ヒューマンインタフェース 38－6
（1991. 9. 12）

英語会話の話題展開と韻律の関係について

中嶋 信弥
NTT ヒューマンインターフェース研究所

ジェームス アレン
ロチェスター大学

1991年8月20日

概要

本報告では，韻律情報と英語対話における話題展開との関係について述べる．まず，会話における基本的単位として，「発話ユニット」という概念を導入する．次に，発話ユニット間の話題のつながり具合を基にして，発話展開の種類を次の4種に分類した；話題変更 (Topic Shift)，話題継続，関連情報の付加，同一発話意図の継続．韻律情報パラメータとしては，始端，第一ピーク，終端ピッチ周波数等を用い，各話題展開境界によって，これらの韻律パラメータがどのような性質をもつか分析した．最後に，この結果から，韻律情報による話題展開の認識アルゴリズムを提案する．

Prosody as a Cue for Discourse Structure

Shin'ya Nakajima
NTT Human Interface Laboratories

James F. Allen
University of Rochester

August 20, 1991

Abstract

In natural conversations, prosodic infromation plays several pragmatic roles. Of these, this paper describes how well prosodic information correlate with the topic structure of discourse. To investigate this correlation systematically, first we introduce the notion of *utterance unit* which can be viewed as a basic unit in conversations. We then define four topic boundary classes; *Topic Shift, Topic Continuation, Elaboration, and Speech-Act Continuation.* The prosodic parameters– onset/first-peak pitch, final pitch, and onset/first-peak pitch ratios– are measured at these topic boundaries to show how the pragmatic roles of prosody are reflected in actual pitch contours. Finally, we propose a schematic algorithm which identifies the topic boundaries via the prosodic parameters.

1 Introduction

The last decade has seen substantial progress in discourse processing and computational linguistic fields. Specifically, the plan recognition approaches based on Austin and Searl's speech-act theory [Austin 62, Searle 69] have been proposed (e.g. [AllPer 80]). However, although a number of analysts have pointed out that prosody plays several important roles in natural conversations (e.g. [BroYul 83, PieHir 90]), there have been very few studies that take account of the prosodic information. In general, the intentional meaning of the utterance in a conversation cannot be determined without refering prosodic information.

Prosodic information plays various pragmatic roles in a conversation; The most salient 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, and we will show how prosodic information can be utilized as a cue for topic structure identification.

In the next section, we introduce our specific task domain– TRAINS world [AllSch 91]– and describe how we have collected natural conversations. We then define the topic structure markers which are based on the notion of *utterance unit*. Finally, we will show how well particular prosodic parameters correlate with the topic structure and propose a schematic algorithm which identifies the topic structure from the prosodic parameters.

2 Speech Data Collection

The map of the TRAINS world is shown in Fig.2-1. The cities in the TRAINS world are connected to each other by rail lines. Each city have either a manufacturing capability (OJ factory or beer factory), or storage capability. Transportation is supplied by engines, boxcars, and tankers which are initially placed at specific cities.

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.

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 of digital audio tape. We collected a total dialogue duration of about one and half hours from six goal-achieving sessions.

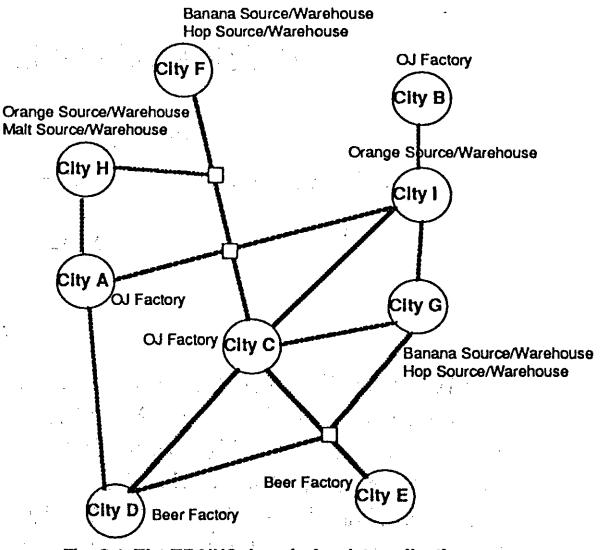

Fig. 2-1 The TRAINS domain for data collection

3 Discourse Structure Marking

3.1 Utterance Unit

Since grammatical units such as *sentences* are absent in the 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)** which can be determined by 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 a basic speech-act. In other words, 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 to discrete UUs by the following two principles.
- **Conversational Principle;** A UU boundary should be placed whenever speaker changes. This includes the case of short acknowledgement such as *hnn-hnn* or *yes*.

