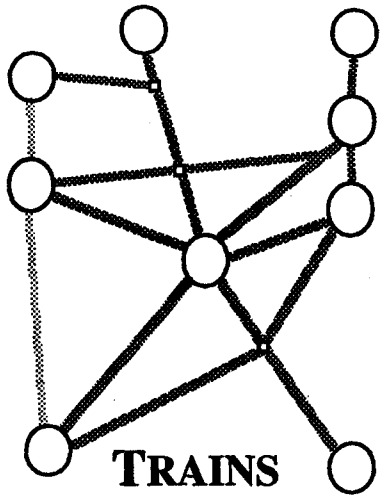

A Study on Prosody and Discourse Structure in Cooperative Dialogues

Shin'ya Nakajima and James F. Allen

TRAINS Technical Note 93-2
September 1993

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1993	technical report

4. TITLE AND SUBTITLE
A Study on Prosody and Discourse Structure in Cooperative Dialogues

5. FUNDING NUMBERS
N00014-92-J-1512

6. AUTHOR(S)
Shin'ya Nakajima and James F. Allen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Computer Science Dept.
734 Computer Studies Bldg.
University of Rochester
Rochester, New York 14627-0226

8. PERFORMING ORGANIZATION REPORT NUMBER
TN 93-2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Information Systems
Arlington, VA 22217

DARPA
3701 N Fairfax Drive
Arlington, VA 22203

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution of this document is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
(see title page)

14. SUBJECT TERMS
TRAINS; dialogue; prosody; discourse structure

15. NUMBER OF PAGES
18

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	UL

A Study on Prosody and Discourse Structure in Cooperative Dialogues

Shin'ya Nakajima
Speech and Acoustics Laboratory
NTT Human Interface Laboratories
Kanagawa 238-03 Japan

James F. Allen
Computer Science Department
University of Rochester
Rochester, New York, U.S.A. 14627-0226

September 1993

This paper describes how well prosodic information correlates with the topic structure of a cooperative dialogue. To investigate this correlation systematically, first we introduce the notion of utterance unit (UU) as a basic unit in conversations. We define the utterance unit by employing four principles. The grammatical principle is a syntactic criterion in which the UU boundary is set wherever a period can be placed. The pragmatic principle says that each UU corresponds to a basic speech act. In other words, if two neighboring phrases correspond to different speech acts (for instance, acknowledgment and request), they should be taken as two different UUs. The conversational principle addresses the turn-taking aspect of conversations. A UU boundary should be placed wherever the speaker changes. Finally, the prosodic principle says that whenever a medium length or longer pause (750 msec) is inserted between two phrases, they are to be taken as two different UUs. We apply these principles to a speech database containing about one and a half hours of collected dialogue to split the dialogues into a sequence of UUs. We then classify the inter-UU boundaries based on the relationship between two neighboring UUs into four semantic categories: topic shift, topic continuation, elaboration (or clarification), and speech-act continuation. The prosodic parameters measured at each boundary are the onset fundamental frequency (F0), the final F0, and the F0 maximal peak declination ratio (the ratio of the current UU's maximal peak to that of the preceding UU). Our study shows how these prosodic parameters vary depending on the topic structure. Our results can be summarized as follows. (1) The onset F0 value tends to be higher when the topic is changed at the UU boundary. (2) The final F0 value indicates finality and is much higher (on average) at speech-act continuation boundaries than at other boundaries. (3) The maximal peak declination ratio reflects the degree of subordination to the preceding UU. That is, this ratio is lowest at elaboration boundaries and highest at topic shift boundaries. Finally, we discuss discourse structure identification via the prosodic parameters.

This research was supported in part by DARPA/ONR under contract N00014-92-J-1512.

Contents

1 Introduction

The last decade has seen substantial progress in discourse processing and computational linguistic fields. Specifically, several plan recognition approaches based on Austin and Searle's speech-act theory [3, 20], in which the speech understanding process is viewed as the speaker's plan recognition problem, have been proposed (e.g. Allen and Perrault[1]). However, although a number of analysts have pointed out that prosody plays several important roles in natural conversations, no study has systematically analyzed prosodic characteristics in spontaneous conversations. Brown and Yule [5], for instance, discussed the correlation between topic shifting and the onset F0 with reference to a number of typical utterances, and Hirschberg and Pierrehumbert [12] investigated the intonational structure of discourse and proposed intonational assignment rules for speech synthesis. Neither of them, however, introduced statistical data from natural conversations.

Prosodic information plays various pragmatic roles in a conversation: the most obvious function of intonation is questioning. That is, by finishing a sentence with rising intonation, we can create a yes-no question. Prosody can also specify the information structure– such as new/old information, and the topic structure. This paper focuses on the latter function of prosody in spontaneous dialogues.

There have been a number of studies on this issue. Hakoda and Sato [11] claimed that when one read aloud written texts, the syntactic structure of each sentence is reflected in the prosodic parameters; onset, peak, and final F0 values, of each intonational phrase. Grosz and Hirschberg [8] analyzed AP news stories spoken by a newscaster and confirmed that there was a correlation between the discourse features such as the discourse segment boundaries and the prosodic features: the F0 range and pause insertion. Fujisaki [7] also investigated the narration of professional announcers and reported a correlation between prosodic phrasing and paragraph structure. All of these focused on professional speakers reading prepared texts. Thus, compared to natural conversations, the prosodic features of speech of this sort tend to be well formulated. In spontaneous conversations, complete *sentences* are seldom found and speech is frequently interrupted by the other speaker(s). Thus, the prosodic features in natural conversations may be much more unstable than found in narrations. The goal of this study was to investigate the correlation between prosody and the discourse structure in spontaneous conversations and to show how prosodic information can be used as a cue for the discourse structure.

