpus and 90 cases in the MTD corpus for analyzing initiative conflicts.

5 Avoiding Initiative Conflicts

In this section, we show that conversants try to avoid initiative conflicts by examining both the offset of initiative conflicts and the urgency levels.

5.1 Offset of Initiative Conflicts

The offset of an initiative conflict indicates where the conflict happens. A short offset indicates that the conflict happens at the beginning of an utterance, while a long offset indicates an interruption in the middle.

Figure 3 shows the cumulative distribution function (CDF) for offsets for both corpora individually. The mean offset is 138ms for the Trains corpus, and 236ms for the MTD corpus. In comparison to the average length of forward utterances (2596ms in the Trains corpus and 1614ms in the MTD corpus), the offset is short. Moreover, in the Trains corpus, 88% of offsets are less than 300ms (and 80% less than 200ms); in the MTD corpus, 75% of offsets are less than 300ms. Thus most initiative conflicts happen at the beginning of utterances.

[1]These cases of negative value preceding-pause are in fact very interesting. They seem to contradict with Sacks et al. (1974)'s model that the hearer has priority to self select to speak. If Sacks et al. is correct, the speaker should wait a certain amount of time in order not to overlap with the hearer, but in these cases we see that the speaker self-selects to speak without taking into account whether the hearer self-selects to speak or not.

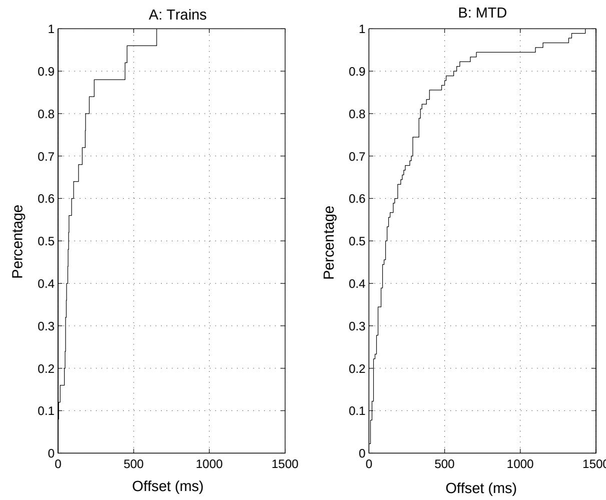

Figure 3: CDF plot for offsets of initiative conflicts

Few initiative conflicts have offsets longer than 500ms. There is one instance in the Trains corpus and eleven in the MTD corpus. Four cases are because the second conversant has something urgent to say. For example, when an interruption game is timing out, conversants would interrupt, sometimes in the middle of an utterance, which results in a long offset. Another six cases are due to miscues. Figure 4 shows an example. Conversant B said "I have two aces" with end-of-utterance intonation, paused for about half a second, and then added "and a seven". The ending intonation and the pause probably misled conversant A to believe that B had finished, and thus A started a new forward utterance, which overlapped with B's extension. A's utterance was then quickly abandoned. In these cases, it is ambiguous whether B's utterance "I have two aces ... and a seven" should be further chopped into two utterances. The final two cases are intrusions, with an example shown in Figure 5. Conversant A cut in probably because he was confident with his decision and wanted to move on to the next card. In such cases, the intruder might be perceived as being rude.

B:	I have two aces and a seven
A:	I have .

Figure 4: Long offset: miscue

B:	well let's just
A:	it's no help I think it goes away

Figure 5: Long offset: intrusion

The preponderance of short offsets provides evidence that conversants try to avoid initiative conflicts. When A detects that B is talking, A should not attempt to show initiative until the end of B's utterance in order to avoid conflicts, unless there is an urgent reason. If conversants did not take into account whether someone else is speaking before attempting initiative, we would see a lot of intrusions in the middle of utterances, which in fact rarely happen in the two corpora. As we have shown, initiative conflicts tend to happen at the beginning of utterances. Thus initiative conflicts occur mainly due to unintentional collision, i.e. both conversants happen to start speaking almost at the same time. The fact that the offset of most initiative conflicts is within 300ms confirms this.[2]

