

## Intonation Recognition for Indonesian Speech Based on Fujisaki Model

Nazrul Effendy<sup>1,2</sup>, Ekkarit Maneenoi<sup>1</sup>, Patavee Charnvivit<sup>1</sup>, Somchai Jitapunkul<sup>1</sup>

<sup>1</sup>Digital Signal Processing Research Laboratory, Department of Electrical Engineering  
Chulalongkorn University, Bangkok, Thailand

<sup>2</sup>Instrumentation Laboratory, Department of Engineering Physics  
Gadjah Mada University, Jogjakarta, Indonesia

Email: nazrul\_e@yahoo.com ekkarit@chula.com patavee@chula.com jsomchai@chula.ac.th

### Abstract

In this paper, we proposed to use the Fujisaki parameter to distinguish between declarative and interrogative intonation in Indonesian speech. Four combinations of Fujisaki parameter were selected as the features to distinguish between declarative and interrogative intonation. The first combination is only the amplitude of last accent command. The second combination consists of the amplitude of last accent command and the magnitude of last phrase command. The third combination consists of Fb, the amplitude of last accent command, and the magnitude of last phrase command. The fourth combination consists of Fb/100, the amplitude of last accent command, and the magnitude of last phrase command.

The recognition rates using the neural network were 83.33 %, 90.00 %, 50.00 %, and 96.67 % for each combination. The highest recognition rate was achieved by using Fb/100, the last accent command amplitude and the last phrase command amplitude as its inputs.

### 1. Introduction

Indonesian language is a non-tonal language. The meaning of an utterance in Indonesian speech does not depend on only the words used by the speaker, but also the intonation expressed by a speaker. Although the words used in the utterance are same, but if the speaker uses different intonation, the meaning of the utterance will be different. [1]

Instead of building detailed mathematical models of F0 by way of STEM-ML (Soft TEMplate Mark-up Language) [2], in this research we proposed to use the Fujisaki model of F0 and use the Fujisaki parameter to distinguish between declarative and interrogative intonation in Indonesian speech. We tried to use four different combinations of Fujisaki parameter as the features to distinguish between declarative and interrogative intonation. The first combination is only the amplitude of last accent command. The second combination consists of the amplitude of last accent command and the magnitude of last phrase command. The third combination consists of Fb, the amplitude of last accent command, and the magnitude of last phrase command. The fourth combination consists of Fb/100, the amplitude of last accent command, and the magnitude of last phrase command.

### 2. Fujisaki Model

From the Fujisaki model shown in Figure 1, the phrase commands are assumed to be impulses applied to the phrase control mechanism to generate the phrase components, while the accent commands are assumed to be positive stepwise functions applied to the accent control mechanism to generate the accent components. Both mechanisms are assumed to be critically damped second-order linear systems, and the sum of their outputs, i.e. the phrase components and the accent components, is superimposed on a baseline value (log Fb) to form an F0 contour, as given by following equation: [3]

$$\log F_0(t) = \log F_b + \sum_{i=1}^I A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^J A_{aj} \{G_a(t - T_{1j}) - G_a(t - T_{2j})\} \quad (1)$$

$$G_p(t) = \begin{cases} \alpha^2 t \cdot \exp(-\alpha t) & , \text{for } t \geq 0 \\ 0 & , \text{for } t < 0 \end{cases} \quad (2)$$

$$G_a(t) = \begin{cases} \min[1 - (1 + \beta t) \cdot \exp(-\beta t), \gamma] & , \text{for } t \geq 0 \\ 0 & , \text{for } t < 0 \end{cases} \quad (3)$$

Where,

- Fb : Baseline value of fundamental frequency
- I : Number of phrase commands
- J : Number of accent commands
- A<sub>pi</sub> : Magnitude of i<sup>th</sup> phrase command
- T<sub>0i</sub> : Timing of i<sup>th</sup> phrase command
- A<sub>aj</sub> : Amplitude of j<sup>th</sup> accent command
- T<sub>1j</sub> : Onset of j<sup>th</sup> accent command
- T<sub>2j</sub> : Offset of j<sup>th</sup> accent command
- α : Natural angular frequency of the phrase control mechanism
- β : Natural angular frequency of the accent control mechanism
- γ : Relative ceiling level of accent components