In the next section, we discuss our specific task domain– TRAINS world – which was originally introduced in Allen and Schubert [2], and we describe how we collected natural conversations. We then define the topic structure markers which are based on the notion of *utterance unit.* Finally, we show how well particular prosodic parameters correlate with the topic structure and discuss discourse structure identification via the prosodic parameters.

2 Speech Data Collection

The map of the TRAINS world is shown in figure 1. A user or Human (hereafter called **H**) should achieve a specific goal by making plans to manufacture and ship various goods to specified cities by the due date. Another person called System (**S**) has up-to-date knowledge on the state of the world and assists **H** in making plans to achieve the given goal. A sample of the problems is;

> *You need to ship a tanker of OJ, a tanker of beer, and two boxcars of bananas to city H by tomorrow evening by 9 p.m., and a tanker of beer to city F by the same time.*

While making plans, **S** and **H** are sitting in different rooms and communicate by using microphones and head phones. The speech of **H** and **S** is recorded on the right and left channel, respectively, of a digital audio tape.

3 Discourse Structure Marking

3.1 Utterance Unit

Since grammatical units such as *sentences* are absent in spontaneous conversations, we must first determine what is the basic unit of conversation to analyze the discourse structure systematically. We refer to this unit as the **utterance unit (UU)**, and define it using the following principles.

- **Grammatical Principle:** Place the UU boundary where a period could be put. In case of sentence conjunction, the UU boundary is set just before the conjunction.
- **Pragmatic Principle:** The UU should correspond to the basic speech-act. In other words, the UU should represent the speaker's basic intention. Please note that this does not rule out the case where one speech act continues over several UUs. Actually, the utterance corresponding to a single speech act can be broken down into discrete UUs by the following two principles.
- **Conversational Principle:** A UU boundary should be placed whenever the speaker changes. This includes the case of short acknowledgements such as *hnn-hnn* or *yes.*
- **Prosodic Principle:** The UU boundary is placed whenever a pause of medium length or longer occurs. The pause threshold is set at 750 msec which is a bit longer than the pauses called *search pauses* or *repair pauses.*

By applying these rules to the speech data, the recorded utterances were split into numbered UUs.

The discourse structure and the prosody analysis discussed in the following sections are based on the above UU definitions. That is, the topic boundary variations are viewed

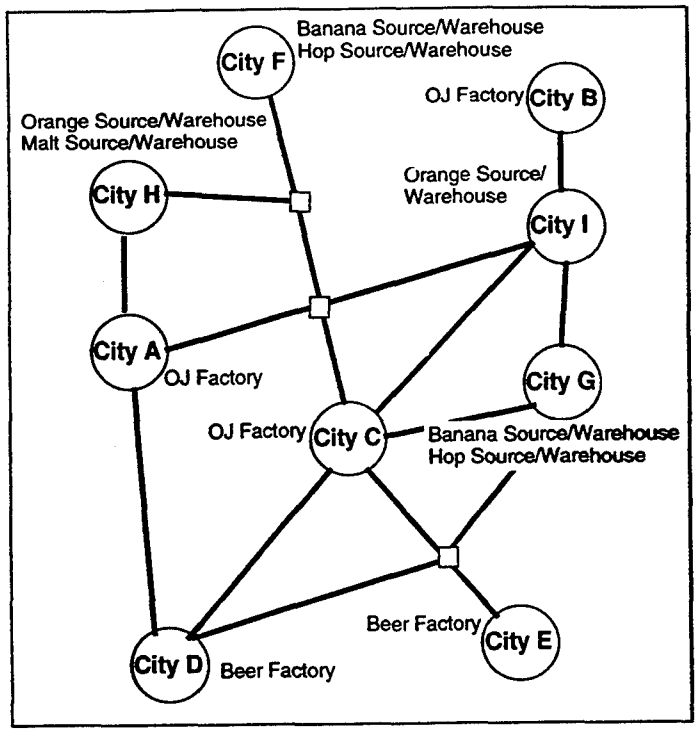

Figure 1: The TRAINS world for speech data collection. The cities (A - G) are connected to each other by rail lines (drawn in bold lines). Each city has either a manufacturing capability (OJ or beer factory), or storage capbility. Transportation is supplied by engines, boxcars, and tankers which are initially placed at specific cities.

as the relationships between the current UU and the previous UU(s), and the prosodic parameters are measured for each UU.

3.2 Topic Boundary Types

The model of discourse structure and the taxonomy for the relation between discourse segments have been discussed elsewhere (e.g. Cohen [6], Hobbs [13], Mann and Thompson [16], Grosz and Sidner [9]). Since our objective here is to investigate the correlation between prosody and the discourse structure, the relations between UUs were simplified and we categorized the topic boundaries into four classes: **Topic Shift, Topic Continuation, Elaboration**, and **Speech-Act Continuation**. These can be defined as follows. (Typical examples in our corpus are shown in figure 2i,ii.)

Topic Shift (TS) This class can be viewed as three subclasses;