- **Prosodic Principle;** The UU boundary is placed whenever a medium length or longer pause occurs. The pause threshold is set to 750 msec which is a bit longer than the pauses called *search pauses* or *repair pauses*.

By appllying these rules to the speech data, the utterances were split into numbered UUs. Ex.3-1 shows typical UU analysis. The utterance in Ex.3-1a is split into two UUs; the first UU, *okay*, is an acknowledgement, and the second UU is WH-question. Ex.3-1b shows the case in which S's acknowledgement *hnn-hnn* is inserted in the middle of H's statement.

H: okay, how long will it take for engine E3 to go to city I

⇓

(H:uu1 okay)
(H:uu2 how long will it take for engine E3 to go to city I)

Ex.3-1a Utterance Unit analysis; including *okay*.

H: let's uhh move engine E3...
S: hnn-hnn
H: to city I

⇓

(H:uu1 le's uhh move engine E3)
(S:uu2 hnn-hnn)
(H:uu3 to city I)

Ex.3-1b Utterance Unit analysis; acknowledgement.

The discourse structure and the prosody analysis discussed in the following sections are based on UU as defined. That is, the topic boundary variations are viewed as the relationships between the current UU and the previous UUs, and the prosodic parameters are measured for each UU.

3.2 Topic Boundary Types

To investigate the correlation between prosody and the discourse structure, we categorized the topic boundary into four classes: **Topic Shift, Topic Continuation, Elaboration,** and **Speech Act Continuation**. These can be defined as follows. (Actual examples are shown in Ex.3-2.)

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

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 Ex.3-2a, after asking some questions, H introduces a new plan at utterance 4.

Topic Development (TD) The topic in the previous utterances is further developed at the current utterance and there might be some weak linkage between them. In Ex.3-2b, 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 Ex.3-2c, utterance 1 is interrupted by S's question.

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 as discussed in the previous utterance. In Ex.3-2d, at utterance 3, H continues to talk about *making beer.*

Elaboration Class (ELB) This class also can be viewed as 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 Ex.3-2e, S informs H of the quantity of the oranges which S believes relevant to H's last question.

Clarification (Clr) The current utterance clarifies some propositions involved in the previous utterances. In Ex.3-2f, 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 Ex.3-2g.

Speech Act Continuation (AC) A single speech act continues over several UUs. Most of them are sequential conjunctions as shown in Ex.3-2h.

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

Ex. 3-2 Examples of each Discourse Segment Boundaries

A. New Topic

1 H: how many boxcars of oranges does it take to produce a tanker of oranges.. orange-juice
2 S: one boxcar uhh of oranges makes a boxcar.. a tanker of orange-juice
3 H: okay
> 4 H: System, should I uhmm.. would you recommend that I uhh use my engine E3 to go to city I ?

B. Topic Development

1 H: is there orange-juice already made at city A ?
2 S: no, there's no orange-juice uhh made at all, right now
3 H: at all, at any of the cityies ?
4 S: that's right
> 5 H: how about uhh bananas, we have bananas at city F and G ?

C. Interruption

1 H: and I would like to brin...
> 2 S: use E3 for that ?
3 H: yes

D. Topic Continuation

1 H: uhmm for beer I need uh hops and malt, is that correct ?
2 S: that's right
3 H: and I need a beer factory ?
4 S: yes, hnn-hnn

E. Elaboration

1 H: are there oranges available in ware houses in both cities H and I uhh let's see
2 S: there're oranges available in uhh yes, in H and in city I
> 3 S: They have oranges in both places, enough for uhh uhm several boxcars of oranges

F. Clarification

1 H: let's do that
> 2 H: let's move E2 to city E

G. Summary

1 S: actually, there's 20 tanker loads at D, I think
2 H: at D
3 S: and uhh something like thirty at E
4 H: E
> 5 S: so plenty of beer

H. Speech Act Continuation

1 H: now let's uhh assume the oranges are already loaded into the boxcar B6
2 S: hnn-hnn
> 3 H: and We'll take the engine that's at city H
> 4 H: we'll move the boxcar with engine down to city A

4 Prosody and Discourse Structure

4.1 Onset and First Peak Pitch Frequencies

A number of analysts have suggested that onset and first peak pitch are raised when the topic of the conversation is changed. (e.g. [BroCur 80]) However, to my best knowledge, clear and reliable confirmation has yet to be shown. In order to clarify how this prosodic tendency reflects on the topic boundary classes of our database where acknowledgements and interruptions are frequently made by the participants, we investigated the onset/first peak pitch frequency at each topic boundary class.

