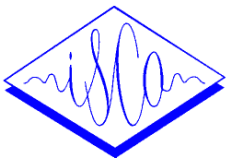

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

Prosody as a Cue for Discourse Structure

Shin'ya Nakajima

NTT Human Interface Laboratories
Take 1-2356, Yokosuka, Kanagawa 238-03, JAPAN

James Allen

The University of Rochester
Rochester, NY 14627, USA

Abstract

This paper describes how well prosodic information correlates with the topic structure of discourse. To investigate this correlation systematically, first we introduce the notion of *utterance unit* which can be viewed as the 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, and final pitch frequencies– are measured at these topic boundaries to show how these prosodic parameters vary with the topic structure. Finally, we propose a schematic algorithm which identifies the topic boundaries via the prosodic parameters.

10.21437/ICSLP.1992-102

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 [3, 14] have been proposed (e.g. [1]). However, although a number of analysts have pointed out that prosody plays several important roles in natural conversations (e.g. [4, 13]), there have been very few studies that analyse spontaneous conversational speech.

Among the various roles of prosody, this paper focuses on the topic structure specification function and shows how prosodic information can be utilized as a cue for topic structure identification.

2 Speech Data Collection

For speech data collection, we use a specific task domain– the TRAINS world[2]. The cities in the TRAINS world are connected to each other by rail lines. Each city has 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, respectively. We collected a total dialogue duration of about one and half hours from six goal-achieving sessions.

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. Note that this does not rule out the case where one speech act continues over several UUs.
- **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 applying these rules to the speech data, the utterances were split into numbered UUs.

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. (Examples of these classes are shown in [12])

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.

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.

Interruption (Int) The previous or simultaneous utterance is interrupted abruptly by the current utterance.

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.

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.

Clarification (Clr) The current utterance clarifies some propositions involved in the previous utterances.

Summary (Summ) The current utterance summarizes the contents of the preceding utterances.

Speech Act Continuation (AC) A single speech act continues over several UUs. Most of them are sequential conjunctive utterances.

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.

4 Prosody and Discourse Structure

4.1 Onset Pitch Frequency

A number of analysts have suggested that onset and first peak pitches are raised when the topic of the conversation is changed. (e.g. [5]) 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 pitch frequency 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. Since we are focusing here on the relationship between topic-shifting and onset pitch, we also excluded simple answer utterances.

Onset pitch average (hereafter Po) at each topic boundary class is shown in Fig.1. The results can be summarized as follows;

- For each speaker, Po declines in the order;

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

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

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

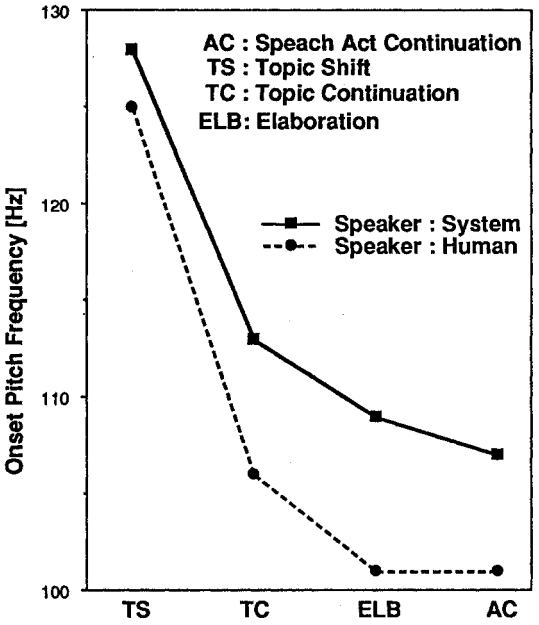

Figure 1: Onset pitch frequency at each topic boundary

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 final pitch of single answers, not followed by any subsequent utterances, are counted together with those of TS boundaries and referred as END class. This is because there is no significant distinction between the isolated answers and the topic shift boundaries. We excluded questioning utterances in which final pitch contours may not signal topic-shifting phenomena.

The average of final pitch frequency at each topic boundary is shown in Fig.2. 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 pitch is 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 pitch value 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.

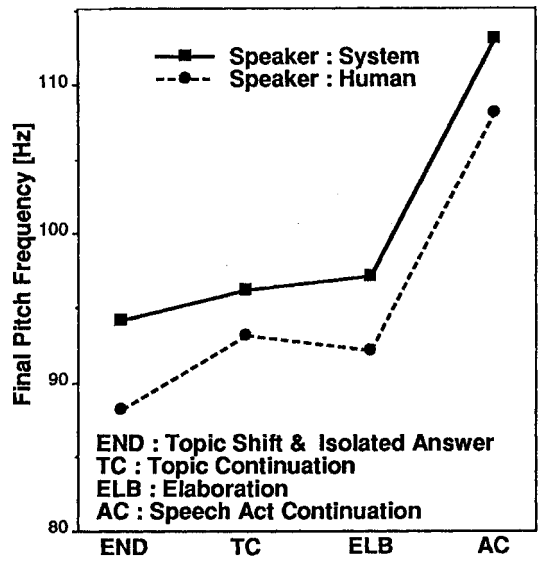

Figure 2: Final pitch frequency at each topic boundary

4.3 Peak Pitch Ratio

It is claimed that within continuous speech, the peak pitch range of each intonational phrase declines towards the end of sentences [8, 11, 10]. [8] 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.