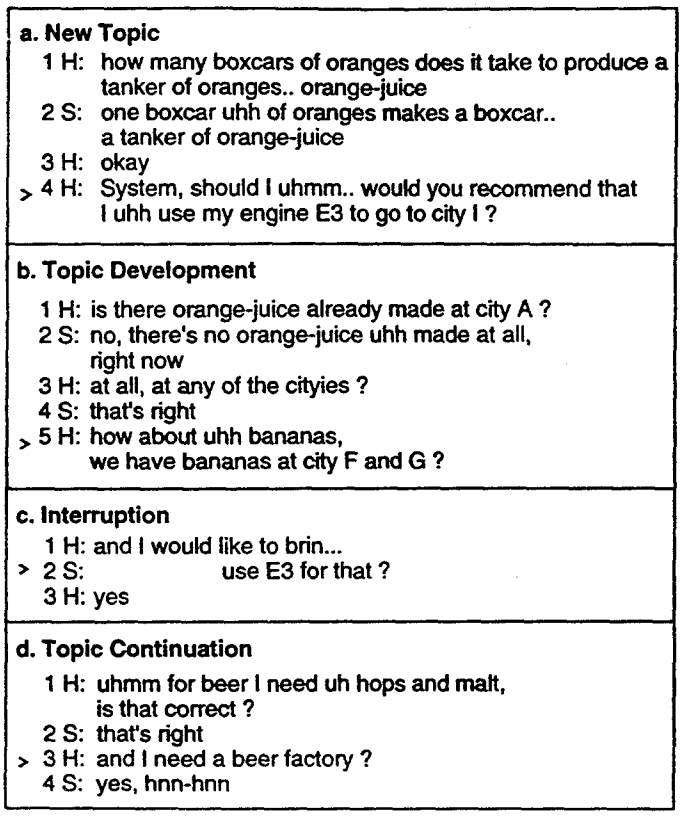

Figure 2i: Typical utterance sequence of each topic boundary class. '>' marks the place where that boundary class occurs. H and S indicates speaker Human (user) and System, respectively.

New Topic (NT) The current UU introduces a new topic. In our TRAINS domain, since **S** and **H** try to cooperate to achieve a particular goal, such utterances on new (sub)goal or new (sub)plan are taken as NT, rather than completely independent topics. In figure 2i-a, after asking some questions, **H** introduces a new plan at utterance 4.

Topic Development (TD) The topic in the previous utterances is further developed in the current utterance and there might be some weak linkage between them. In figure 2i-b, at utterance 5, **H** shifts his focus from the orange juice to the bananas, but there is a shared topic between them, namely, *search for resources involved in the goal.*

Interruption (Int) The previous or simultaneous utterance is interrupted abruptly by the current utterance. In figure 2i-c, utterance 1 is interrupted by **S**'s ques-

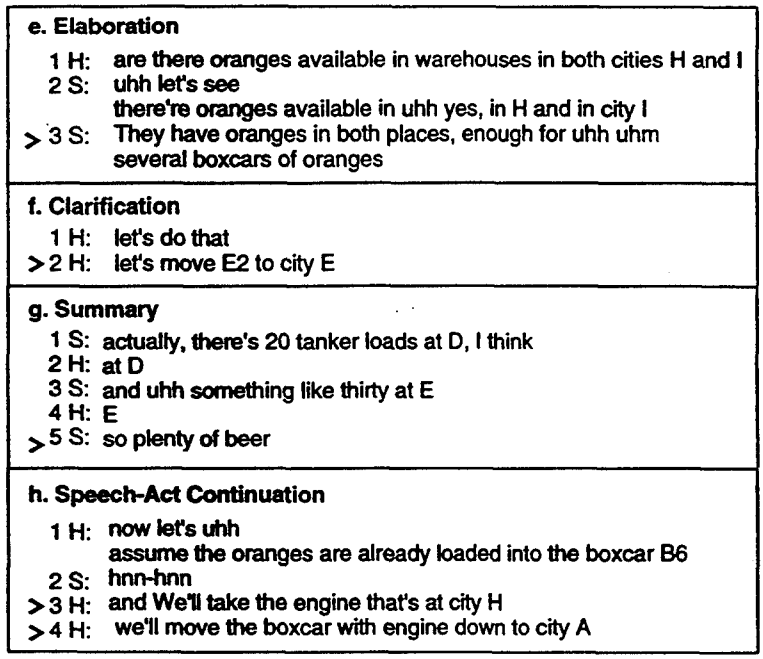

Figure 2ii: Typical utterance sequence of each topic boundary class. '>' marks the place where that boundary class occurs. H and S indicates speaker Human (user) and System, respectively.

tion.

Topic Continuation (TC) The linkage between the current topic and the previous one is comparatively strong. The current utterance may be talking about the same plan or the same entity discussed in the previous utterance. In figure 2i-d, at utterance 3, **H** continues to talk about *making beer*.

Elaboration Class (ELB) This class also can be viewed as covering three subclasses. The general interpretation of this class is that, the current utterance adds some relevant information to the previous utterance(s).

Elaboration (Elab) The current utterance adds some relevant information to the previous statement. In figure 2ii-e, **S** informs **H** of the quantity of the oranges which **S** believes is relevant to **H**'s last question.

Clarification (Clr) The current utterance clarifies some propositions made in the previous utterances. In figure 2ii-f, **H** restates his proposal while clarifying what *do that* really means.

Summary (Summ) The current utterance summarizes the contents of the preceding utterances, as shown in figure 2ii-g.

Speech Act Continuation (AC) A single speech act continues over several UUs. Most of them are sequential conjunctions as shown in figure 2ii-h.

In the following section, we describe how some prosodic parameters depend on the topic boundary classes and how the variation can be interpreted from the pragmatic viewpoint.

4 Prosody and Discourse Structure

By using the recording setup described in the previous section, we collected a total dialogue duration of about one and half hours from 5 goal-achieving sessions which were performed by two male speakers (both were native speakers of English but not professional). The dialogues consisted of 1025 utterance units. The topic boundaries were marked by both authors and those UUs whose topic boundaries could not be determined by either of the authors were excluded from the analysis. Fundamental frequencies were measured by using a KAY sonograph. The points at which the prosodic parameters could not be measured stably were also ignored. As a result, about 500 UUs were used in the following analysis.

4.1 Onset Fundamental Frequency

A number of analysts have suggested that onset F0 is raised when the topic of the conversation is changed. (e.g. Brown, Currie, and Kenworthy [4]) However, to the best of our knowledge, clear and reliable confirmation has yet to be shown. In order to clarify how this prosodic tendency is reflected in the topic boundary classes of our database where acknowledgements and interruptions are frequently made by the participants, we investigated the onset F0 at each topic boundary class.