The measuring points of onset pitch (Po) and first peak pitch (Pp) are illustrated in Fig.4-1. For analysis consistency, we excluded the cases in which a single gramatical 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:uu1 from city...) [1 sec. pause] (H:uu2 G) were excluded. Since we are focusing here on the relationship between topic-shifting and onset/peak pitch, we also excluded simple answer utterances.

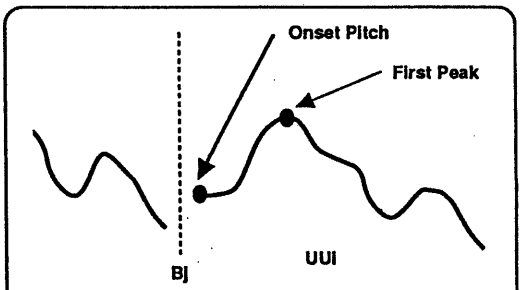

Fig.4-1 The measuring points of onset /first peak pitches

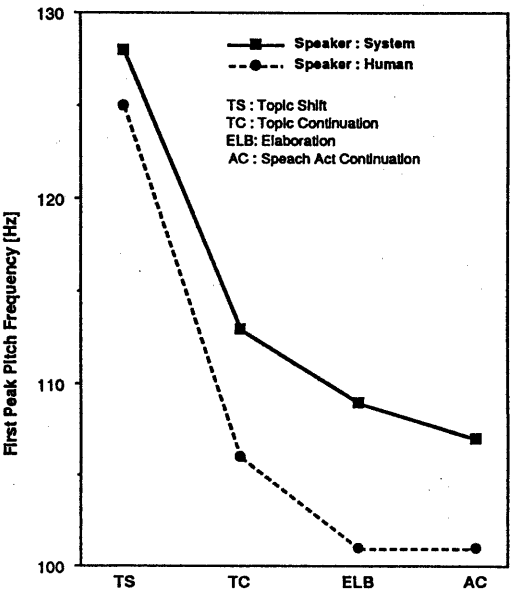

Fig. 4-2A Onset Pitch Frequency at each Topic Boundary

Onset/first peak pitch averages at each topic boundary class are shown in Fig.4-2. The results can be summarized as follows;

- For each speaker, both Po and Pp decline in the order;

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

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

- Po/Pp at ELB boundary and those at AC boundary are almost identical for both speakers. This result suggests that as far as Po and Pp are concerned, the prosodic connection between the previous and the current elaboration utterance is as strong as that of speech act continuation utterances.

- From the T-distribution tests, the statistical significance of Po is higher than Pp's for all cases. That is, onset pitch is a more reliable parameter than the first peak pitch, at least in terms of topic boundary class identification.

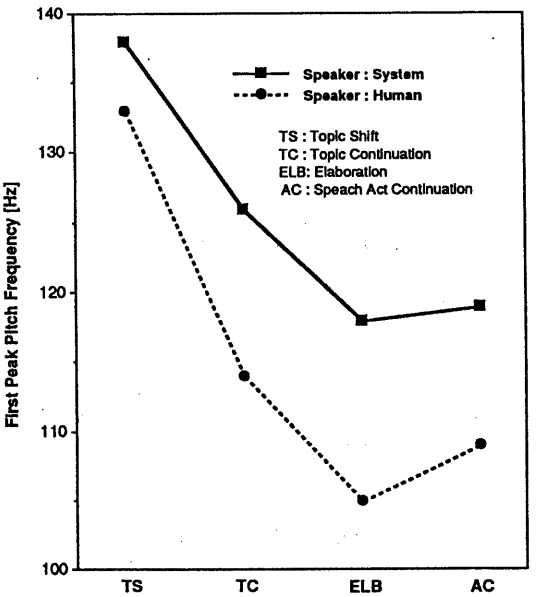

Fig.4-2B First Peak Pitch at each Topic Boundary

4.2 Final Pitch 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 pitch frequency (Pf) and topic boundary class to show how this tendency is reflected in actual pitch contour.