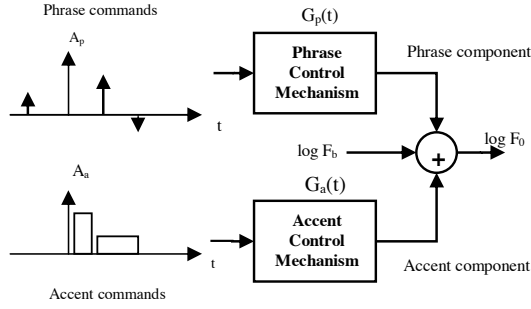


Figure 1: Fujisaki model

### 3. The Fujisaki Parameter Extraction

The F0 contour used in our research will be extracted by using PRAAT (©P.Boersma) [4], and then we got the Fujisaki parameter by using automatic Fujisaki parameter extractor proposed by Mixdorff and his colleagues [5]. Although, the automatic Fujisaki parameter extractor was specifically designed for German, we tried to adapt it for Indonesian language. Then, we chose some of the Fujisaki parameters to be used as the inputs of the neural network.

### 4. Experiment

We used speech data base consisting of 60 utterances from a native Indonesian male which consists of 30 declarative intonation utterances and 30 interrogative intonation utterances. The utterances were recorded at 16 kHz sample rate, and 16 bit resolution.

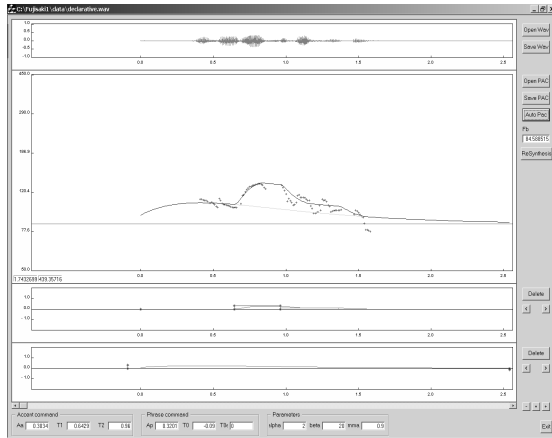


Figure 2: Example of automatic Fujisaki parameter extraction of Indonesian Declarative intonation utterance: "Saudara sudah makan?" (You already had lunch)

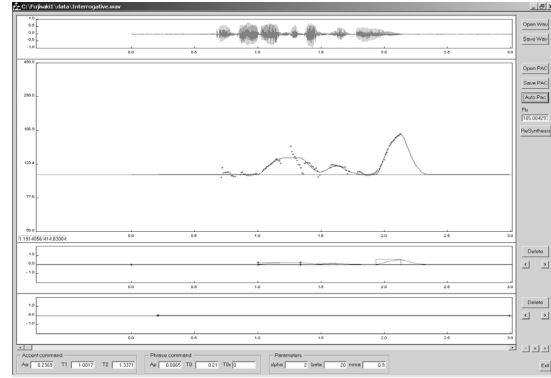


Figure 3: Example of automatic Fujisaki parameter extraction of Indonesian interrogative intonation utterance: "Saudara sudah makan?" (Have you had lunch?)