For analysis consistency, we excluded the cases in which a single grammatical phrase (e.g. noun-phrase, prepositional-phrase, and so on) is split into several UUs via the prosodic principle. For instance, the cases like (H: "from city...") [1 sec. pause] (H: "G") were excluded. Since we are focusing here on the relationship between topic-shifting and onset F0, we also excluded simple answer utterances.

Average onset F0 (hereafter $F0_S$) at each topic boundary class is shown in figure 3, and the number of samples, averages, and standard deviations are given in table 1. The results can be summarized as follows;

- For each speaker, $F0_S$ value declines in the order;

  $$TS > TC > ELB \approx AC$$

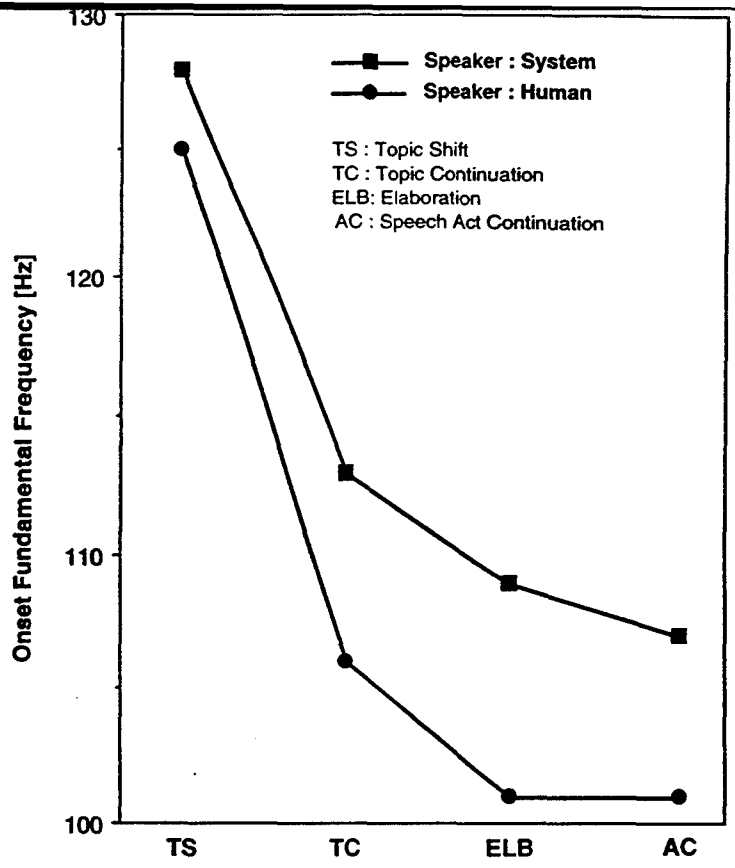

Figure 3: Onset fundamental frequency average at each topic boundary of both speakers (Human and System).

Table 1: Number of samples (N), average, and standard deviation (s.d.) of onset F0 at each topic boundary; TS– Topic Shift, TC– Topic Continuation, ELB– Elaboration, AC– Speech-act Continuation.

	System			Human		
boundary type	N	average	s.d.	N	average	s.d.
TS	24	128	9.7	38	125	11.3
TC	41	113	5.6	33	106	5.6
ELB	52	109	4.9	25	101	6.2
AC	119	107	5.6	68	101	5.9

In particular, for both speakers, the distinction between TS and other boundary classes is much more significant than the other differences.

- Average $F0_S$ value at the ELB boundaries and that at the AC boundaries are

almost identical for both speakers. This result suggests that as far as the onset F0 is concerned, the prosodic connection between the previous and the current elaboration utterance is as strong as that of speech act continuation utterances.

In the above analysis, $F0_S$ values were simply measured at the beginning of the first stable part of F0 contours, rather than at first stressed syllable; the aim was automatic topic boundary identification via prosody. In fact, some of the measured points were stressed and some were not. These results suggest that in spontaneous conversations, the onset F0 values, even at not-stressed syllables (stable enough to measure), can be correlated to some extent with the topic boundary classes.

4.2 Final Fundamental Frequency

As suggested in the literature, the final boundary tone reflects *finality* or *completeness* of the statement in declarative sentences. We investigated the correlation between final F0 ($F0_F$) and topic boundary class to show how this tendency is reflected in actual F0 contours.

Table 2: Number of samples (N), average, and standard deviation (s.d.) of final F0 at each topic boundary; END– Topic Shift and end of isolated answer, TC– Topic Continuation, ELB– Elaboration, AC– Speech-act Continuation.

	System			Human		
boundary type	N	average	s.d.	N	average	s.d.
END	81	94	3.3	44	88	5.9
TC	28	96	5.1	19	93	8.5
ELB	34	97	6.8	17	92	7.9
AC	147	113	15.7	51	108	7.5

The $F0_F$ of single utterance answers, not followed by any subsequent utterances, were counted together with those of TS boundaries and treated as constituting the END class. This is because there is no significant distinction between isolated answers and topic shift boundaries.

The average $F0_F$ value at each topic boundary is shown in figure 4, and the number of samples, averages, and standard deviations of final F0 frequency are shown in table 2.

As can be seen in the figure, for both speakers **S** and **H**, final F0 is much higher at AC boundaries than at other boundaries. Moreover, $F0_F$ values at boundaries other than AC are almost identical. Thus, final fundamental frequency can be taken as a good cue for discriminating AC boundaries from other boundaries.