The measuring point of Pf is illustrated in Fig.4-3. The final pitch of single answers, not followed by any subsequent utterances, are counted together with those of TS boundaries and refered as END class. This is because there is no significant distinction between the isolated answers and the topic shift boundaries.

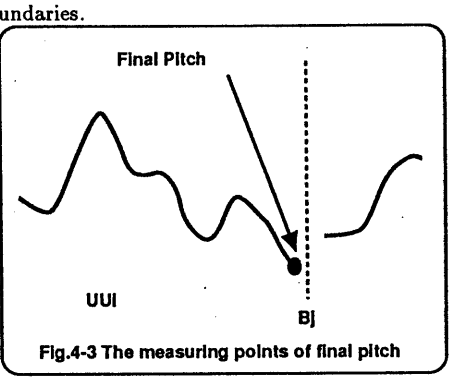

Fig.4-3 The measuring points of final pitch

The average of final pitch frequency at each topic boundary is shown in Fig.4-4.

As can be seen in the figures, for both speakers S and H, final pitch is much higher at AC boundaries than at other boundaries. Moreover, Pfs at boundaries other than AC are almost identical. Thus, final pitch frequency can be taken as a good cue for discriminating AC boundaries from other boundaries.

The previous results suggest that as far as onset and first peak pitch are concerned, the prosodic connection at the elaboration boundary is as strong as that of speech-act continuation, whereas the final pitch result indicates considerable isolation between the previous and elaboration utterances. However, this phenomena 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 pitch lowering and the following relevant utterance influences on the onset and first peak pitch values of UU_1.

We'd like to note that when measuring the final pitch frequencies, we do not discriminate rising tones from falling tones. Actually, however, while rising tones are the most typical pitch contours at AC boundary, we have found some so called *half completion* falling contours [Gussenhoven], where the pitch

falls to mid-level. This fall can be also taken as indicating non-finality of the utterance.

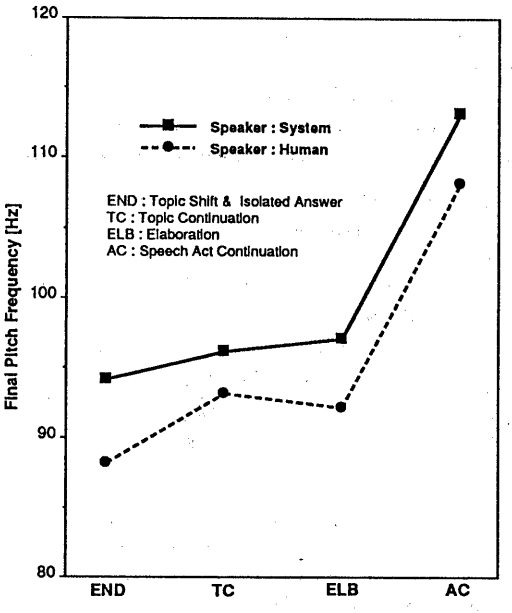

Fig. 4-4 Final pitch frequency at each topic boundary

4.3 Onset and First Peak Ratio

It is claimed that within a continuous speech, the peak pitch range of each intonational phrase declines towards the end of sentences [HakSat 80, LiePie 84, Ladd 84]. [HakSat 80] also suggested that as the grammatical connection between two neighboring phrases increases, the peak of the second phrase is suppressed more relative to the first phrase.

In this section, we extend the application of this tendency, from sentence speech to a sequence of linked utterance units, and show how this phenomenon is reflected in each topic boundary class.

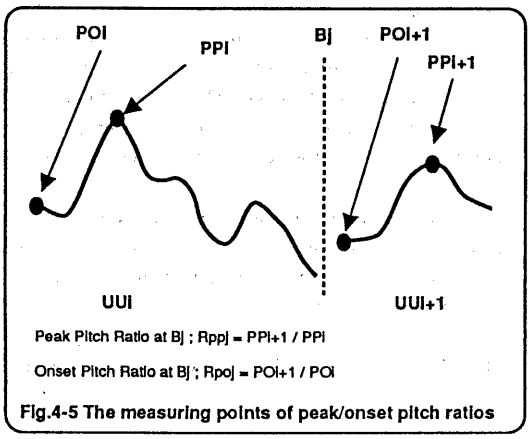

Fig.4-5 The measuring points of peak/onset pitch ratios

To investigate the degree of declination, we use the ratio of the current UU's first peak pitch (or onset pitch) to that of the previous one. The measuring points are illustrated in Fig.4-5. As illustrated in the figure, both onset and first peak frequencies of the current UU_1 (Po1, Pp1) and the previous (same speaker's) UU_0 (Po0, Pp0) are measured. Then the declination ratios of onset pitch (Rpo) and first peak pitch (Rpp) at boundary B_j are computed as follows.

