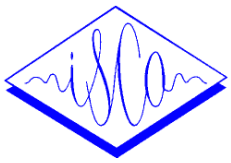
ISCA Archive
http://www.isca-speech.org/archive



3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

# Heuristics for Generating Acoustic Stress in Dialogues and Examination of their Validity

**Shozo NAITO and Akira SHIMAZU**
NTT Software Laboratories and Basic Research Laboratories
email: naito@atom.ntt.jp and shimazu@atom.ntt.jp
3-9-11 Midori-cho, Musashino-shi, Tokyo 180, Japan

## Abstract

This paper examines the validity of five types of heuristics for generating acoustic stresses which had been proposed based on the analysis of cooperative dialogues. The five types of heuristics are, 1) noun that first appears in dialogues, 2) verb, adjective, or adjectival verb that first appears in dialogues, 3) new topics in dialogues, 4) sentence final particle of interrogative or rhetorical interrogative sentences, and 5) contrasting. These heuristics have been incorporated into an experimental natural language generation system.

## 1 Introduction

A speaker, in general, intentionally and locally modifies prosodical features, such as pitch, power, speed, and pause duration, and thereby makes important portions of utterances more prominent than others. These prosodical features are affected by factors concerning speaking ranging from discourse to the physical level[2][7] [8]. For example, Hirschberg et al. analyzed the intonation patterns of 100 occurrences of the English word 'now' in dialogues and tried to classify its functions into two basic categories of use: the indication of time; and the indication of a change in topic[3].

Knowledge of the communication functions of phonemes has deepened with advances in research on natural language such as speech act theory[11][13]. However, when compared with the advancement of research on phonemes, many questions about the functions of prosody such as acoustic stresses concerning discourse, and the mechanisms of prosody generation/understanding in spoken language remain unanswered[1], though prosodical phenomena such as catathesis[10] and metrical boost[5] concerning syntax are reported.

With this premise, our purpose is to analyze functions of acoustic stresses and compile basic data for the construction of an utterance generation and understanding model that can be processed by a computer. Takeda et al. analyzed prominence rules of spoken Japanese and suggested methods for their production[12]. In their research, they assumed that prominent portions in utterances were predetermined, but they did not investigate heuristics for planning prominence in the process of generating utterances.

This paper examines the validity of heuristics for generating acoustic stresses in Japanese utterance generation in cooperative dialogues. The validity of these heuristics is examined against the dialogue data and these heuristics are incorporated into a natural language generation system[9]. In the following, Section 2 explains five types of heuristics to be examined. Section 3 describes the method of examination and its results.

## 2 Heuristics for Prominence Generation

We have examined five types of heuristics shown in Table 1, which have been proposed through the analysis of fifteen dialogues listed in Table 2.

Table 1: Heuristics for Generating Prominences

| Label | Heuristicss |
|---|---|
| h1 | Noun that first appears in dialogue |
| h2 | Verb, adjective, or adjectival that first appears in dialogues |
| h3 | New topics in dialogue |
| h4 | Sentence final particle of interrogative or rhetorical interrogative sentences |
| h5 | Contrasting |

### (1) Noun that first appears in dialogue

In general, by introducing a noun as well as verb, adjective, or adjectival verb into a dialogue, individuals, sets, or events (actions or situations) are introduced into the dialogue for the first time. The most basic and important process in a dialogue is for its participants to accurately register these new elements when they appear. The simplest and most effective way of ensuring that this happens is to use acoustic stresses in dialogues. In a telephone dialogue, confirming the identity of the person on the other end of the line is the most basic process in the dialogue. How this process unfolds through the use of prominence is demonstrated in the following dialogue in which one speaker (*Sakai*) attempts to confirm an identity and the other speaker (*Aizawa*) introduces himself[1].

A: *A moshimoshi*
Hello.
B: *Hai.*
Yes.
A: **Sakai***san.*
Sakai-san?
B: *A Sakai de gozaimasu.*

[1]The boldface indicates prominent words under discussion.