The results in section 4.1 suggest that as far as onset F0 is concerned, the prosodic connection at elaboration boundaries is as strong as that of speech-act continuation, whereas

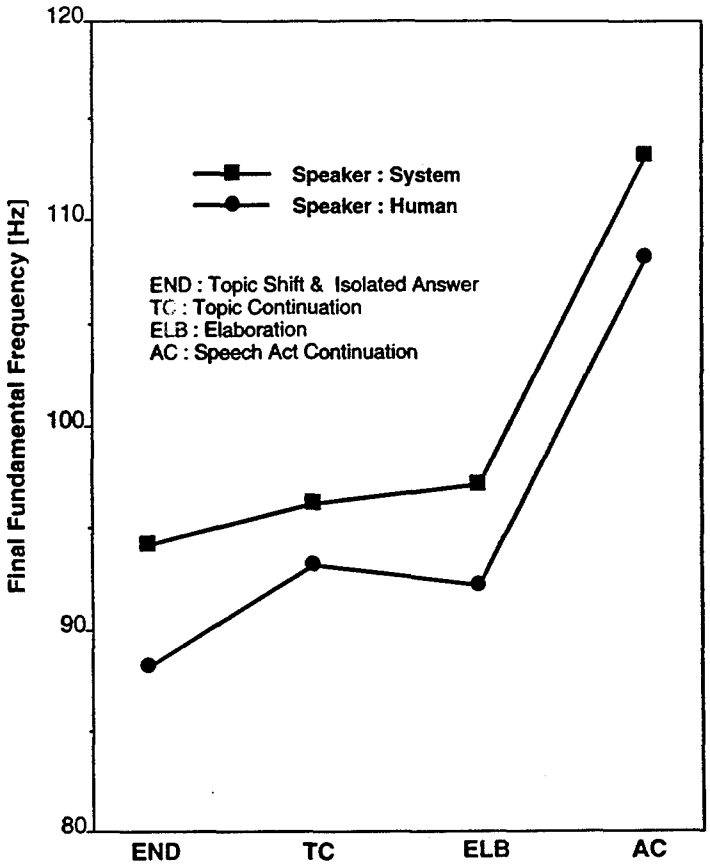

Figure 4: Final fundamental frequency average at each topic boundary of both speakers (Human and System).

the final F0 result indicates a considerable difference between speech-act continuation and elaboration utterances. However, this phenomenon can be explained by the semantic definition of elaboration class boundary and the pragmatic roles of prosody. At an elaboration boundary, the previous utterance UU_0 *per se* completes a particular statement, and the succeeding elaboration utterance UU_1 adds some relevant information to UU_0. So, the completeness of UU_0 leads to the final F0 lowering and the following relevant utterance influences the onset F0 value of UU_1.

We note that when measuring the final F0 values, we do not distinguish rising tones from falling tones. Actually, however, while rising tones are the most typical F0 contours at AC boundary, we have found some *half completion* falling contours (term comes from Gussenhoven [10]), where the F0 falls to mid-level. The $F0_F$ values of this sort at AC boundaries also pulled up the average and can be taken as indicating the non-finality of

the utterance.

4.3 Peak F0 Ratio

It is claimed that within continuous speech, the peak F0 range of each intonational phrase declines towards the end of sentences (e.g. Hakoda and Sato [11], Liberman and Pierrehumbert [15], Ladd [14]). Hakoda and Sato [11] also suggested that as the grammatical connection between two neighboring phrases increases, the peak F0 of the second phrase is suppressed more relative to the first phrase. In this section, we examine this tendency in a sequence of linked utterance units, and show how it is reflected in each topic boundary class.

To investigate the degree of declination, we use the ratio of the current UU's maximal peak F0 to that of the previous one. That is, the maximal peak F0 of the current UU_1 ($F0_{P1}$) and that of the same speaker's previous UU_0 ($F0_{P0}$) are measured. The declination ratio of maximal peak F0 (R_P) is then computed as follows. (Hereafter, we call this parameter simply *the peak F0 ratio.*)

$$R_P = \frac{F0_{P1}}{F0_{P0}}$$

The average peak F0 ratio is shown in figure 5, while the number of samples, averages, and standard deviations are shown in table 3.

Table 3: Number of samples (N), average, and standard deviation (s.d.) of peak F0 ratio at each topic boundary; TS– Topic Shift, TC– Topic Continuation, ELB– Elaboration, AC– Speech-act Continuation.

	System			Human		
boundary type	N	average	s.d.	N	average	s.d.
TS	27	1.15	0.12	32	1.17	0.16
TC	34	1.00	0.10	44	0.97	0.12
AC	121	0.95	0.09	58	0.94	0.10
ELB	34	0.89	0.07	21	0.89	0.07

The results can be summarized as follows;

- For both speakers, the peak F0 ratio declines in the order;

  TS > TC > AC > ELB

- The peak F0 ratio is around 1.15 at TS boundaries, and is around 1.0 at TC boundaries. This suggests that if the topic changes, the speaker starts speaking with

