PROSODY MODELING OF IMPERATIVE SENTENCE IN BAHASA INDONESIA USING MATLAB

Arum Tri Iswari Purwanti
Universitas Gunadarma
Email: arum_tri@staff.gunadarma.ac.id

Muhammad Subali
Sekolah Tinggi Teknik Multimedia Cendekia Abditama (STTC)
Email: muhammadsubali@yahoo.com

ABSTRACT

Speech synthesis or text to speech is a speech processing technology and can change text to speech form. Speech synthesis applications are also available in Bahasa Indonesia which is known as TTS Indo. In these applications, there are still many shortcomings such as the sound produced using a sentence with a flat tone (without prosody), whereas in a sentence that prosody used to distinguish the meaning of the sentence. The sentences which are discussed in this journal are imperative sentences. Imperative sentence modeling on this research is basically seeing pitch (fundamental frequency) of the speaker by using MATLAB software. By looking at the results of the imperative sentence for Bahasa Indonesia, either one word or two words, both have a high tendency of intonation in the final seconds (at the end of the sentence). At the last, the prosody modeling analysis that inform the alteration of pitch can be used for prosody model of TTS Indo.

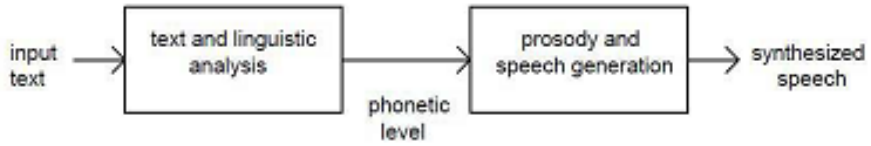
Key words: Pitch, Bahasa Indonesia, Imperative Sentence, Prosody Modeling.

Introduction

A. Introduction

Text-to-speech is widely used for all people in this world. Text-to-speech or which is known as speech synthesis is an application that converts an input of text and output of speech. Generally the use of speech synthesis is the text pronunciation aided for people who have difficulty in speaking. A simplified version of the procedure is presented in Figure 1. Improvement quality of speech synthesis system has been more remarkable. The system can produce output that sounds relatively close to human language. For the ears, two more important elements highlighted in this regard are the improvement of signal quality and improvement in the coherence and naturalness. The improvements in signal quality of good contemporary systems are mainly due to the use and improved control over concatenative speech technology, while the greater coherence and naturalness of synthetic speech are primarily a function of much improved prosodic modeling [Keller, 2002].

Figure 1. Simple text-to-speech synthesis procedure [Lemmetty, 1999]



Converting input of text into output of speech relies heavily on the rules which apply in a language. Language itself is a tool used by humans to be able to communicate with other humans. Language has a characteristic that also depends on one's area of origin. Not only where the area of someone is, but also how a person uttering a word also helped determine. The symbols are inputted text will naturally be represented in sound in a language [Arman, 2008]. Now, high-quality speech synthesizer has been available for some languages, e.g. English, French, Dutch, German and several other languages. Speech synthesis applications are also available in Bahasa Indonesia, where an application is often known as Indo TTS developed by Arry Akhmad Arman. But after trying this application there are still many shortcomings such as the sound produced using a sentence with a flat tone (without prosody). Whereas in a sentence has prosody which is used to distinguish the meaning of the sentence [Novianti, 2009].

The objective of this research is to analyze the intonation (prosody) of Imperative Sentences for Bahasa Indonesia. By analyzing the intonation of Imperative Sentences one by one, it will get the prosody modeling of the Imperative Sentences on Bahasa Indonesia. In the process of analyzing the Imperative Sentences used MATLAB Simulation

Software. The use of this software is to help obtaining the pitch values for every sentence, then processing it into a prosody form of Imperative Sentences on Bahasa Indonesia.