Table 2: Data for Analysis

| No. | Theme | Duration (sec.) | Turns | Length of transcript (bytes) |
|---|---|---|---|---|
| 1 | Changing the time of a meeting | 98 | 62 | 2835 |
| 2 | Interviewing an applicant as a secretary | 188 | 89 | 4720 |
| 3 | Scheduling a game of mah-jongg | 154 | 66 | 4048 |
| 4 | Selling English conversation tapes | 220 | 106 | 5756 |
| 5 | Asking for directions to a flower shop | 420 | 251 | 10929 |
| 6 | Giving drawing instructions (cat) | 244 | 70 | 4979 |
| 7 | Giving drawing instructions (crow) | 245 | 65 | 4822 |
| 8 | Applying for an audition (female) | 207 | 135 | 5548 |
| 9 | Buying a ticket | 189 | 112 | 5646 |
| 10 | Applying for an audition (male) | 271 | 84 | 5865 |
| 11 | Talking with a child (at school) | 184 | 78 | 3511 |
| 12 | Talking with a child (in a park) | 82 | 27 | 1480 |
| 13 | Speculating about a fortune | 216 | 125 | 6048 |
| 14 | Talking about roles and weights | 244 | 169 | 7472 |
| 15 | Talking about how to avoid colds | 245 | 149 | 7171 |
| Total | | 3207 | 1588 | 80830 |
| Average | | 214 | 106 | 5389 |

Yes, Sakai speaking.
A: *A* **Aizawa** *desu ga ne.*
Eh, this is Aizawa speaking.

**(2) Verb, adjective, or adjectival verb that first appears in dialogue**

The function of introduction of events (actions or situations) into the dialogue is already described above. Furthermore, in spoken dialogues, subjective information held by the speaker, which is classified as a modality, is often transmitted to the hearer. Subjective information often has an emotional dimension and is closely linked with the appropriate form of prosody. When there is discordance between the subjective information and the prosody, the hearer must infer the intentions of the speaker and prosody becomes an ineffective tool for transmitting of meaning. In the following dialogue, prominence is placed on the word used to express an apology (*sumanai*).

A: *Aa sorekara.*
And...
B: *Hai.*
Yes.
A: *Watashi no hoono aa youken de nobashita koto wa kureguremo* **sumanai** *to.*
I am sorry that our meeting will have to be delayed on my account.
B: *Hai.*
Yes.
A: *Soredake wa tsukekuwaete oite kuretamae.*
Please convey my apologies to him for me.

**(3) New topics in dialogue**

After introducing discourse elements, a topic is chosen from these elements and participants in the dialogue then exchange information about the topic. Therefore, it is important for both participants to understand which discourse element has been chosen as the topic. In the following dialogue, *eigo* (English) is chosen as a topic.

A: *Haha naruhodo ne.*
Ah, I see.
B: *Sorekara eeto* **eigo** *wa shooshoo.*
And some English.
A: *Un*
Um?
B: *Hai.*
Yes?
A: *Eeto shooshoo to iu to dono teido ka na.*
Well, what do you mean by some?

Prominence is a simple and effective way of indicating a new topic to the hearer. The identification of a topic are basic issues in the use of language[6]. In this paper, a new topic is assumed to be identified by the procedure described below. A phrase that satisfies one of the following conditions is considered to be the topic phrase for each point in time.

1. A phrase with *wa*, which is a particle; or a phrase without a particle, but in which *wa* is judged to have been omitted.
2. A phrase referred to by a pronoun (including zero pronouns).

A new topic is acknowledged as having been introduced when the topic changes through one of the above phrases. In the above example, speaker B mentions ability in English as one of her skills. Speaker A then asks speaker B about the extent of her English ability. Speaker B introduces the topic *eigo* (English) by *wa*. At this point, condition 1 above is satisfied, so the point at which speaker B introduces the word *eigo* into the conversation is acknowledged to be the point at which he introduces the new topic of *eigo*. Speaker A then retains this topic by a zero pronoun referring to it.

**(4) The sentence final particle of interrogative or rhetorical interrogative sentences**

At the end of an interrogative sentence, the pitch usually rises (intonation) and at the end of a declarative sentence, the pitch usually falls. In the following dialogue, in addition to pitch, power is also added to place prominence on the word *ka* at the end of the utterance.

A: *A moshimoshi.*
Hello.
B: *Hai.*
Yes.
A: *Bucho iru* **ka** *na.*
Is the manager in?
B: *A hai Sakai de gozaimasu.*
Eh, this is Sakai speaking.

**(5) Contrasting**

When a sentence contains contrasting parts, the particle *wa* is used in the part the speaker intends to stress. In the following example, speaker A confirms that the time of the meeting has been changed, but that the place of the meeting has not. To contrast these two elements, speaker A places prominence on *basho* (the meeting place), which is one of the contrasting elements.

