ミャンマー語の声調の規則合成とテキスト音声変換システム

Chapter 1
Introduction

1.1 Background

There has been ample research on speech processing in some tonal languages such as Chinese, Thai and Vietnamese. Nowadays, many applications in various information systems incorporate digital speech technologies, such as telecommunications, security and warning system, telephone inquiry system, multi-media applications, electronic mail reading, aids for the blind and people with speech impairments. This type of communication is progressing [1]. For Myanmar language, keeping track of world trends in this field is important for fostering cooperation in the field of global technology. For this purpose, the Myanmar TTS system is essential. While Text-To-Speech systems for major world languages are quite advanced, there has not been any Myanmar TTS system and research on speech synthesis is lacking.

Myanmar (Burmese) is official language in Myanmar. We choose the formal register for Myanmar speech synthesis

1.2 Myanmar speech synthesis by rule

In this dissertation, Myanmar Text To Speech system with rule-is designed. The MyanmarTTS system is a rule-based speech synthesis system, in which fundamental speech units are demisyllables with Level tone. We use Myanmar (Burmese) official language for the TTS system. To construct the TTS system, monosyllabic words are

analyzed and the parameters are obtained for synthesis of Myanmar tones. This TTS system uses a source filter model and furthermore a Log Magnitude Approximation Filter.

In Myanmar tones, F_0 patterns are very simple and which is unique comparing with other tonal language such as Vietnamese, Chinese and Thai. Standard Myanmar is used by 8 main races and sub races under the main races as an official language. It is spoken in most of the country with slight regional variations. In addition, there are other regional variants that differ from standard Myanmar in pronunciation and vocabulary. Myanmar tones are unique in their simplistic pattern not only related to F_0 but also more specifically and importantly to length. This is the basis for the proposed linear pattern for tone rule using normalized F_0 and length of each tone among the utterances.

1.3 Thesis's objective

The purpose of this research is

-to design the first Myanmar TTS system

-to study the tone synthesis for tonal language

-to implementing tone synthesis by rule

-to examining normalization in fundamental frequency and length of syllables

1.4 Thesis's outline

The thesis includes 6 chapters; the rest is organized as follows.

Chapter 2 covers the overview of Myanmar language. Then, chapter 3 mentions speech analysis and synthesis. Chapter 4 introduces Myanmar Text-To-Speech System, Rule-based Tone Synthesis, and tone rule with F_0 linear pattern. Next, chapter 5 raises Tone synthesis, normalization in F_0, normalization in syllable length. Chapter 6 reviews the main aspects and concludes the thesis.

Chapter 2
An Overview of Myanmar language

Myanmar belongs to the Lolo-Burmese sub-branch of the Tibeto-Burmese branch of the Sino-Tibetan language family. Myanmar script draws its source from Brahmi script which flourished in India from about 500 B.C. to over 300AD. Myanmar is a tonal language. This means that all syllables in Myanmar have prosodic features that are an integral part of their pronunciation. Prosodic contrasts involve not only pitch, but also phonation. Standard Myanmar is based on the dialect spoken in the lower valleys of the Irrawaddy and Chindwin rivers. It is spoken in most of the country with slight regional variations. In addition, there are other regional variants that differ from standard Myanmar in pronunciation and vocabulary. All dialects are mutually intelligible. In addition, there are two registers: a formal and a colloquial one. The formal register is used in official publications, radio and TV broadcasts, literary works, and formal speech. The colloquial register is used in daily communications. In Myanmar there are 8 main races and 135 sub races under the main races. Myanmar (Burmese) is official language in Myanmar. We choose the formal register for Myanmar Speech Synthesis.

ဘ	ခ	ဂ	ဃ	င
စ	ဆ	ဇ	ဈ	ည
ဋ	ဌ	ဍ		ဏ
တ	ထ	ဒ	ဓ	န
ပ	ဖ	ဗ	ဘ	မ
ယ	ရ	လ	ဝ	သ
	ဟ	ဠ	အ	

Fig. 2-1 Myanmar Consonantal letters

2.1 Consonants

There are 33 consonantal letters in Myanmar. which are generally classified as plosive and affricate, nasal, fricative, approximant and lateral. The approximant /ɹ/ is rare, and is only used in place names that have preserved Sanskrit or Pali pronunciations. The 33 consonants are represented by 26 phonemes since some some consonantal letters represents the same phoneme. For example the consonants /ဂ/ and /ဃ / represent the same phoneme/g/, the consonant /ဒ/ and /ဓ / represent the same phoneme /d/. The list of Myanmar consonantal letters and their corresponding symbols in International Phonetic Alphabets (IPA) are shown in table 2-1. The IPA equivalent Myanmar consonantal letters classified in the place articulation and the manner of articulation are as shown in table 2-2.