In this paper, we describe Bahasa Indonesia especially Imperative Sentence and intonation modeling as supporting theoritical. For the methodology research, Imperative Sentence has been chosen to define the prosody. It continued by recording each sentences which has samples and processing of data analysis. The main components of this research are the pitch obtaining process, filter definition, and fundamental frequency estimation. In result and analysis, the data obtained in Bahasa Indonesia are one word and two words of Imperative Sentence. Each data sentences has each samples which is continued by processing the data synchronization and data sampling. Thus, the conclusion declare the imperative sentence in Bahasa Indonesia tends to have high intonation at the end of the sentence and this data can be used to further research to enrich the language library especially in the prosody of Bahasa Indonesia.

B. Review Literature

Languages use pitch variation contrastively for the expression of discoursal meaning and for marking phrases. One of the important points developed in is that intonation is structural [Ladd, 2008], just as lexical tone is structural, or morphological paradigms are. In principle, an intonation contour has two structures: a morphological one, which identifies the morphemes and thus gives the meaning of the contour; and a phonological one, which gives its tones [Gussenhoven, 2004] [Clark, 2003].

Most of the predicate that states the order of verbs or verbal phrases, it is usually not accompanied by the subject [Sugono, 1999]. In other words, all verbs can express command. One of the command verb features has not a beginning. Another feature of Imperative Sentence is that if it be written at the end of sentence with an exclamation mark (!) [Indonesia, 2000].

1. Tembak!
2. Catat semua keterangan saksi!
3. Dengar baik-baik!

In addition to such basic verbs, derived verbs can express command. In this case also the prefix me(N)- which states actively uninstalled [Irwan, 2009].

1. Perbaiki mobil itu!
2. Bersatu!
3. Bukukan kisah itu!

In addition, the predicate that states the command is marked also by the particle -lah. Commands that use these particles feel more assertive, as shown in the following example.

1. Bacalah buku!
2. Tuntutlah ilmu!
3. Berdoalah!

In reality it is often refined form of the command so that it becomes an invitation, request, or prohibition. This type is usually preceded by words such as mari, mohon, silakan, harap, or jangan.

The related research has been discussed in Prototype Prosody in Speech Disorder Children. [Subali and friends, 2013] This study used the Dynamic Time Warping (DTW) method to calculate the distance between two time series data of the sound signal. Testing of these applications is done by finding the ratio error matching that state the probability of matching errors in the system. In another research, Analisis Frekuensi Dasar dan Frekuensi Formant dari Fonem Huruf Hijaiyah untuk Pengucapan Makhraj dengan Metode DTW has been discussed. The sound will be extracted to get the value of the fundamental frequency and formant frequency. After getting both frequencies, it will be obtained analysis of the similarities and differences in the fundamental frequency and formant frequencies of speech beginner and expert and it will shows matching distance of both speech. The result is all of speech beginner and expert based on makhraj pronunciation have different values of fundamental frequency and formant frequency.

C. Methodology

Imperative Sentence modeling on this research is basically seeing pitch of the speaker. Pitch has been described in the previous chapter which is the auditory sensation of tonal height. A collection of pitch is then processed to obtain a prosody modeling of the words entered by the speaker. The Prosody Modeling Research Process Diagram of Imperative Sentence can be shown in figure 2.

The recording process begins by making the Imperative Sentence which is used as prosody modeling. As remembered the Imperative Sentence can be made only by a predicate word, the sentences for one word are as follows [Bahasa, 2008]:

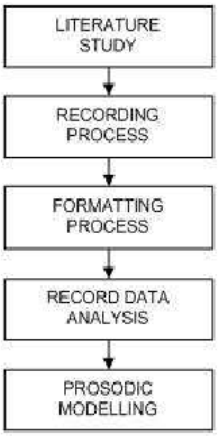
1. Baca!
2. Catat!