5.2 Urgency Level and Initiative Conflicts

To further support the hypothesis that conversants avoid initiative conflicts except for urgent reasons, we examined the MTD corpus for the correlation between the urgency levels of the interruption game and initiative conflicts. For the urgency level of 10 seconds, conversants started 33 interruption games, 8 of which were introduced via initiative conflicts. For 25 seconds, conversants started 36 interruption games, 5 introduced via initiative conflicts. For 40 seconds, conversants started 33 interruption games, 3 introduced via initiative conflicts. Thus the percentages of initiative conflicts for the three urgency levels are 24% for 10 seconds, 14% for 25 seconds, and 9% for 40 seconds. The urgency level of 10 seconds requires conversants to start the interruption game very quickly in order to complete it in time. On the other hand, the urgency level of 40 seconds allows conversants ample time to wait for the best time to start the game (Heeman et al., 2005). Thus we see the percentage of initiative conflicts decreases as it becomes less urgent to the interruption game. These results suggest that conversants try to avoid initiative conflicts if they can, unless there is an urgent reason.

6 Resolving Initiative Conflicts

In this section, we present evidence that initiative conflicts, if they occur, are resolved very quickly using simple devices.

[2]This 300ms might be related to human reaction time.

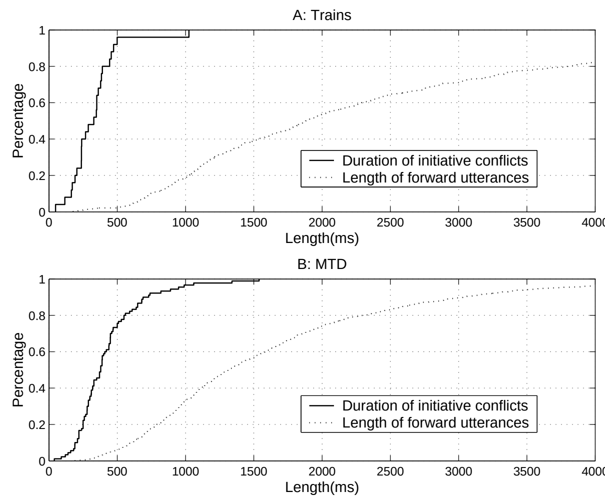

Figure 6: CDF plot for durations of initiative conflicts together with lengths of forward utterances

6.1 Duration of Initiative Conflicts

The duration of an initiative conflict, as defined in Section 4, indicates how quickly the conflict is resolved. Figure 6 shows the cumulative distribution function of durations of initiative conflicts and the lengths of forward utterances in the two corpora. The mean duration is 328ms in the Trains corpus and 427ms in the MTD corpus. From Figure 6 we see that the duration is much shorter than the length of forward utterances, which have the mean length of 2596ms in the Trains corpus and 1614ms in the MTD corpus. The difference between duration of initiative conflicts and length of forward utterances is statistically significant ($p < 10^{-5}$, ttest). On average, the duration of initiative conflicts is about 1/8 the length of forward utterances in the Trains corpus and about 1/4 in the MTD corpus. The short durations suggest that initiative conflicts are resolved very quickly.

According to Crystal and House (1990), the average length of CVC syllable is about 250ms. Thus on average, the length of initiative conflicts is about one to two syllables.[3] In fact, 96% of conflicts in the Trains corpus and 73% in the MTD corpus are resolved within 500ms. These observations are consistent with one of Schelogff's (2000) claims about turn-taking conflicts, that they usually last less than two syllables to resolve.

6.2 Resolution of Initiative Conflicts

From our definition of initiative conflict, at least one of the speakers has to back off. For expository ease, we re-

[3]It would be interesting to examine the length of initiative conflicts based on syllable. However currently we do not have syllable-level alignment for the two corpora. We leave this for future research.

fer to the person who gets the turn to contribute as *the winner*, and the other who fails as *the yielder*. There are two cases in the Trains corpus and three cases in the MTD corpus in which both speakers abandoned their incomplete utterances, paused for a while, and then one of them resumed talking. These five cases are treated as ties: no winners or yielders, and are excluded from our analysis here.