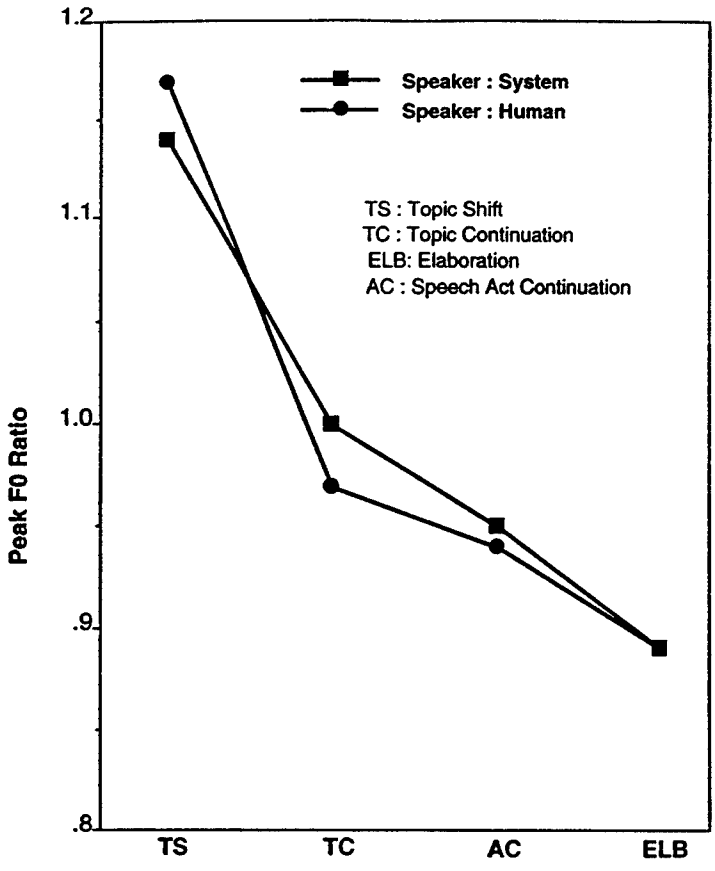

Figure 5: Peak F0 ratio average at each topic boundary of both speakers (Human and System).

higher peak F0 range and that if there's no salient relationship, and no abrupt topic shift between the two utterances, the speaker utters them with the same peak F0 range.

- For both speakers, the peak F0 ratio at ELB boundaries is lower than that at AC boundaries. This result can be interpreted as follows; the relationship between two utterances at an AC boundary is mostly coordinate, whereas elaboration utterances are sometimes subordinate to the previous ones. This subordination suppresses elaboration utterances more than coordination utterances.

- As can be inferred from figure 5, the peak F0 ratio is a reliable parameter with which to discriminate ELB boundaries from TC boundaries.

5 Summary and Discussions

5.1 Summary of the Results

The prosodic characteristics of each topic boundary class can be summarized as follows;

- When the topic changes, the onset F0 is high, the final F0 of previous utterance is low, and the maximal peak F0 is raised considerably (the ratio > 1.1).
- When the topic continues and there's no salient relation between previous and current utterances, the values of prosodic parameters are similar to that of the topic-shift, except that the onset F0 and the peak F0 ratio are slightly lower than in the topic-shift case.
- At speech-act continuation boundaries, the onset F0 is lower and the final F0 of the previous utterance is much higher than in other cases.
- Elaborating utterances are characterized by low onset and low final F0 values, and the maximal peak F0 is normally suppressed.

These results are listed in table 4.

Table 4: Prosodic characteristics at each topic boundary class; TS– Topic Shift, TC– Topic Continuation, ELB– Elaboration, AC– Speech-act Continuation.

boundary type	*Onset*	*Final*	*Peak Ratio*
TS	**High**	Low	**High** (> 1.1)
TC	**Mid**	Low	**Mid** (≈ 1.0)
AC	Low	**High**	**Mid** (≈ 0.95)
ELB	Low	**Low**	**Low** (< 0.9)

5.2 Discourse Structure Identification via Prosody

One application of these results is discourse structure identification via prosody, which is an important process for speech understanding systems. In table 4, the features typed in bold face are the key to boundary type discrimination. In Nakajima and Allen [18, 19], a boundary type discrimination tree was proposed.

Discourse structure identification can be viewed as having 2 levels: global and local. The global level is concerned with topic changes, that is, the discrimination between TS or TC. The local level corresponds to the identification of the fine structure of UUs which are uttered for the same discourse goal (by the same speaker). This level of identification ought to include not only the relation between UUs but also the hierarchical structure

of UUs. Since global level identification is fairly straightforward, we discuss local level identification in the rest of this section.

The utterance sequences shown in figures 6, 7, and 8 have typical discourse structures. Please note that the F0 contours in the figures are stylized by three parameters; onset, maximal peak, and final F0 values, and that they were extracted from an actual speech database.

Figure 6 shows a typical speech-act continuation utterance sequence. The maximal peak of each U_i declines towards the end of the sequence, indicating that the topic is not changed. The final F0 values other than that of U_3 are higher, showing that the relations between the UUs are speech-act continuations.

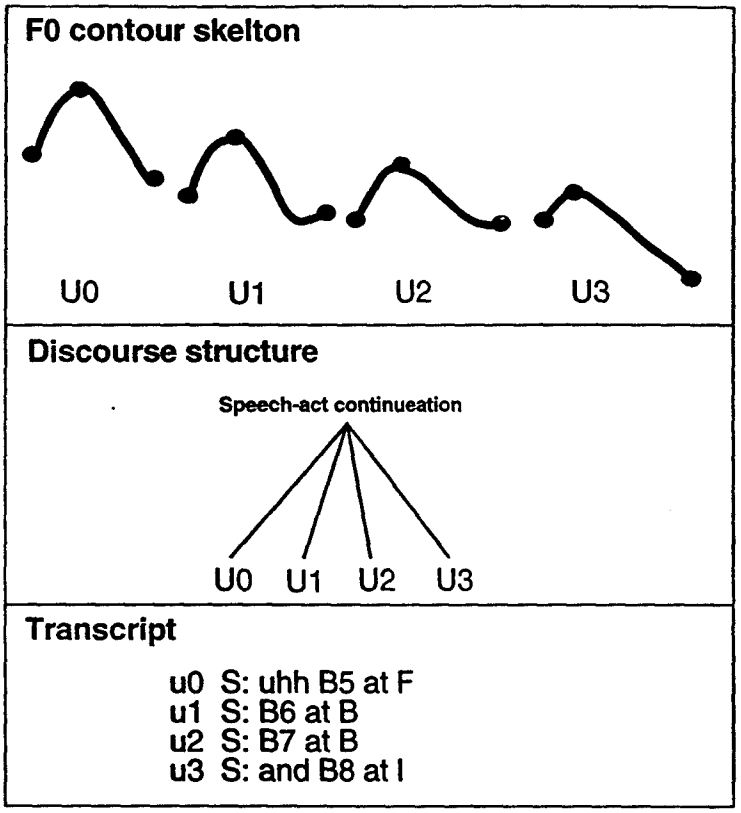

Figure 6: An utterance unit sequence sample; F0 contour, dicourse structure, and transcript of speech-act continuation. The F0 contour is stylized by 3 parameters: onset, peak, and final F0 values.