3. Duduk!
4. Jalan!
5. Makan!
6. Minum!
7. Pergi!
8. Tembak!

Furthermore, the sentences for two words of Imperative Sentence used a refined form. A refined form has another purpose in the form of an invitation, request, or prohibition. This type of sentence is preceded by words such as "mari, mohon, silakan, harap, or jangan". Then the sentence used for this study are as follows:

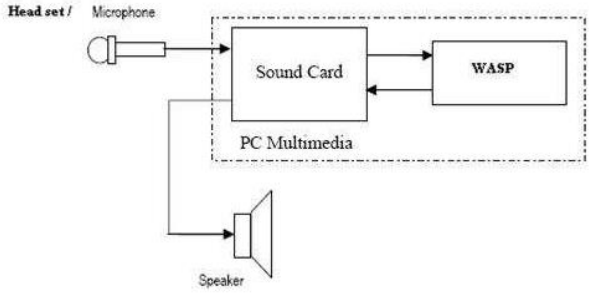
1. Ayo baca!
2. Ayo makan!
3. Harap tenang!
4. Jangan lompat!
5. Jangan pergi!
6. Jangan tembak!
7. Mari duduk!

Figure 2. The Prosody Modeling Research Process Diagram of Imperative Sentence



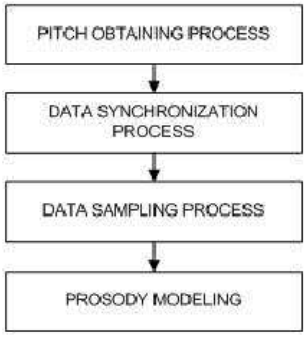
In addition to hardware, this research also using software to help the recording process (figure 3). Software that supports this research is the SFS / WASP version 1.5. WASP is a free program for the recording, display and analysis of speech. It can record and replay speech signals, save them and reload them from disk, edit annotations, and display spectrograms and a fundamental frequency track which is also designed to be compatible with the Speech Filing System (SFS) tools for speech research [of Psychology and Sciences, 2008].

Figure 3. Recording Tools [Novianti, 2009]



Analyzing the data has been obtained aiming to get prosody modeling of the Imperative Sentence. This process is performed after executing the formatting process. The data analysis processes consisted of pitch obtaining process, data synchronization process, data sampling process, and produce the desired prosody modeling in Figure 4.

Figure 4. Data Analysis Process



Pitch Obtaining Process

Pitch obtaining process which is known as the estimation of fundamental frequency, or pitch, is an essential component of many speech processing applications [Gonzalez and Brookes, 2011]. The proposed method used in this research is adapted of "A Pitch Estimation Filter Robust To High Levels Of Noise (PEFAC)". For a perfectly periodic source at frequency f0, our signal model at time t in the power spectral density domain is

$$Y_t(f) = \sum_{k=1}^{K} a_{k,t}\delta(f - kf_0) + N_t(f) \quad (1)$$

where N_t (f) represents the power spectral density of the unwanted noise and $a_{k,t}$ the power of the kth harmonic. In the log-frequency domain, the signal model can be expressed as

$$Y_t(q) = \sum_{k=1}^{K} a_{k,t}\delta(q - \log k - \log f_0) + N_t(q) \quad (2)$$

where q = log f. In this domain, the spacing of the harmonics is independent of f_0 and their energy can therefore be combined by convolving $Y_t(q)$ with a filter with impulse response

$$h(q) = \sum_{k=1}^{K} \delta(q - \log k) \quad (3)$$

The convolution $Y_t(q) * h(q)$ will include a peak at $q_0 = \log f_0$ and additional peaks corresponding to simple rational multiples of f_0.