A: *Ee goji kara ichijikan areba sumu to.*
Well, one hour should be enough,
starting from five o'clock.
B: *Hai acho.*
Oh, yes.
A: *Desukara* **basho** *wa kawarazu.*
So, the place is the same.
B: *Basho wa kawarazu.*
The same place.
A: *Un.*
Yes.

# 3 Validity Examination of Heuristics

## 3.1 Method of Examination

The validity of the heuristics is examined by the following two-step procedures: marking of words in dialogue transcripts which are predicted to be prominent by the heuristics, and confirmation of prominence by reference to the power graphs shown in Figure 1. New topics are marked by the procedures described in Section 2. We determined whether the marked parts in the transcripts were actually prominent by reference to power graphs and assigned to each predicted word a corresponding prominence level selected from those shown in Table 3. Level 1 corresponds to the case that some portion of the predicted part is included in the range of prominence (that is, between the maximum power and 6 db lower), level 2 covers the range to 3 db lower than the prominence level, and level 3 covers other cases. The prominence range was determined to be 6 db in the examination for two reasons: (1) The step size of the data for examination (examplified in Figure 1) is 3 db and phonemes in the dialogues rarely fall in this step power range, (2) Measurements of natural dialogues have shown that power generally fluctuates between 20 db and 30 db[4] and preliminary experiments for recorded dialogues showed that prominences are well discriminated from non-prominences with this power fluctuation range.

Table 3: Check of Prominent Level

| Label | Level |
|---|---|
| 1 | Some portion of the predicted word is included in the range of prominence (6db lower than maximum power) |
| 2 | 3 db lower than level 1 |
| 3 | Otherwise |

Table 4: Validity Examination Result

| Heuristics | Prominent/Non-prominent Levels | | | |
|---|---|---|---|---|
| Label | 1 | 2 | 3 | Total |
| h1 | 218(58%) | 84(23%) | 71(19%) | 373 |
| h2 | 55(39%) | 41(29%) | 46(32%) | 142 |
| h3 | 29(76%) | 8(21%) | 1(3%) | 38 |
| h4 | 8(20%) | 9(22%) | 24(58%) | 41 |
| h5 | 9(60%) | 2(13%) | 4(27%) | 15 |
| | 319(52%) | 144(24%) | 146(24%) | 609 |

## 3.2 Data for Examination

The people engaged in the dialogues were assigned roles such as that of a person providing information or a person asking for information. The conversations were continued until their intended goals were achieved. The dialogues were recorded with a two-channel tape recorder, so that each speaker could be recorded individually, and transcribed (Table 2). In the present analysis, power parameters were used to identify prominences in the dialogues, but pitch and speed fluctuations were not taken into consideration because they are strongly influenced by lexical information such as accent. Figure 1 shows an example of the data for analysis; the figure has phonemic labels which correspond to the dialogues. The data is for the power parameter of the dialogues and it is arranged in a time series, so it is henceforth called a power graph. The average duration of the fifteen dialogues was 214 seconds, the average number of utterance turns was 106, and the average number of bytes of the transcripts was 5,389.

## 3.3 Examination Results

The result of the validity examination is shown in Table 4. We obtained the following results.

1. When nouns or verbs occur in the dialogue for the first time, nouns are more frequently stressed compared with verbs. It has been suggested that the reason for this is that nouns are in general introduced as topics in dialogues and information about these topic nouns are communicated betweeen dialogue participants.
2. When nouns are introduced into dialogues as new topics, 76% of the total occurrences are prominent. When the threshold of prominence is low-