$$\mathrm{Rpp} = \frac{\mathrm{Pp}_1}{\mathrm{Pp}_0}, \quad \mathrm{Rpo} = \frac{\mathrm{Po}_1}{\mathrm{Po}_0}$$

We refer to the former as peak pitch ratio (Rpp), and the latter as onset pitch ratio (Rpo).

The averages of peak pitch ratio and onset pitch ratio are shown in Fig.4-6. The results can be summarized as follows;

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

  TS > TC > AC > ELB

  The onset pitch ratio also shows a similar tendency, but the distinction between the boundary classes other than TS is less significant than in the case of the first peak ratio.

- Both peak and onset pitch ratios are larger than 1.0 at TS boundaries. This result means that these parameters are raised at TS boundaries (about 1.15 times) relative to those of previous utterance. The peak pitch ratio at TC boundaries is around 1.0, so, this suggests that if there's no salient relationship and no abrupt topic shifting between two utterances, the speaker utters them with the same peak pitch range.

- For both speakers, both ratios (Rpp, Rpo) at ELB boundaries are slightly lower than those 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 cordination utterance.

- As can be infered from Fig.4-6, the peak pitch ratio is a more reliable parameter than onset pitch ratio in terms of topic boundary identification. In other words, declination or suppression tendency is more salient on the *top line*

than on the *base line*. Moreover, only the peak pitch ratio can discriminate ELB boundaries from TC boundaries reliably.

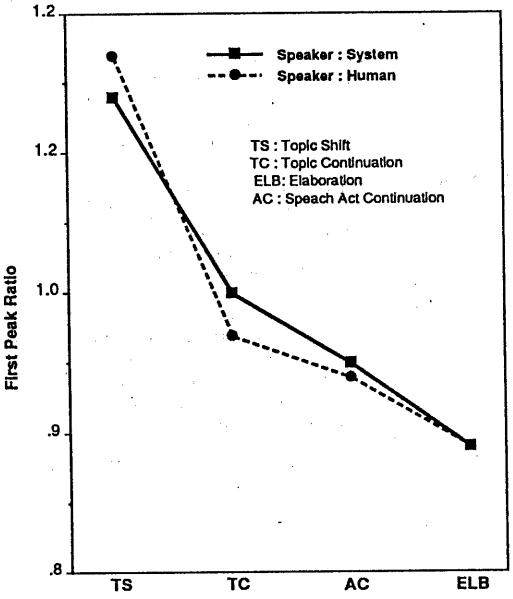

Fig. 4-6A First Peak Ratio at each Topic Boundary

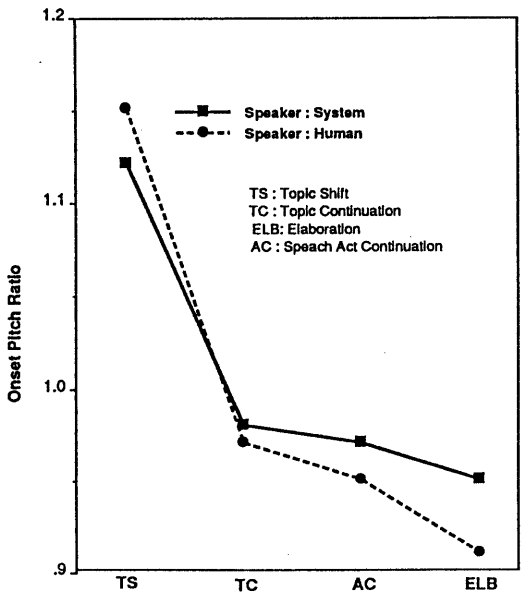

Fig. 4-6B Onset Pitch Ratio at each Topic Boundary

4.4 Topic Boundary Identification via Prosody

In this section, we discuss how our results can be utilized for topic boundary identification. From this point of view, the results shown above can be summarized as follows;

- Onset pitch is the best parameter to discriminate topic shift boundaries.
- Final Pitch is the best parameter to locate speech-act continuation boundaries.
- To discriminate elaboration boundaries from topic continuation boundaries, peak pitch ratio can be used.