Filter Definition

In practice, the width of each harmonic peak will be broadened due to the analysis window and the rate of change of f_0. Accordingly we use a filter with broadened peaks having the impulse response

$$h(q) = \beta - \log(\gamma - \cos(2\pi e^q)) \quad (4)$$

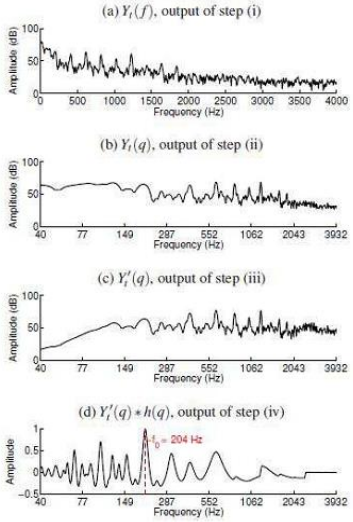
for $\log(0.5) < q < \log(K+0.5)$ and $h(q) = 0$ otherwise. γ is an algorithm parameter that controls the peak width while β is chosen so that $\int h(q)dq = 0$. The number of peaks, K, is restricted to 10 in order to reduce the response of $Y_t(q) * h(q)$ at values of q corresponding to sub-harmonics of f_0.

Fundamental Frequency Estimation

The complete PEFAC (Pitch Estimation Filter with Amplitude Compression) therefore comprises the following steps whose outputs are shown in figure 5 for a single voiced frame corrupted by car noise:

1. Transform the input signal to the time-frequency domain using the short-time Fourier transform (STFT), $Y_t(f)$
2. Interpolate the power spectral density (PSD) of each frame onto a log-spaced frequency grid, $Y_t(q)$
3. Find at α_t (q) so that the normalized smoothed spectrum $Y'_t(q)$ equals L(q) and calculate the compressed PSD, $Y'_t(q)$
4. Convolve the compressed PSD, Y't (q), with the analysis filter, h(q), and select the highest peak in the feasible range as the estimated pitch.

Figure 5. Algorithm processing steps for a single voiced frame of speech corrupted with car noise. (a) PSD in dB, (b) PSD in dB in a log-frequency grid, (c) compressed PSD in dB in a log-frequency grid, and (d) normalized output of the filter and fundamental frequency, f_0.



D. Result And Analysis

After getting the raw data of 10 times (pronunciation) in a time interval of 2 seconds, it continued by splitting the data into 10 different data. Each data (samples) used as input in Matlab to seek the pitch (fundamental frequency). Some figures below show sample data of Imperative Sentence.

The data obtained of Imperative Sentence in Bahasa Indonesia for two words

The data obtained of Imperative Sentence in Bahasa Indonesia for one word obtain time of 0.7553 seconds. The data obtained have not been carried out the data synchronization so that data obtained are not uniform (due to data separation). It seen that the frequency of the sample 1 to sample 10 does not have the same frequency at first. For that, emphasized the importance of synchronization for the raw data. Data synchronization is a data analysis based on the distance (interval) and time.

Figure 6. Data Sample 1 of word "Ayo Baca"

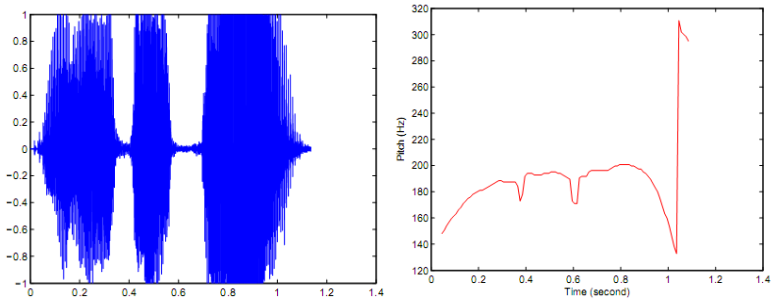


Table 1. Raw Data Sample of word "Ayo Baca"