Given how quickly initiative conflicts are resolved, we examined whether the resolution process might be dependent on factors presented before the conflict even begins, namely who was speaker in the previous utterance, and who was interrupted. If we predict that the conversant who spoke prior to the conflict (speaker of u262 in Figure 2) loses, we get 55% accuracy in the Trains corpus and 61% accuracy in the MTD corpus. If we predict the conversant who spoke first in the overlap (speaker of u263 in Figure 2) wins, we get 60% accuracy in the Trains corpus and 53% accuracy in the MTD corpus. These low percentages suggest that they are not robust predictors.

We next examined how conversants resolve the conflicts using devices such as volume, pitch, and others.

6.2.1 Volume

For a stretch of speech, volume is calculated as the mean energy of the spoken words. For each initiative conflict, we calculated each conversant's volume during the overlap, and then normalized it with respect to the conversant's volume throughout the whole conversation.[4] We refer to this as *relative volume*. In the Trains corpus, the average relative volume of the winner is 1.06; the average relative volume of the yielder is 0.93. The difference is statistically significant ($P < 0.01$, anova). In the MTD corpus, the average relative volume of the winner is 1.12; the average relative volume of the yielder is 0.98. The difference is also statistically significant ($p < 10^{-6}$, anova). These results show that the winner is the one speaking at a higher relative volume.

To strengthen our argument, we also calculated *volume ratio* as the relative volume of the winner divided by the yielder. The average volume ratio in the Trains corpus is 1.16 and in the MTD corpus is 1.18. If a classifier always chooses the speaker with higher relative volume to be the winner, we achieve about 79% accuracy in both corpora, which is a 29% absolute improvement over random prediction. These results further confirm that the conversant who speaks at a higher relative volume wins the initiative conflicts.

Given the importance of volume in the resolution process, we examined whether it has an impact on the duration of initiative conflicts. Figure 7 plots the relation

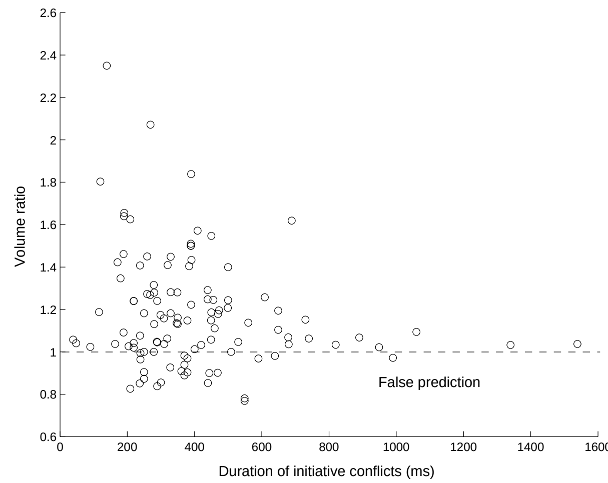

Figure 7: Volume ratio and duration of conflicts

between volume ratio and duration of conflicts for all the cases in the two corpora. For reference, the dotted line divides the data points into two groups: under the line are what volume ratio fails to predict the winner, and above the line are success. If we look at the points where volume ratio succeeds, we see that when duration of initiative conflicts is long, volume ratio tends to be small: in fact, the average volume ratio for initiative conflicts shorter than 600ms is 1.27; for long than 600ms is 1.13; and the difference is statistically significant (ttest, $p < 0.01$).

To further understand how volume is used in the resolution procedure, we examined how volume changes during the overlap. For initiative conflicts whose duration is longer than 600ms, we cut the overlapped speech evenly in half, and calculated the relative volume for each half individually. For the first half, the average relative volume of the winner is 1.03, and the yielder is 1.02. The difference is not statistically significant ($p = 0.93$, paired ttest). For the second half, the average relative volume of the winner is 1.20, and the yielder is 1.02. The difference is statistically significant ($p < 0.001$, paired ttest). The fact that these long initiative conflicts are not resolved in the first half is probably partially due to the close relative volume.

We then calculated *volume increment* as subtracting the relative volume of the first half from the second half. The average volume increment of the winner is 0.17; the average volume increment of the yielder is 0. The difference is statistically significant ($p < 0.001$, paired ttest). These results show that the range of volume increment during the overlap by the winner is larger than the yielder. The behavior of increasing volume during overlap to win the fight suggests that conversants use volume as a device to resolve initiative conflicts.