These conclusions lead to the topic boundary discrimination tree described in Fig.4-7.

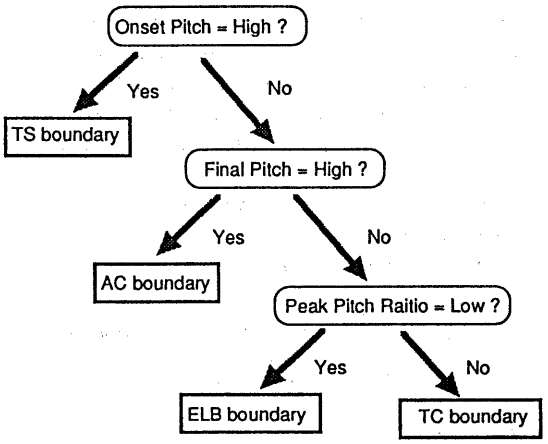

Fig.4-7 The topic boundary discrimination tree

5 Discussions

To develop a practical topic boundary discrimination algorithm, two problems must be overcome. First, as we have seen in the previous results, there is a considerable difference in pitch range depending on the speaker. Therefore, a sort of normalizing technique should be utilized to eliminate this effects. Another problem is that, since the prosodic phenomena described above reflect statistical effects, literal information should be also taken into account together with prosody. The following literal information will useful in identifying the topic 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.

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

In this paper, we have been focusing on the correlation between prosodic information and the topic boundaries. However, threre might be a more microscopic view of discourse structure analysis. For instance, a speaker sometimes uses a number of structured UUs to convince his interlocuter to do some particular actions. In such cases, the first UU may presummarize the speaker's proposal, the second UU may talk about his main plan, and the last UU may show the alternative plans. The prosodic information can be also used as a cue for this sort of structure; called *argumentative structure* [Cohen 87] or *coherent structure* [Hobbs 79], and [NakAll 91] discusses this issue with showing some typical examples.

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

参考文献

[AllPer 80] Allen, J.F. & Perrault, C.R. *Analyzing intention in utterances.* Artificial Intelligence 15, 1980.

[AllSch 91] Allen, J.F. & Schubert, L.K. *The* TRAINS *project,* TRAINS Technical Note 91-1, Computer Science Dept, University of Rochester, 1991.

[Austin 62] Austin, J.L. *How to do things with words.* Oxford University Press, 1962.

[BroYul 83] Brown, G. & Yule, G. *Discourse analysis.* Cambridge University Press, 1983.

[BroCur 80] Brown, G., Currie, K.L. & Kenworthy, J. *Questions of intonation.* Croom Helm, 1980.

[Cohen 87] Cohen, Robin. *Analyzing the structure of argumentative discourse.* Computational Linguistics 13, 1987.

[Guss 83] Gussenhoven, C. *On the grammar and semantics of sentence accents.* Language Sciences 16, 1983.

[HakSat 80] Hakoda, K. & Sato, H. *Prosodic rules in connected speech synthesis.* Trans. of the Institute of Electronics and Communication Engineers 63-D, 1980.

[Hobbs 79] Hobbs, J. *Coherence and coreference.* Cognitive Science, 3(1), 1979.

[Ladd 84] Ladd, D.R. *Declination: a review and some hypotheses.* Phonology Yearbook I, 1984.

[LiePie 84] Lieberman, M. & Pierrehumbert, J.B. *Intonational invariance under changes in pitch range and length,* in M. Aronoff and R.T. Oehrle (eds.) *Language sound structure.* MIT Press, 1984.

[NakAll 91] Nakajima, S. & Allen, J.F. *A study of pragmatic roles of prosody in the* TRAINS *dialogs.* TRAINS technical note, Computer Science Dept, University of Rochester, forthcoming.

[PieHir 90] Pierrehumbert, J. & Hirschberg, J. *The meaning of intonational contours in the interpretation of discourse,* in P.R. Cohen, J. Morgan, & M.E. Pollack (eds.) *Intentions in communication.* MIT Press, 1990.

[Searle 69] Searle, J.R. *Speech Acts.* Cambridge University Press, 1969.