	A	B	C	D	E	F	G	H	I	J	K	L	M
5	Time	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10		Final Baca
6	0.0453	125.40630	124.6883	124.6883	127.5853	126.1285	124.6883	129.8022	125.4063	126.8548	129.0590		124.6883
7	0.0553	126.85480	126.1285	127.5853	127.5853	127.5853	126.8548	131.3014	128.3200	128.3200	129.8022		127.5853
8	0.0653	128.32000	127.5853	129.0590	129.0590	128.3200	129.8022	132.0575	129.8022	129.8022	130.5496		129.8022
9	0.0753	129.05900	129.0590	130.5496	130.5496	130.5496	132.0575	132.8180	131.3014	131.3014	132.0575		130.5496
10	0.0853	130.54960	130.5496	132.8180	132.0575	132.8180	133.5829	133.5829	132.8180	132.8180	133.5829		132.0575
11	0.0953	132.05750	132.0575	134.3521	133.5829	134.3521	135.9039	135.1258	133.5829	134.3521	135.1258		133.5829
12	0.1053	269.14740	133.5829	135.9039	135.9039	135.9039	137.4737	135.9039	135.1258	135.9039	136.6866		135.9039
13	0.1153	273.82400	135.9039	138.2654	138.2654	138.2654	139.8624	138.2654	136.6866	138.2654	139.0616		138.2654
14	0.1253	278.58180	138.2654	143.1120	141.4778	141.4778	141.4778	140.6678	138.2654	140.6678	141.4778		141.4778
15	0.1353	142.29260	140.6678	146.4371	144.7650	144.7650	143.1120	142.2926	140.6678	144.7650	143.9361		144.7650
16	0.1453	144.76500	142.2926	148.9815	147.2803	146.4371	143.9361	144.7650	142.2926	148.1285	146.4371		146.4371
17	0.1553	147.28030	144.7650	151.5701	148.9815	148.1285	144.7650	146.4371	143.9361	149.8394	148.1285		148.1285
18	0.1653	149.83940	146.4371	154.2037	149.8394	148.9815	146.4371	147.2803	145.5986	151.5701	149.8394		149.83940
19	0.1753	151.57010	148.1285	155.9849	151.5701	148.9815	147.2803	148.9815	146.4371	152.4430	150.7023		151.57010
20	0.1853	152.44300	149.8394	158.6952	153.3208	149.8394	147.2803	150.7023	148.1285	153.3208	151.5701		153.3208
21	0.1953	154.20370	150.7023	160.5282	154.2037	149.8394	148.1285	151.5701	149.8394	155.0917	152.4430		154.2037
22	0.2053	155.98490	151.5701	161.4526	154.2037	150.7023	148.1285	152.4430	150.7023	156.8831	153.3208		150.7023

Data Synchronization Process

As already explained, the data analysis based on distance and time is a data synchronization process to obtain absolute valued. After the data synchronization process is expected that the resulting data can be directly used as the primary data for further analysis. The record data that have been processed are analyzed to determine the length/distance and determine the sample points based on the time and frequency. Data synchronization is done to equalize the initial frequency to the time. Before the data synchronization process, data is very irregular. This occurs due to the data separation done before. Then this data synchronization is to smoothing the irregular data so that the next analysis can proceed. In other words, data synchronization is done so that the data obtained more "apple to apple" so it can be processed to obtain a most appear value of the frequency for one unit of time.

Data Sampling Process

After executing the sampling process, it followed by finding the average value of sampling data. There are 8 Imperative Sentence of one word and 7 Imperative Sentence of two words that have previously been through the stage of the sampling process. The results of the sampling process have a prosody modeling for each Imperative Sentence.

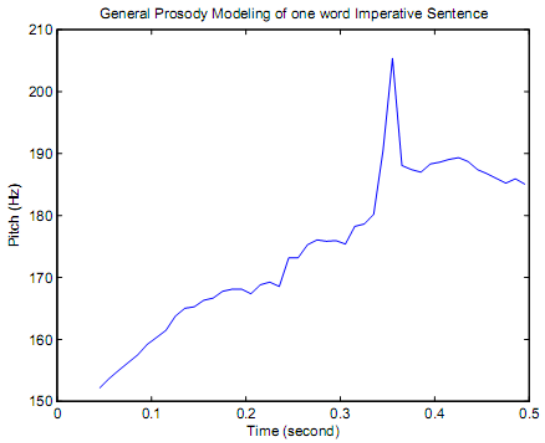
Prosody modeling is a data collection of frequency was selected after performing the sampling process. Prosody modeling of each Imperative Sentence is collected to find the average value. Collection of prosody modeling of every Imperative Sentence aims to seek a prosody modeling to the Imperative Sentence that has been used as input. In other words, for imperative sentence of one word and two words, the result of prosody modeling is different each other. Prosody modeling is the average data set of each Imperative Sentence for Bahasa Indonesia. The Figure 8 show a data set of each Imperative Sentence.