Table 1: The example of input data for neural network

| No | Fb     | A <sub>pl</sub> | A <sub>aj</sub> | Intonation    |
|----|--------|-----------------|-----------------|---------------|
| 1  | 107.07 | 0.2242          | 0.19            | declarative   |
| 2  | 132.83 | 0.4012          | 0.60            | interrogative |
| 3  | 85.69  | 0.0924          | 0.19            | declarative   |
| 4  | 119.13 | 0.1399          | 0.55            | interrogative |
| 5  | 120.02 | 0.3485          | 0.97            | declarative   |
| 6  | 117.70 | 0.3149          | 0.51            | interrogative |
| 7  | 86.47  | 0.4673          | 0.24            | declarative   |
| 8  | 108.37 | 0.1033          | 0.53            | interrogative |
| 9  | 97.81  | 0.3776          | 0.06            | declarative   |
| 10 | 131.11 | 0.1877          | 0.25            | interrogative |
| 11 | 89.55  | 0.3473          | 0.16            | declarative   |
| 12 | 132.96 | 0.0288          | 0.35            | interrogative |
| 13 | 106.02 | 0.4212          | 0.33            | declarative   |
| 14 | 103.98 | 0.4814          | 0.53            | interrogative |
| 15 | 121.72 | 0.0518          | 0.27            | declarative   |
| 16 | 116.54 | 0.043           | 0.53            | interrogative |
| 17 | 86.01  | 0.0025          | 0.47            | declarative   |
| 18 | 98.88  | 0.2707          | 0.76            | interrogative |
| 19 | 98.18  | 0.489           | 0.20            | declarative   |
| 20 | 106.67 | 0.2981          | 0.32            | interrogative |
| 21 | 89.46  | 0.42            | -0.24           | declarative   |
| 22 | 103.85 | 0.02            | 0.59            | interrogative |
| 23 | 97.47  | 0.22            | 0.17            | declarative   |
| 24 | 98.64  | 0.1882          | 0.46            | interrogative |
| 25 | 105.44 | 0.0008          | 0.14            | declarative   |
| 26 | 136.54 | 0.4257          | 0.45            | interrogative |
| 27 | 105.97 | 0.4241          | 0.21            | declarative   |
| 28 | 84.87  | 0.0328          | 0.87            | interrogative |
| 29 | 10.14  | 2.662           | 1.72            | declarative   |
| 30 | 122.54 | 0.293           | 0.57            | interrogative |

A<sub>aj</sub>: the amplitude of last accent command

A<sub>pl</sub>: the magnitude of last phrase command

Fb: Baseline value of fundamental frequency

Unlike [2], the utterances we used in this research did not contain the syllables with the same long. Figure 2 and 3

depicted examples of Indonesian declarative and interrogative intonation. These figures show that in Indonesian speech, the declarative intonation has the falling shape in the final part of the utterance, and the interrogative intonation has the rising shape in the final part of the utterance. Examples of Fujisaki parameters of declarative and interrogative intonation are given in Table 1. 90 % of the data used in this research has lower value of the amplitude of last accent command of declarative intonation than that of interrogative intonation, and 10 % of the data has higher value. 80 % of the data has higher value of the magnitude of the last phrase command of declarative intonation than that of interrogative intonation, and 20 % of the data has lower. 86.67 % of the data has lower value of Fb of declarative intonation than that of interrogative intonation, and 13.33 % of the data has higher value.

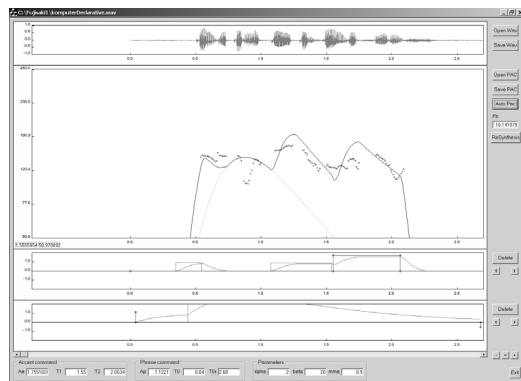


Figure 4: Example of automatic Fujisaki parameter extraction of Indonesian Declarative intonation utterance: "Komputer itu terjangkit virus" (That computer was infected by virus)

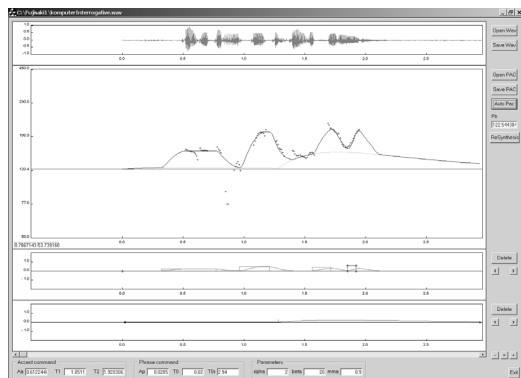


Figure 5: Example of automatic Fujisaki parameter extraction of Indonesian Interrogative intonation utterance: "Komputer itu terjangkit virus?" (Was that computer infected by virus?)