[4]Normalization is necessary particularly as conversants heard each other via headsets, and the microphones were not calibrated to have exactly the same gains.

6.2.2 Pitch

We used the tool *WaveSurfer* (Sjölander and Beskow, 2000) to extract the f0 from the audio files. We calculated *relative pitch* similarly as we did for volume.

In the Trains corpus, the average relative pitch of the winner is 1.02; the average relative pitch of the yielder is 0.96. The difference is not statistically significant ($P = 0.54$, anova). In the MTD corpus, the average relative pitch of the winner is 1.09; the average relative pitch of the yielder is 0.98. The difference is statistically significant ($p < 0.001$, anova). If we choose the speaker with higher pitch to be the winner, we achieve about 65% accuracy in the Trains corpus and 62% in the MTD corpus. These results suggest that pitch alone is not robust for predicting the winner of initiative conflicts, at least not as predictive as volume, although we do see the tendency of higher pitch by the winner.

We also examined pitch range in the window of 100ms and 300ms respectively. We calculated the pitch range of the overlapping speech and then normalized it with respect to the conversant's pitch range throughout the whole conversation. We did not see a significant correlation between pitch range and the winner of initiative conflicts. Thus pitch does not seem to be a device for resolving initiative conflicts.

6.2.3 Role of Conversants

Human-computer dialogues often have a user interacting with a system, in which the two have very different roles. Hence, we investigated whether the conversant's role has an effect in how initiative conflicts are resolved. We focused on the Trains corpus due to both its rich discourse structure and the difference in the roles that the system and the use have.

In the Trains corpus, if we predict that the initiator of a discourse segment wins the conflicts, we get 65% accuracy. In system-initiated segments, the system wins all eight conflicts; however, in user-initiated segments, the user only wins seven and system wins eight. The user does not have an advantage during initiative conflicts in its segments. Moreover, if the initiator had an advantage, we would expect the system to have fought more strongly in the user-initiated segments in order to win. However, we do not see that the relative volume of the system winning in user-initiated segments is statistically higher than in system-initiated segments in this small sample size ($p = 0.9$, ttest). The initiator does not seem to have a privileged role in the resolution process.

From the above analysis, we see that the system wins the conflicts 16 out of 23 times. Thus if we predict that the system always wins the conflicts, we achieve 70% accuracy. This is not surprising because the system has all the domain information, and is more experienced in solving goals. If the system and user want to speak at the same time, both would know that the system probably has a more significant contribution. That the system wins most of the initiative conflicts agrees with Guinn (1998) that capability plays an important role in determining who to show initiative next.

7 Discussion

In this paper, we present our empirical study of human behavior in initiative conflicts. Our first finding is that conversants try to avoid initiative conflicts. The consequence of initiative conflicts is that at least one of the conversants would have to back off, which makes their effort of contributing in vain. Moreover, the effort of resolving initiative conflicts is overhead to the dialogue. According to the theory of *least collaborative effort* by Clark and Wilkes-Gibbs (1986), it only makes sense for conversants to interrupt when the loss of not interrupting is higher than the cost of an initiative conflict. Thus the theory of least collaborative effort is consistent with our conclusion that most initiative conflicts are unintentional collisions, except where conversants interrupt in the middle of an utterance for urgency reasons.

The second finding of our research is that initiative conflicts, when they occur, are efficiently resolved. We found that volume plays an important role: the louder speaker wins. We also show how conversants change their volume to resolve initiative conflicts. Conversants probably identify their eagerness of speaking, confidence in what they want to say, and capability of achieving the current goal by means of volume, which resolves the initiative conflicts very quickly.

Domain settings obviously have an impact on conversants' initiative behavior. There are more frequent initiative conflicts in the MTD corpus than in the Trains corpus. Moreover, the roles of the conversants also affect their initiative behavior as we found that the system wins more initiative conflicts in the Trains corpus. In a teacher-student conversation, one would expect to see that the teacher interrupts the student more often than vice versa, but also that the teacher wins more initiative conflicts. Capability, culture, and social relationship probably are some underlying elements that influence when and under what conditions conversants would seek initiative, while volume is a device for resolving initiative conflicts.