Table 2. Data collection of two words in Imperative Sentence

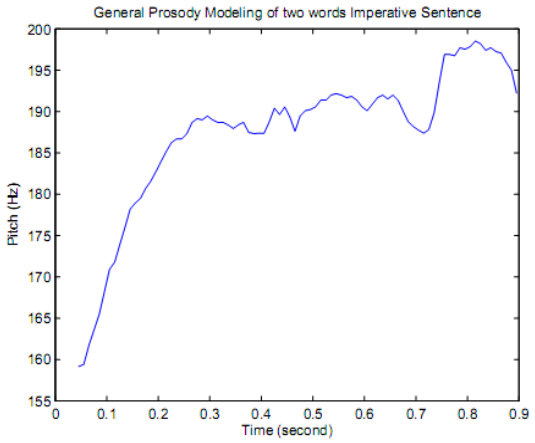
	A	B	C	D	E	F	G	H	I
5	Time	Final Ayo Baca	Final Ayo Makan	Final Harap Tenang	Final Jangan Lompat	Final Jangan Pergi	Final Jangan Tembak	Final Mari Duduk	Average Total
6	0.0453	155.9849	144.765	151.5701	164.2579	169.0422	162.3823	166.1552	159.1654
7	0.0553	156.8831	147.2803	152.443	165.2038	170.0157	155.9849	168.0743	159.4122
8	0.0653	159.609	150.7023	154.2037	167.112	171.9794	158.6952	170.0157	161.7596
9	0.0753	162.3823	153.3208	155.9849	169.0422	173.9658	159.609	170.9947	163.6142
10	0.0853	164.2579	155.0917	157.7865	172.9698	174.9676	161.4526	171.9794	165.5008
11	0.0953	166.1552	157.7865	160.5282	173.9658	175.9752	170.0157	172.9698	168.1995
12	0.1053	169.0422	164.2579	166.1552	174.9676	176.9886	170.9947	173.9658	170.9103
13	0.1153	171.9794	163.3174	162.3823	175.9752	179.0329	173.9658	175.9752	171.8040
14	0.1253	172.9698	169.0422	163.3174	180.0639	181.1008	173.9658	176.9886	173.9212
15	0.1353	174.9676	171.9794	169.0422	181.1008	181.1008	176.9886	176.9886	176.0240
16	0.1453	180.0639	173.9658	170.0157	182.1437	181.1008	179.0329	181.1008	178.2034
17	0.1553	178.0078	175.9752	170.0157	182.1437	182.1437	181.1008	183.1926	178.9399
18	0.1653	179.0329	174.9676	170.9947	182.1437	183.1926	182.1437	184.2475	179.5318
19	0.1753	180.0639	175.9752	173.9658	184.2475	184.2475	183.1926	183.1926	180.6979
20	0.1853	180.0639	178.0078	175.9752	185.3085	184.2475	184.2475	183.1926	181.5776
21	0.1953	181.1008	179.0329	180.0639	186.3756	185.3085	183.1926	184.2475	182.7603
22	0.2053	186.3756	180.0639	181.1008	186.3756	185.3085	182.1437	186.3756	183.9634
23	0.2153	190.7059	181.1008	182.1437	186.3756	186.3756	183.1926	186.3756	185.1814
24	0.2253	191.8041	182.1437	185.3085	186.3756	186.3756	184.2475	187.4489	186.2434
25	0.2353	191.8041	183.1926	185.3085	188.5284	187.4489	185.3085	185.3085	186.6999
26	0.2453	190.7059	184.2475	184.2475	189.614	188.5284	185.3085	184.2475	186.6999
27	0.2553	190.7059	185.3085	185.3085	189.614	189.614	185.3085	185.3085	187.3097
28	0.2653	189.614	186.3756	187.4489	190.7059	190.7059	186.3756	189.614	188.6914
29	0.2753	188.5284	187.4489	188.5284	190.7059	190.7059	186.3756	191.8041	189.1567