Table2-1. Myanmar phonemes and consonantal letters used in the system

Phonemes	Myanmar letters	Character input
ə	အ	a
k	က	ka
k^h	ခ	kha
g	ဂ	ga
ŋ	င	nga
s	စ	sa
s^h	ဆ	sha
z	ဇ	za
ɲ	ည	nya
t	တ	ta
t^h	ထ	hta
d	ဒ။ဓ	da
n	န။ဏ	na
p	ပ	pa
p^h	ဖ	hpa
b	ဗ။ဘ	ba
m	မ	ma
j	ယ	ya
ɹ	ရ	ra
l	လ	la
w	ဝ	wa
θ	သ	tha
h	ဟ	ha
ʧ	ကျ ။ ⿰	ca
ʧʰ	ချ ။ ⿰	cha
ʤ	ဂျ ။ ⿰	gya

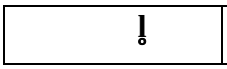
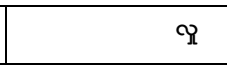
		hla

Table 2-2. The 33 consonantal letters of Myanmar (IPA)

	Bilabial	Dental	Alveolar	Postalveolar and palatal	Velar and labiovelar	Glottal	Placeless
Plosive and Affricate	**pʰ p b**	**tʰ t d**		**tɕʰ tɕ dʑ**	**kʰ k ɡ**	**ʔ**	
Nasal	**m̥ m**	**n̥ n**		**ɲ̥ ɲ**	**ŋ̊ ŋ**		**N**
Fricative		**θ (ð)**	**sʰ s z**	**ʃ**		**h**	
Approximant		**(ɹ)**		**j**	**(w̥) w**		
Lateral		**l̥ l**					

2.2 Vowels

Myanmar has basically 12 vowels, 8 monophthongs and 4 diphthongs. Table 2-3 shows Myanmar monophthongs and dipthongs classified according to tongue heights and positions.

Table 2-3. The 12 vowels of Myanmar

Tongue Height	Monophthongs		Diphthongs	
	Front	Back	Front offglide	Back offglide
Close	i	u		
Close-mid	e	o	ei	ou
Mid	ə			
Open-mid	ɛ	ɔ		
Open		a	ai	au

2.3 Tones

Myanmar has 4 tones and a simple syllable structure that consists of an initial consonant followed by a vowel with an associated tone. This means all syllables in Myanmar have prosodic features. Different tone makes different meanings for syllables with the same structure of phonemes. In the Myanmar writing system, a tone is presented by a diacritic mark. The four Myanmar tones are shown in Table 2-4. Fig.2-2 and Fig.2-3 shows an example of F_0 contour of four Myanmar tones with syllable /ma/. The Falling and checked tones belongs to the short-tone group and high falling and level tones belongs to the long-tone group. Checked tone is accompanied by a glottal stop. More details on Myanmar language can be referred to in [5, 6, and 7].

Table 2-4 The Example four Myanmar tones

Tone name	Symbol	Description
Level	လာ	**/là/ - moon**
Falling(Creaky)	လ	**/la/ - come**
High Falling(High)	လား	**/lá/ dray horse**
High Extremely short(Checked)	လတ်	**/laʔ/- fresh**