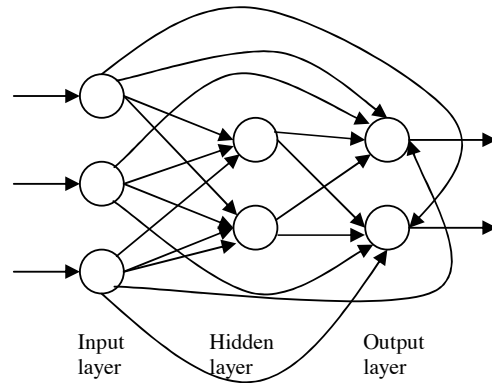


Figure 6: The architecture of neural network used in this research was fully connected feed forward neural network

In this research, we used a fully connected feed forward neural network such as depicted in Figure 6 to distinguish between declarative intonation and interrogative intonation. The number of nodes in input layer can be changed to be arbitrary positive integer. We used one; two or three nodes in input layer depend on the number of Fujisaki parameter we used. Both hidden layer and output layer consist of 2 nodes. The targets of the neural network are (0 1) for declarative intonation and (1 0) for interrogative intonation.

We used 30 utterances as training set of neural network, and 60 utterances which consisted also the training files as the testing set. At first time, we only used the amplitude of last accent command as an input of the neural network. We got the recognition rate of the neural network as 83.33 %. By using the amplitude of last accent command and the magnitude of last phrase command as the inputs of the neural network, we could increase the recognition rate to 90 %. After that, we tried to use Fb parameter as one of the inputs of neural network, but the result was deteriorated to only 50 %. However, we tried to reduce the effect of Fb by dividing all value of Fb by 100, and using together with the amplitude of last accent command and the magnitude of last phrase command as the inputs of neural network, we could get the highest recognition rate at 96.67 %. The summary of the recognition rate of the neural network was given in Table 2.

Table 2: The Recognition Rate of neural network

| Input of Neural Network          | Recognition Rate |
|----------------------------------|------------------|
| $A_{aJ}$                         | 83.33 %          |
| $A_{aJ}$ and $A_{pI}$            | 90 %             |
| $Fb$ , $A_{aJ}$ and $A_{pI}$     | 50 %             |
| $Fb/100$ , $A_{aJ}$ and $A_{pI}$ | 96.67 %          |

In this research, neural network could not recognize all the patterns correctly because of the resemblance of F0 contours. For example, Figure 4 and 5 depicted a declarative intonation

and interrogative intonation respectively. The F0 contour in Figure 4 we got from automatic Fujisaki parameter extraction was more likely an interrogative intonation than a declarative intonation. So, to increase the recognition rate of neural network by using the parameter proposed in this paper, in future, we plan to design a better new Fujisaki parameter extractor.

## 5. Conclusions

We proposed to use Fujisaki parameter as the features to classify intonation information of Indonesian speech. The highest recognition rate of neural network was achieved by using a fraction value of Fb: Fb/100, the amplitude of last accent command, and the magnitude of last phrase command as its inputs. To increase the recognition rate of neural network by using the parameter proposed in this paper, in future, we plan to modify a better new Fujisaki parameter extractor.

## 6. Acknowledgements

The authors would like to thank to the AUN/SEED Net for the scholarship of doctoral degree program to the first author, and Digital Signal Processing Research Laboratory, Chulalongkorn University for providing facilities.

## 7. References

- [1] J.W.M., Verhaar, *Pengantar Linguistik*, Gadjah Mada University Press, Jogjakarta, 1992.
- [2] J. Yuan, C., Shih and G.P., Kochanski, *Comparison of Declarative and Interrogative Intonation in Chinese*, in Proceedings of the Speech Prosody 2002 Conference, 2002
- [3] H., Fujisaki, and S., Ohno, *The use of a generative model of F0 contours for multilingual speech synthesis*, Proceedings of ICSP, pp. 714-717, 1998
- [4] P. Boersma and D., Weenink, *Praat: doing phonetics by computer*, University of Amsterdam, <http://www.praat.org>, 2004
- [5] H. Mixdorff, H. Fujisaki, G.P., Chen, and Y., Hu, *Towards the Automatic Extraction of Fujisaki model Parameters for Mandarin*, in proceedings of Eurospeech, Geneva, 2003