The processes carried out in this chapter are also done on other Imperative Sentence. As already known before, this research uses 8 one word and 7 two words. The sentences for one word are as follows, " Baca!, Catat!, Duduk!, Jalan!, Makan!, Minum!, Pergi!, Tembak!", and then the sentence used for two words are as follows, "Ayo baca!, Ayo makan!, Harap tenang!, Jangan lompat!, Jangan pergi!, Jangan tembak!, Mari duduk!".

Figure 7. (a) Prosody Modeling of one word in Imperative Sentence (b) General Prosody Modeling of two words in Imperative Sentence



(a)



(b)

E. Conclusion

The research that has been done, it started by learning literature study, collecting the Imperative Sentence for only one word and two words, processing the data recording, preparing the recording equipment, processing the record data formatting, and finally analyzing the record data. In analyzing the data, some of which stage of the process to obtain the pitch (fundamental frequency), the data synchronization process, the sampling data process included calculating the average value of sampling data, and finally to the ultimate goal of prosody modeling for one word and two words. By looking at the results of the Imperative Sentence for Bahasa Indonesia in the previous chapter, either one word or two words, both have a high tendency of intonation in the final seconds. This indicates the Imperative Sentence that tends to have high intonation at the end of sentence. Emphasis on the Imperative Sentence which written frequently by an exclamation mark (!) show that the presence of a high intonation on the sentence.

For further research, this data can be used for implementing Imperative Sentence in Speech synthesis applications in Bahasa Indonesia, where an application is often known as Indo TTS. Exclamation mark can be distinguish the Imperative sentence itself. When exclamation mark is used after the word, the pronounciation of the sentence should be changed as following this final pitch data.