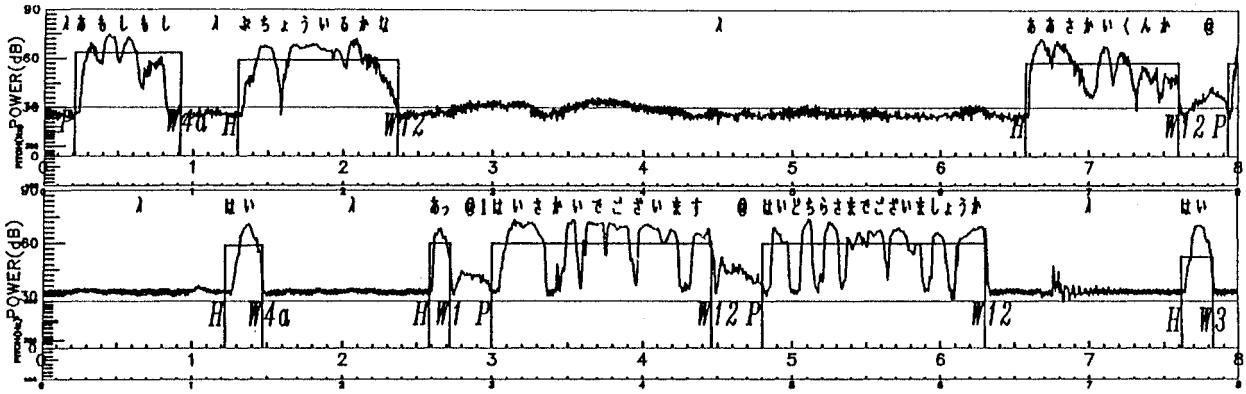


Figure 1: Example of power graph

ered by 3db, 97% of the total occurrences become prominent.

3. The interrogative sentence-final particle *ka* is not necessarily prominent according to the power parameter. When the threshold is expanded by 3db, 40% of the total occurrences fall in the prominence range. The interrogative sentence-final particle *ka* is in general achieved by intonation, so we assume that power becomes redundant. There are no occurrences of the rhetorical interrogative utterances in the analyzed data.
4. The number of contrasting occurrences is small, and 60% of the total occurrences are prominent. In Japanese contrasting expressions, prominence may fall on two contrasting nouns or the modal particle *wa* following the contrasting nouns.
5. Reasons for cases where the heuristics fail can be considered. For example, in the following dialogue, although *yooken* (matter) is introduced into discourse for the first time, it is not prominent. In cooperative dialogues, it is in general requested that the reason will be given when a shared plan is interrupted[?]. In the utterance stating a reason for asking the hearer to change the beginning time of the meeting in the following dialogue, the speaker places prominence on *hazusenai* (urgent) modifying the *yooken* instead of placing it on *yooken*, which is abstract and does not convey important information.

A: *Ee sanji kara chotto uchiawase ga attan daga nee*
Well, we will have a meeting from three o'clock.
B: *Hai*
Yes.
A: *Dooshitemo watashi hazusenai* **yooken** *ga attene*
I have an uegent matter.

## 4 Concluding Remarks

In this paper we examined the validity of the proposed heuristics for planning acoustic stresses in utterance generation. These heuristics have been incorporated into an experimental utterance generation system[9]. In the future, detailed analyses of acoustic stresses using other parameters such as pitch and speed should be conducted.

## Acknowledgments

We would like to thank Naotoshi Osaka for offering dialogue data. We are also grateful to Shinobu Takechi for analyzing dialogue data.

## References

[1] Bolinger, D.: Accent is predictable (if you're a mindreader), *Language*, 48, pp. 633-644, 1972.
[2] Grosz, B. and Sidner, C.: Attention, intention and the structure of discourse, *Computational Linguistics*, 12(3), pp. 175–204, 1985.
[3] Hirschberg, J. and D. Litman: Now let's talk about 'now': Identifying cue phrases intonationally. *Proc. of 25th ACL*, pp.163-171, 1987.
[4] Koike, T. et al: Onsei jouhou kougaku (Speech information technology), (in Japanese) NTEC, 1987.
[5] Kubozono, H.: The Organization of Japanese Prosody, University of Edinburgh Ph.D Thesis, 1987.
[6] Kuno S.: Danwa no bunpou (Discourse grammar), (in Japanese) Taishuu-kan, 1978.
[7] Levelt, Willem, J. M.: Speaking, MIT Press, 1989.
[8] Maynard, S. K.: Kaiwa bunseki (Conversational analysis), (in Japanese) Kuroshio-shuppan, 1993.
[9] Naito, S. and A. Shimazu: Planning Utterances with Prominence, Proc. of PACLING, pp.242-250, 1993.
[10] Poser, W.: The Phonetics and Phonology of Tone and Intonation in Japanese, MIT Ph.D. Thesis, 1984.
[11] Searle, J.: Speech Acts: An Essay in the Philosophy of Languages, Cambridge University Press, 1969.
[12] Takeda, S. and Ichikawa, A.: Production and evaluation of prominence rules for spoken Japanese sentences (in Japanese), *The Journal of the Acoustic Society of Japan Vol. 47, No. 6*, 1991,6, pp. 397-404.
[13] Thomason, R. H.: Accomodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics Cohen, Morgan, and Pollack eds. *Intentions in COMMUNICATION*, pp.325-363, The MIT Press, 1990.