8 Future Work

In this paper we focused on initiative conflicts in dialogue where two conversants cannot see each other. In face-to-face conversation, there might be other cues, such as eye-contact, head-nodding, and hand gesture, that conversants use in initiative conflicts. Moreover, in a multi-party conversation, a conversant might talk to different people on different topics, and get interrupted from time to time, which leads to an initiative conflict involving multiple speakers. In our future work, we plan to examine initiative conflicts in face-to-face multi-party conversation, such as the ICSI corpus (Shriberg et al., 2004).

Inspired by the findings on human behavior of initiative conflicts, we speculate that conversants might also have a mechanism to even minimize unintentional initiative conflicts, which probably includes devices such as volume, pause, and other prosodic features. The speaker uses these devices, as opposed to explicitly informing each other of their knowledge to evaluate capability (Guinn, 1998), to implicitly signal his or her eagerness, confidence and capability. The hearer then compares his or her own eagerness with the speaker's, and decides whether to just make an acknowledgement (allowing the speaker to continue the lead) or to take over the initiative when taking the turn to speak. In our future work, we plan to build an initiative model to capture this negotiation process.

References

Jennifer Chu-Carroll and Michael K. Brown. 1998. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User Adapted Interaction*, 8:215–253.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognitive Science*, 22:1–39.

Robin Cohen, C. Allaby, C. Cumbaa, M. Fitzgerald, K. Ho, B. Hui, C. Latulipe, F. Lu, N. Moussa, D. Pooley, A. Qian, and S. Siddiqi. 1998. What is initiative? *User Modeling and User Adapted Interaction*, 8:171–214.

Mark G. Core and James F. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Cambridge.

Thomas H. Crystal and Arthur S. House. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of Acoustical Society of America*, 88:101–112.

Starkey Duncan. 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 2:161–180.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Curry I. Guinn. 1998. An analysis of initiative selection in collaborative task-oriented discourse. *User Modeling and User Adapted Interaction*, 8:255–314.

Susan Haller and Timothy Fossum. 1999. Using protocols to model mixed initiative interaction. In *Proceedings of AAAI Workshop on Mixed Initiative Intelligence*.

Peter A. Heeman and James F. Allen. 1995. The Trains spoken dialogue corpus. CD-ROM, Linguistics Data Consortium.

Peter A. Heeman, Fan Yang, Andrew L. Kun, and Alexander Shyrokov. 2005. Conventions in human-human multithreaded dialogues: A preliminary study. In *Proceedings of Intelligent User Interface (short paper session)*, pages 293–295, San Diego CA.

Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of CHI*, pages 159–166, Pittsburgh PA.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Emanuel A. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29:1–63.

E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. 2004. The ICSI meeting recorder dialog act corpus. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*.

Kåre Sjölander and Jonas Beskow. 2000. WaveSurfer: An open source speech tool. In *Proceedings of ICSLP*, pages 4:464–467, Beijing China.

Ronnie W. Smith. 1993. Effective spoken natural language dialogue requires variable initiative behavior: an empirical study. In *AAAI93 Fall Symposium On Human-Computer Collaboration*.

Susan E. Strayer, Peter A. Heeman, and Fan Yang. 2003. Reconciling control and discourse structure. In J. Van Kuppevelt and R. W. Smith, editors, *Current and New Directions in Discourse and Dialogue*, chapter 14, pages 305–323. Kluwer Academic Publishers.

Siew Leng Toh, Fan Yang, and Peter A. Heeman. 2006. An annotation scheme for agreement analysis. In *Proceedings of INTERSPEECH*, Pittsburgh PA.

Marilyn Walker and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of 28th ACL*.

Steve Whittaker and Phil Stenton. 1988. Cues and control in expert-client dialogue. In *Proceedings of 28th ACL*, pages 123–130.

Fan Yang, Peter A. Heeman, Kristy Hollingshead, and Susan E. Strayer. 2007. Dialogueview: Annotating dialogues in multiple views with abstraction. *Natural Language Engineering*. To appear.