A SCHEMA FOR ILLOCUTIONARY ACT IDENTIFICATION WITH PROSODIC FEATURE

Masafumi TAMOTO and Takeshi KAWABATA

NTT Basic Research Laboratories
3-1 Morinosato Wakamiya, Atsugi-city
Kanagawa, 243-0198, Japan
URL: http://www.brl.ntt.co.jp/info/dug/
mailto: tamoto@idea.brl.ntt.co.jp

ABSTRACT

We propose a new discrimination schema for illocutionary acts using prosodic features based on experimental results.We performed a series of experiments in which subjects were asked to identify the sentence type and intonation contour of given stimuli. Given the transcribed sentence with contextual information, the subjects were able to identify correctly the sentence type of 85% of 290 sentences. With information about the intonation contour types, they could correctly identify 90% of speech acts. We find evidence that illocutionary acts can be signaled by specific contour types. These typical contours are realized in the sentence final boundary tone; a neutral or falling tone for assertion and request, a rising tone for question. An intonation contour is then identified using an algorithm that calculates the range and slope of the upper and lower bounds of unwarped segmental contour, and matches these against predefined contour templates. This algorithm could correctly recognize 78% of the pitch contour types in the utterances. Furthermore, this automated intonation contour classification, nearly 90% of speech acts could be correctly identified.

1. INTRODUCTION

Speech conversation is our usual communication method and is most comfortable man-machine interface. For constructing an effortless speech conversation system, it is necessary to implement the coordination mechanism for dialogs.

Prosodic information contributes to identifying speech acts of utterance when linguistic information is missing due to omission or obscure utterance. We aim to construct a system that successfully incorporates prosody, and to analyze the dialog coordination in human-machine conversation. The most commonly known contribution of prosody to speech communication is at the sentential pragmatic and intentional level. That is, analyzing syntactic and intonational properties of utterances, and relationship among sentence type, pitch contour and illocutionary act, intonation can be effectively used for disambiguation in mapping an utterance to these three types of illocutionary acts.

Several simplifications are introduced in the course of our experiment, speech acts are represented as three basic categories, the illocutions of assertion, question and request. Similarly, sentence types are represented as declarative, interrogative and imperative. Intonations are classified into rise-up, fall-down and neutral pitch contour. For investigating how prosodic information incorporates with linguistic information to identify speech acts, we performed a series of experiments.

1) We make sure of the contribution of prosodic information to identify speech acts of utterance through an experiment that evaluates precision/recall of speech act identification by human subjects using sentence type, pitch contour or both information. On the assumption that three sentence types of declarative, interrogative and imperative express the illocutions of assertion, question and requesting, respectively, then a successful mapping of illocutions might be expected to predict the predominance of these three sentence types. Contrary to the expectation, speech act identification can be improved by applying prosodic information in a manner of specific combination of sentence types and pitch contour types.

2) We propose a new algorithm that classify pitch contour to three intonation types for the purpose of automated speech act identification. These intonations are largely realized in the utterance final boundary tone. An intonation contour is identified using an algorithm that calculates the range and slope of the upper and lower bounds of unwarped segmental contour of the last one mora of the utterance, and matches these against predefined contour templates. Problems, however, remain in the course of this process, such as reliable extraction of an intonation contour processing.

3) With this automated intonation contour classification, the same examination as the first experiment that evaluates precision/recall of speech act identification based on the automated pitch contour classification and sentence identification by human subjects is performed.

2. PROBLEM SETTING

We consider an illocutionary act set, containing the illocution of assertion, question and request; and combination of a sentence type set, consisting three basic sentence types of declarative, interrogative and imperative, and a pitch contour type set, consisting of neutral, rising and falling. To find the efficient combinations that identify illocutionary act from sentence type a pitch contour type, we used a measurement of precision/recall.

Likewise, we find the feature that efficiently classify pitch contours into a pitch contour type of neutral, rising and falling. Our research addresses this problem and uses the schema to identify illocutionary act of the given utterance.