To investigate the degree of suppression, we use the ratio of the current UU's first peak pitch to that of the previous one. The first peak pitch frequency of the current UU (Pp1) and that of the same speaker's previous UU (Pp0) are measured. The suppression ratio of first peak pitch (Rpp) is then computed as follows. (Hereafter, we call this parameter *the peak pitch ratio.*)

$$\mathrm{Rpp} = \frac{\mathrm{Pp}_1}{\mathrm{Pp}_0}$$

The averages of Rpp are shown in Fig.3. The results can be summarized as follows;

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

  TS > TC > AC > ELB

- The peak pitch ratio is larger than 1.0 at TS boundaries, and is around 1.0 at TC boundaries. This suggests that if the topic changes, the speaker starts speaking with a higher peak pitch range and 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, Rpp at ELB boundaries is lower than that at AC boundaries. This can be interpreted as follows; the relationship between two utterances at an AC boundary is mostly coordinate, whereas elaboration utterances are often subordinate to the previous ones. This subordination suppresses elaboration utterances more than cordination utterances.

- As can be inferred from Fig.3, the peak pitch ratio is a reliable parameter with which to discriminates ELB boundaries from TC boundaries.

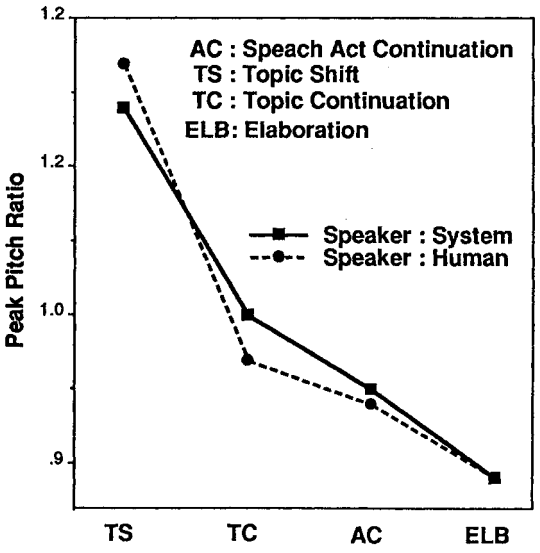

Figure 3: Peak pitch ratio at each topic boundary

4.4 Topic Boundary Identification

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. Although onset pitch can be also used for this discrimination, as can be inferred from Fig.3 comparing with Fig.1, peak pitch is more reliable parameter. (this is also confirmed via T-distribution test.)

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

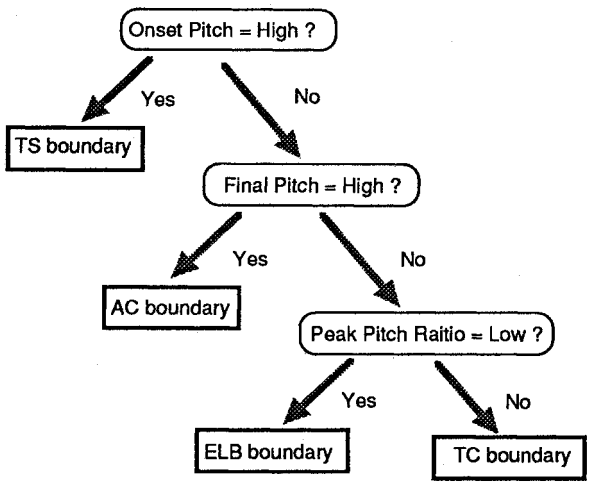

Figure 4: Topic boundary discrimination tree

5 Discussion

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 be 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 summarization.
- 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 are left as future tasks.

This paper has been focused on the correlation between prosodic information and topic boundaries. However, there might be a more microscopic approach to discourse structure analysis. For instance, a speaker sometimes uses a number of structured UUs to convince his interlocutor to do some particular action. In such cases, the first UU may presummarize the speaker's proposal, the second UU may introduce his main plan, and the last UU may show alternative plans. The prosodic information can be also used as a cue for this sort of structure. The analysis of this sort is called *argumentative structure* in [6] and *coherent structure* in [9], and [12] discusses this issue by showing some typical examples.

6 Conclusion

This paper describes how well prosodic parameters correlate with the topic structure of discourse. In order to show this correlation we have introduced the notion of *Utterance Unit* and defined four topic boundary classes. The prosodic parameters such as *onset pitch*, *final pitch*, and *peak pitch ratio* are measured at each topic boundary, and it has been shown that these parameters can be used for discriminating topic boundary classes.

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, 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. & Yule, G. *Discourse analysis.* Cambridge University Press, 1983.

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

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

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

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

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

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

[11] 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.

[12] 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.

[13] 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.

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