Figure 7 shows an elaboration-continuation hybrid case. U_1 and U_2's prosodic parameters suggest that the relation between them is speech-act continuation. U_0's final F0 shows the finality of its proposition and the maximal peak declination between U_0 and U_1 suggests that their relation may be elaboration. Consequently, these inferences lead

to the structure shown in the figure.

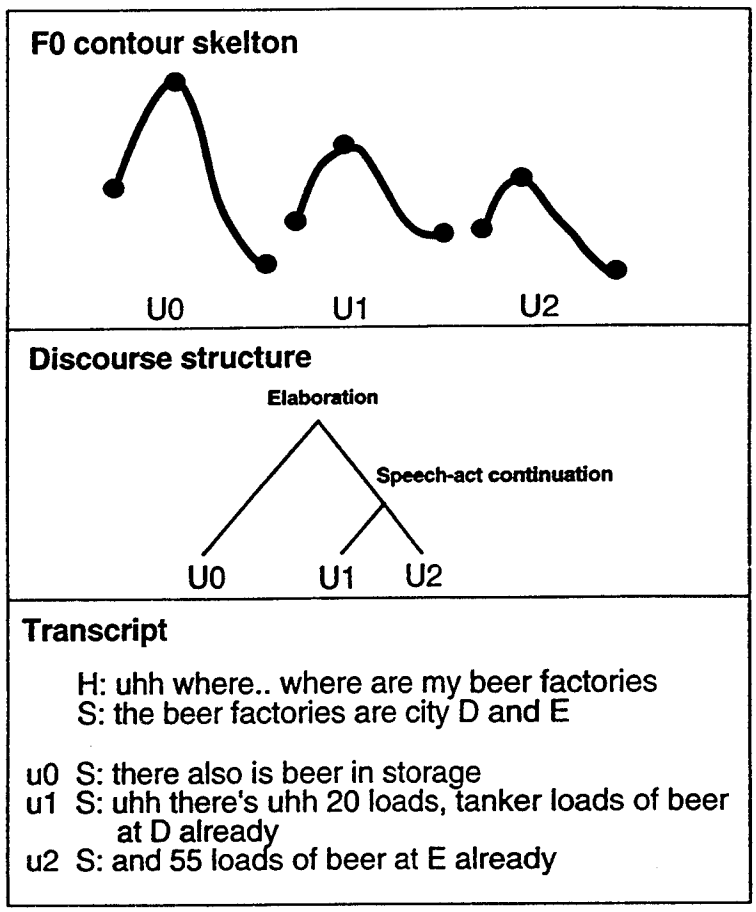

Figure 7: An utterance unit sequence sample; F0 contour, dicourse structure, and transcript of elaboration-speech-act continuation. The F0 contour is stylized by 3 parameters: onset, peak, and final F0 values.

The final sample shown in figure 8 is more complicated. U_1 and U_3 are elaborations of preceding utterances – U_0 and U_2, respectively– and the relation between U_0, U_2, and U_4 is continuation. In this case, discourse structure identification might be more complicated. U_1's first peak is largely suppressed, indicating that it is completely subordinate to U_0. Because of this subordination, U_0's high final F0 can be taken as indicating a continuation to U_2 rather than to U_1, and U_2's slightly lower maximal peak F0 also supports this inference. A similar analysis can be done for U_2, U_3, and U_4. To identify the structures of this sort, the order of the identification should be managed and a recursive mechanism should be utilized.

In order to develop a practical discourse structure identification algorithm, two problems must be overcome. First, as we have seen in the previous results, there is considerable

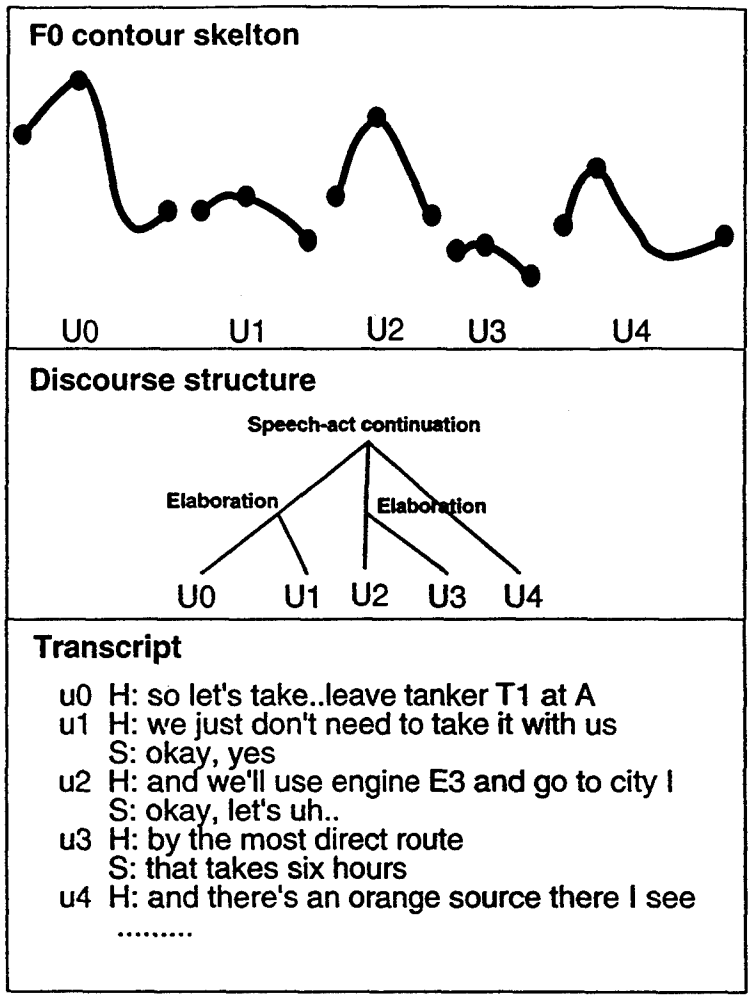

Figure 8: An utterance unit sequence sample; F0 contour, dicourse structure, and transcript of parenthetical (inserted) elaboration. The F0 contour is stylized by 3 parameters: onset, peak, and final F0 values.