3. ILLOCUTIONARY ACT IDENTIFICATION BY HUMAN SUBJECTS

In this section, we perform an experiment to investigate the scheme of illocutionary act identification. This is an illocutionary act identification experiment for investigating how the illocutionary acts are appropriately classified by a human applying syntactic and intonational properties.

3.1. Speech data

The speech data for the experiment was produced in the following way. Two participants perform two coordinative tasks. One is the "map task". One participant informs a given route on map to the other, exchanging information of their map and route. The other task is "maze task". Each of the two participants has one half of a maze divided into two halves. They have to find a passage of the maze through exchanging information of their piece of maze with each other. These dialogs are 30 minutes length totally, sampled at 12kHz and transcribed and divided into chunks. All chunks had to form complete meaningful utterances.

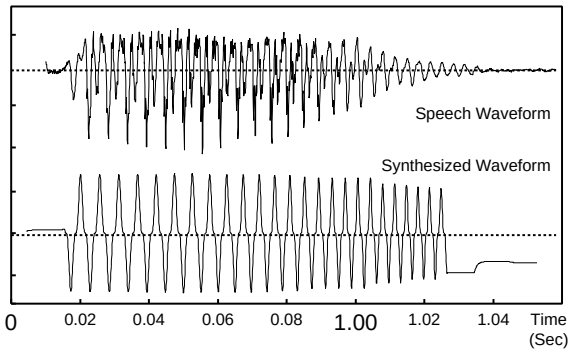

Figure 1. Pitch contour synthesis

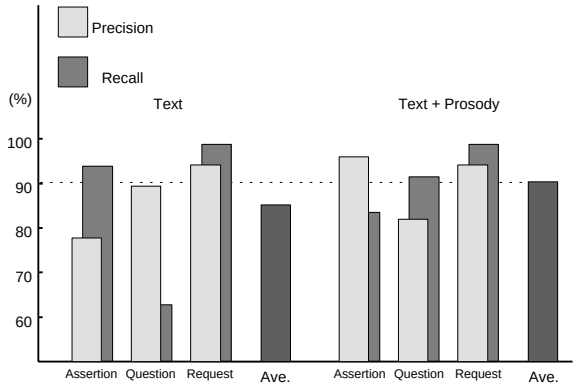

Figure 2. Estimation efficiency of speech act by sentence type

3.2. Method and Procedure

First, sentence type identification is performed by human subjects on the basis of transcribed chunks of text. Then, hearing F_0 contours, subjects classify these boundary tones into neutral, rising and falling pitch contour types. The tone signal that denotes F_0 contour is synthesized by modulating sinusoid according to F_0 frequency and square power of utterance. Finally, the transcriber determines illocutionary act of each chunk listening to utterances in sequence. This means the transcriber is permitted to use backward contextual information. Assuming three basic sentence types – interrogative, imperative, and declarative express the illocutionary acts of questioning, requesting and asserting, respectively, identification error rate is 15%. These classification errors are caused by lost particles that denote interrogative moods.

Assuming three basic pitch contour types – rise-up, neutral and fall-down – express these illocutionary acts, identification error rate is 50%.

As the result of classification errors, the combinations that efficiently identify illocutionary act from sentence type and pitch contour type is :

assertion declarative except for rise-up boundary tone

question interrogative or declarative with rise-up boundary tone

request imperative

3.3. Results and Discussion

Figure 2 shows precision/recall rates of the illocutionary acts identification obtained for each method of classification. Using this classification rules, the error rate drops to 10%. This figure also shows the recall rate of identifying the illocution of question is improved remarkably. This is considered due to the joining of declarative sentence with rise-up boundary tone to the illocution of question.