References

[Arman, 2008] Arman, A. A. (2008). Text to speech bahasa indonesia.
[Arum, 2012] Arum Tri, I., P. (2012). Prosody Modeling of Imperative Sentence in Bahasa Indonesia Using MATLAB. Master thesis, Department of Electrical Engineering, Universitas Gunadarma.
[Bahasa, 2008] Bahasa, T. P. (2008). Kamus Bahasa Indonesia. Pusat Bahasa: Departemen Pendidikan Nasional.
[Brookes, 1997] Brookes, M. (1997). Voicebox: A speech processing toolbox for
matlab.
Clark, 2003] Clark, R. A. J. (2003). Generating Synthetic Pitch Contours Using Prosodic Structure. PhD thesis, The Department of Linguistics, University of Edinburgh.
[D. Byrne, 1994] D. Byrne, H. Dillon, K. T. e. a. (1994). An international comparison of long-term average speech spectra. In J. Acoust Soc Amer, volume 96, pages 2108–2120.
[Gussenhoven, 2004] Gussenhoven, C. (2004). The Phonology of Tone and Intonation. Cambridge University Press.
[Hirst and Cristo, 1998] Hirst, D. and Cristo, A. D. (1998). A survey of intonation systems. In Hirst, D. and Cristo, A. D., editors, Intonation Systems: A Survey of Twenty Languages, pages 1–44. Cambridge University Press.
[Indonesia, 2000] Indonesia, P. P. B. (2000). Pedoman Umum Ejaan Bahasa Indonesia Yang Disempurnakan. Pusat Bahasa: Departemen Pendidikan Nasional.
[Irwan, 2009] Irwan (2009). Pola kalimat perintah dalam bahasa angkola mandailing. Master's thesis, Universitas Sumatera Utara.
[Johan 't Hart and Cohen, 1990] Johan 't Hart, R. C. and Cohen, A. (1990). A perceptual study of intonation: An experimental-phonetic approach to speech melody. Cambridge University Press.
[Keller, 2002] Keller, E. (2002). Towards greater naturalness : Future directions of research in speech synthesis. In E. Keller, G. Bailly, A. M. J. T. and Huckvale, M., editors, Improvements in speech synthesis. COST 258: The naturalness of synthetic speech, pages 3–17. Chichester: John Wiley and Sons.
[Ladd, 2008] Ladd, D. R. (2008). Intonational Phonology. Cambridge University Press, second edition edition.
[Ladefoged, 1993] Ladefoged, P. (1993). A Course In Phonetics. Harcourt Brace College, third edition edition.
[Lemmetty, 1999] Lemmetty, S. (1999). Review of speech synthesis technology. Master's thesis, Helsinki University of Technology: Department of Electrical and Communications Engineering.
[Levine, 1977] Levine, A. (1977). English Intonation And Computerized Speech Synthesis. PhD thesis, Stanford University.
[Myers, 1995] Myers, R. E. W. R. H. (1995). Ilmu Peluang dan Statistika Untuk Insinyur dan Ilmuwan. ITB Bandung, fourth edition edition.
[Novianti, 2009] Novianti, D. (2009). Analisis model prosodi untuk kalimat tanya pada bahasa indonesia. Master's thesis, Gunadarma University.
[of Psychology and Sciences, 2008] of Psychology, U. D. and Sciences, L. (2008). Sfs/wasp version 1.5. Speech, Hearing and Phonetic Sciences.
[Smith, 2003] Smith, S. G. J. (2003). Predicting query types by prosodic analysis. PhD thesis, School of Electronic, The University of Birmingham.
[Subali and Novianti, 2009] Subali, M. and Novianti, D. (2009). Prosody model analysis of bahasa indonesia speech synthesizer using speech filing system.
[Subali and friends, 2013] Subali, M., Kholid, Muh., and Ningsih., T.W.R. (2013). Prototype Prosody in Speech Disorder Children. Language in India.
[Subali Miftah Chris, 2015] Subali, M., Andriansyah, M., and Sinambela, C. (2015). Analisis Frekuensi Dasar dan Frekuensi Formant dari Fonem Huruf Hijaiyah untuk Pengucapan Makhraj dengan Metode DTW. Prosiding PESAT (Psikologi, Ekonomi, Sastra, Arsitektur & Teknik Sipil). Universitas Gunadarma.
[Sugono, 1999] Sugono, D. (1999). Berbahasa Indonesia Dengan Benar. Puspaswara.
[Supranto, 2000] Supranto, J. (2000). Statistik: Teori dan Aplikasi. Erlangga, sixth edition edition.
[Tsegaye, 2005] Tsegaye, N. (2005). Speech to text conversion using amharic characters. Master's thesis, Addis Ababa University.
[van Heuven and van Zanten, 2007] van Heuven, V. J. and van Zanten, E. (2007). Prosody in Indonesian Languages. Leiden University Centre for Linguistics.
[Walpole, 1995] Walpole, R. E. (1995). Pengantar Statistika. Gramedia Pustaka Utama, third edition edition.
[Xia, 2004] Xia, L. (2004). Background research of speech recognition with modeling analysis. Master's thesis, University of Nottingham Malaysia.
[Gales, 2012] Gales, Mark. (2012). Acoustic Factorisation for Speech Recognition and Speech Synthesis. Cambridge University Engineering Department.
[Glass, 2012] Glass, James. (2012). Towards Unsupervised Speech Processing. The 11th International Conference on Information Sciences, Signal Processing and their Applications: Main Tracks. IEEE.