difference in the F0 range depending on the speaker. Therefore, a normalizing technique should be utilized to eliminate this effect. Another problem is that, since the prosodic phenomena described above reflect statistical effects, literal information such as cue phrases should be also taken into account together with prosody. The following literal information will be useful in identifying the discourse structure.

- Clue words; *okay, so, now, well*
  If used with falling intonation, these clue words are often used as topic shift markers, and deaccented *so* is a good cue for indicating summarisation.
- Vocative; *System*

In our speech database, vocative *System* is always used at topic shift boundaries

- Form of question;
  Wh-questions are frequently used at topic shift boundaries, and declarative/tag-questions are normally used at topic continuation boundaries.

Investigating such literal cues and showing how they can be used in combination with the prosodic cues are beyond this article and are left as a future task.

5.3 Prosodic Parameter Generation

Another application of the results is to develop prosodic parameter generation rules for speech synthesis. The simplest generation method is to use a table such as table 4. For instance, if the first utterance UU_0 introduces a new topic, the onset and maximal peak F0 values are higher, and if UU_1 elaborates UU_0, UU_1's maximal peak F0 value should be 90% of the value of UU_0, and so on. A similar analysis of cooperative Japanese dialogues led to the more detailed prosodic parameter generation rules proposed in Nakajima [17].

6 Conclusion

In natural conversations, the speaker uses prosodic features to convey structural information. When the topic changes, the speaker starts speaking with raised onset and peak F0 values, and when the topic continues, but there's no specific relationship between current and previous utterances, the speaker produces them with the same peak F0 range. By using higher final F0 and slight declination of peak F0, the speaker indicates that the propositional contents of the utterances are not finished, and by lowering the final F0 and following it by an utterance with suppressed peak F0, the speaker suggests that this utterance elaborates the previous utterance(s).

Acknowledgements

Many thanks to Tim Becker for kindly being our subject, and also to David Traum for his fruitful suggestions on discourse marking.

References

[1] Allen, J.F.; Perrault, C.R.: *Analyzing intention in utterances.* Artificial Intelligence 15: 143-178 (1980).

[2] Allen, J.F.; Schubert, L.K.: *The* TRAINS *project.* TRAINS Technical Note 91-1, Computer Science Dept, University of Rochester (1991).

[3] Austin, J.L.: *How to do things with words.* Oxford University Press (1962).

[4] Brown, G.; Currie, K.L.; Kenworthy, J.: *Questions of intonation.* Croom Helm (1980).

[5] Brown, G.; Yule, G.: *Discourse analysis.* Cambridge University Press (1983).

[6] Cohen, R.: *Analyzing the structure of argumentative discourse.* Computational Linguistics 13: (1987).

[7] Fujisaki, H.: *An analysis and model for Japanese prosody: the relation between prosody and the syntactic/discourse structure.* in Sugitoh, M (ed.): Japanese and Japanese Education 2.: 266-297 (1989), (in Japanese).

[8] Grosz, B.; Hirschberg, J.: *Some intonational characteristics of discourse structure.* Proceedings of International Conference on Spoken Language Processing: 429-432 (1992).

[9] Grosz, B.; Sidner, C.: *Attention, intentions, and the structure of discourse.* Computational Linguistics 12(3): 175-204 (1986).

[10] Gussenhoven, C.: *On the grammar and semantics of sentence accents.* Language Sciences 16: (1983).

[11] Hakoda, K.; Sato, H.: *Prosodic rules in connected speech synthesis.* Systems Computers Controls 11-5: 28-37 (1980).

[12] Hirschberg, J.; Pierrehumbert, J.: *The intonational structuring of discourse.* Proceedings of the Twenty-fourth Annual Meeting, Association for Computational Linguistics: 136-144 (1986).

[13] Hobbs, J.: *Coherence and coreference.* Cognitive Science 3(1): (1979).

[14] Ladd, D.R. *Declination: a review and some hypotheses.* Phonology Yearbook I: (1984).

[15] Liberman, M.; Pierrehumbert, J.B.: *Intonational invariance under changes in F0 range and length,* in Aronoff M.; Oehrle R.T.(eds.): *Language sound structure.* MIT Press (1984).

[16] Mann, W.C.; Thompson, S.A.: *Rhetorical structure theory: description and construction of text structures.* in Kempen, G.(ed.): Natural Language Generation.: 85-96, Martinus Nijhoff Publishers (1986).

[17] Nakajima, S.: *Some prosodic characteristics of Japanese cooperative dialogues.* Proc.of Fall Meeting 1-1-10, The Acoustical Society of Japan (1992), (in Japanese).

[18] Nakajima, S.; Allen, J.F.: *Prosody as a cue for discourse structure.* SIG-Human Interface 38-6, Information Processing Society of Japan (1991).

[19] Nakajima, S.; Allen, J.F.: *Prosody as a cue for discourse structure.* Proceedings of International Conference on Spoken Language Processing: 425–428 (1992).

[20] Searle, J.R: *Speech Acts.* Cambridge University Press (1969).