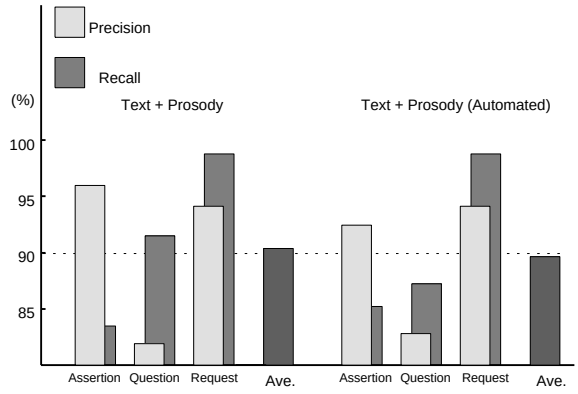

Figure 3. Estimation efficiency of Speech act by sentence type and automated intonation type discrimination

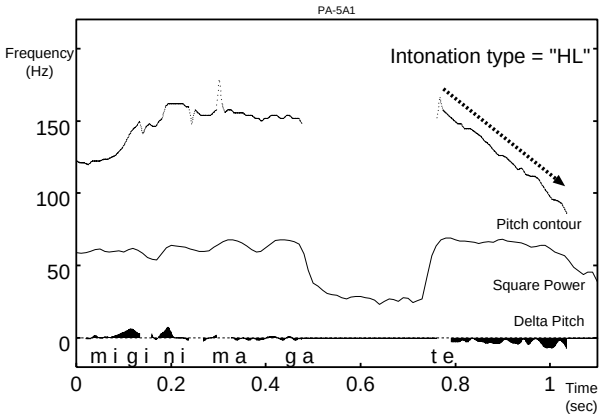

Figure 4. Falling intonation

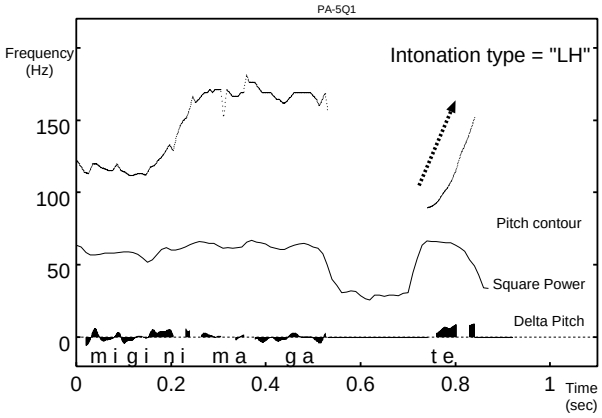

Figure 5. Rising intonation

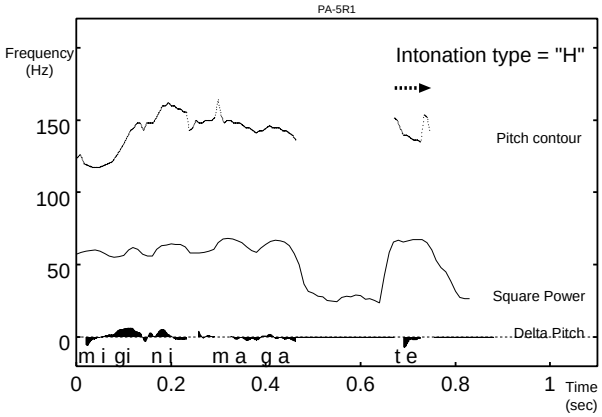

Figure 6. Neutral intonation

4. ILLOCUTIONARY ACT IDENTIFICATION WITH AUTOMATED PITCH CONTOUR CLASSIFICATION

In this section, we report an experiment for evaluating an automated pitch contour classification method and the illocutionary act identification proposed above.

4.1. Method and Procedure

Based on the pitch contour classification criteria performed by human subjects, the automated pitch contour classification criteria was produced following way.

Pitch contour extraction

- Using pitch tracking software that is based loosely on an algorithm by Medan, Yair and Chazan and also utilizes dynamic programming, rough pitch contour is extracted.
- Post-processing to determine voiced/unvoiced decision and to omit glitches including half-tone and overtone. The method to omit glitches is as follows: when square power of input speech becomes less than the threshold, output pitch frequency is regarded as unreliable; when the pitch frequency changes rapidly more than the specified rate, pitch frequency is recalculated as less than over-tone and higher than half-tone of recent frequency.

Pitch contour approximation

- Dividing unwarped pitch contour into monotonous segments from utterance end. Deviations that cannot affect to pitch contour determination are ignored. Average pitch slopes are calculated.
- Repeat the above procedure until vowel boundary appears.

Figure 4,5 and 6 show how pitch contour is extracted by proposed procedure, and the classification to pitch contour types. All these figure show utterances of the same phrase by same speaker in different illocutions. These phrases mean “Turn right”.

4.2. Results and Discussion

Figure 3 shows precision/recall rates of the illocutionary acts identification obtained for each methods of classification with automated pitch contour classification. The pitch deviation threshold between rise-up and neutral is $0.38 octave/sec$, and threshold between fall-down and neutral is $-1.43 octave/sec$, to fit the classification that human subjects performed. Using automated pitch contour classification, the error rate drops slightly, less than 1%, against pitch contour classification by human subjects.

This figure shows the recall rates of identifying the illocution of question decrease due to classification errors between neutral and fall-down pitch contour. This is considered due to the fact that the slow but long fall-down pitch contour is classified as neutral pitch contour, whereas fall-down pitch contour with sustained power is recognized as neutral pitch contour by human subjects.

5. SUMMARY AND FUTURE WORK

We have proposed a new discrimination schema for illocutionary acts using prosodic features based on experimental results. An intonation contour is identified using an algorithm that calculates the range and slope of the upper and lower bounds of unwarped segmental contour, and matches these against predefined contour templates. This algorithm could correctly recognize 78% of the pitch contour types in the utterances. Furthermore, by this automated intonation contour classification, nearly 90% of speech acts could be correctly identified.

There are a number of problems that need to be solved in the future. First, in order to reduce the number of intonation contour identification errors due to inappropriate analysis, a new method is expected with an algorithm that explicitly uses pitch contours in a parametric form (e.g., by integration into distance scores) for more precise identification, and that calculates the cross correlation factor between two neighboring frames so that the probability density of fundamental frequency may be derived.

In order to implement this algorithm into a speech understanding system, we have to make a module based on the result of these experiments that determines the type of a sentence by analyzing its syntactic and intonational properties. A sentence type is identified using a parser or template matching.

6. ACKNOWLEDGEMENT

Acknowledgement is made to Dr. Ken'ichiro Ishii, the Executive Manager of the Information Science Research Laboratory of NTT Basic Research Laboratories, for his support. The authors also acknowledge their debt to the other members of the Dialog Understanding research group.

7. REFERENCES

[1] A. Black. Predicting the intonation of discourse segments from examples in dialogue speech. In Y. Sagisaka, N. Campbell, and N. Higuchi, editors, *Computing Prosody*, pages 117–128. Springer-Verlag, 1997.

[2] B. Grosz, J. Hirschberg, and C. Nakatani. Some intonational characteristics of discourse structure. In *Proceedings of ICSLP*, 1992.

[3] N. Kaiki and Y. Sagisaka. Pause characteristics and local phrase-dependency structure in Japanese. In *Proceedings of ICSLP*, pages 357–360, 1992.

[4] Willem J. Levelt. From intention to articulation. In *Speaking*. The MIT Press, 1989.

[5] Stephen C. Levinson. *Pragmatics*. Press Syndicate of the University of Cambridge, 1983.

[6] Y. Medan and E. Yair. Pitch synchronous spectral analysis scheme for voiced speech. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-37(9):1321, 1989.

[7] N Minematsu and K Hirose. Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *THE JOURNAL of the Acoustical Society of Japan (E)*, pages 311–320, 1995.

[8] Shin'ya Nakajima and J. F. Allen. Prosody as a cue for discourse structure. In *Proceedings of ICSLP*, pages 425–428, 1992.

[9] J Pierrehumbert and J Hirschberg. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 271–312. The MIT Press, 1989.

[10] J. Pitrelli, E. Beckman, and J. Hirschberg. Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proceedings of ICSLP*, pages 123–126, 1995.

[11] L. R. Rabiner, M. J. Cheng, and A. E. Rosenberg. A comparative performance study of several pitch detection. In *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 24, pages 399–418. IEEE, 1976.

[12] Marc Swerts and Mari Ostendorf. Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication*, 22:25–41, 1997.

[13] Alex Waibel. *Prosody and Speech Recognition*. Morgan Kaufmann, 1988. .