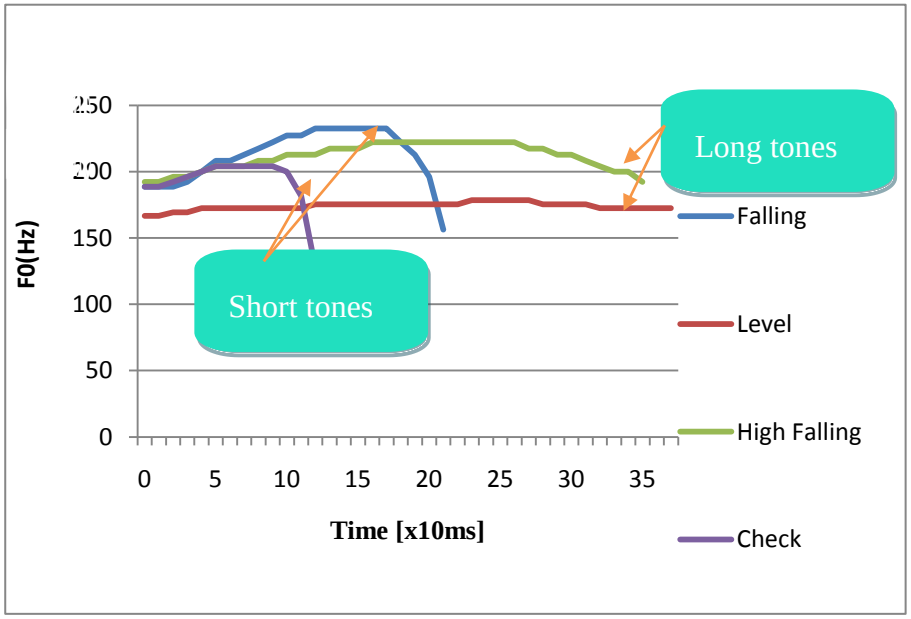


Figure 2-2 The example of F_0 contours of four Myanmar Tones with syllables /ma/ uttered by female speaker

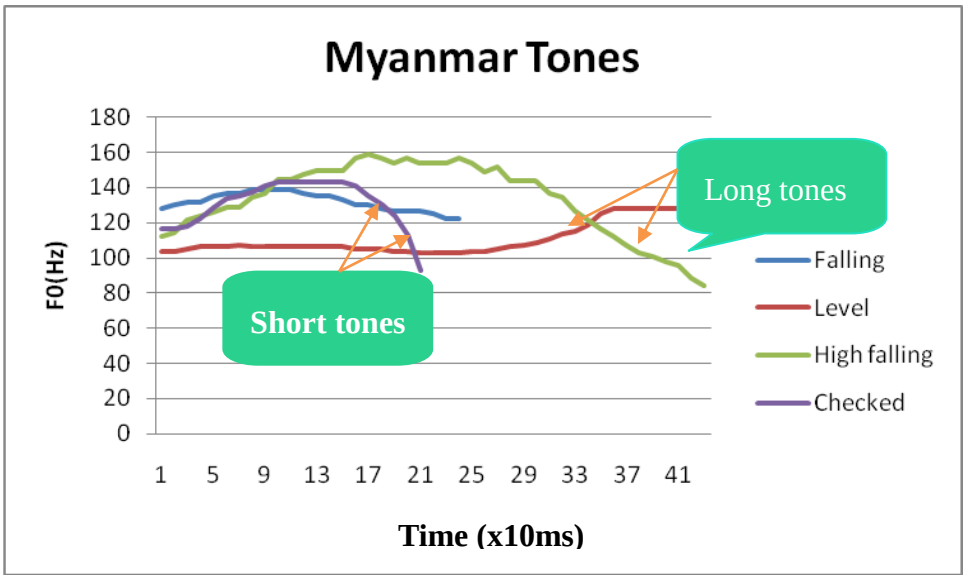


Figure 2-3 The example of F_0 contours of four Myanmar Tones with syllables /ma/ uttered by male speaker

Chapter 3
Speech analysis and synthesis

In this his chapter the basic concept of speech analysis and synthesis of Myanmar is explained. We developed the first Myanmar speech synthesis system using cepstral method.

3.1 Speech Analysis

The Myanmar synthesis system adopts short-time cepstral analysis with the frame length is 25.6ms and the frame interval or frame shifting time is 10ms. The cepstral analysis is a non-linear signal processing technique. A time–domain Hamming window with a length of 25.6 ms is used in our analysis system.

The expression of Hamming window is:

$$w(n) = 0.54 - 0.46 \text{ Cos } \left(\frac{2\pi}{N-1} n\right) \qquad (3\text{-}1)$$

N is 256 which equal to the length of frame L = NT = 25.6 ms at 10 kHz sampling frequency. The frequency resolution Δf is:

$$\Delta f = \frac{1}{NT} - \frac{1}{256\times10^{-4}} \approx 39.1\ [Hz] \qquad (3\text{-}2)$$

The cepstrum is defined as the inverse Fourier transform of the short time logarithmic

amplitude spectrum [3]. For signal $x(n) = e(v) * v(n)$, where $e(v)$ and $v(n)$ are th excitation function and the vocal track response, respectively, cepstrum is calculated as:

$$\begin{aligned} c(m) &= \text{IDFT}(\log|\text{DFT}(e(n) * v\,(n))| \\ &= \text{IDFT}(\log|E(k)v\,(k)|) \\ &= \text{IDFT}(\log|E(K) + \text{Log}|v\,(k)|) \end{aligned} \qquad (3\text{-}3)$$

Cepstral analysis has advantage that it can separate the spectral envelope part and excitation part. Process of the cepstral analysis is illustrated in the Fig.3-1. The resulting parameter of speech unit include the number of frames and, for each frame, voiced/unvoiced (V/UV) decision, pitch period and cepstral coefficients $c(m)$, $0 \leq m \leq 29$.

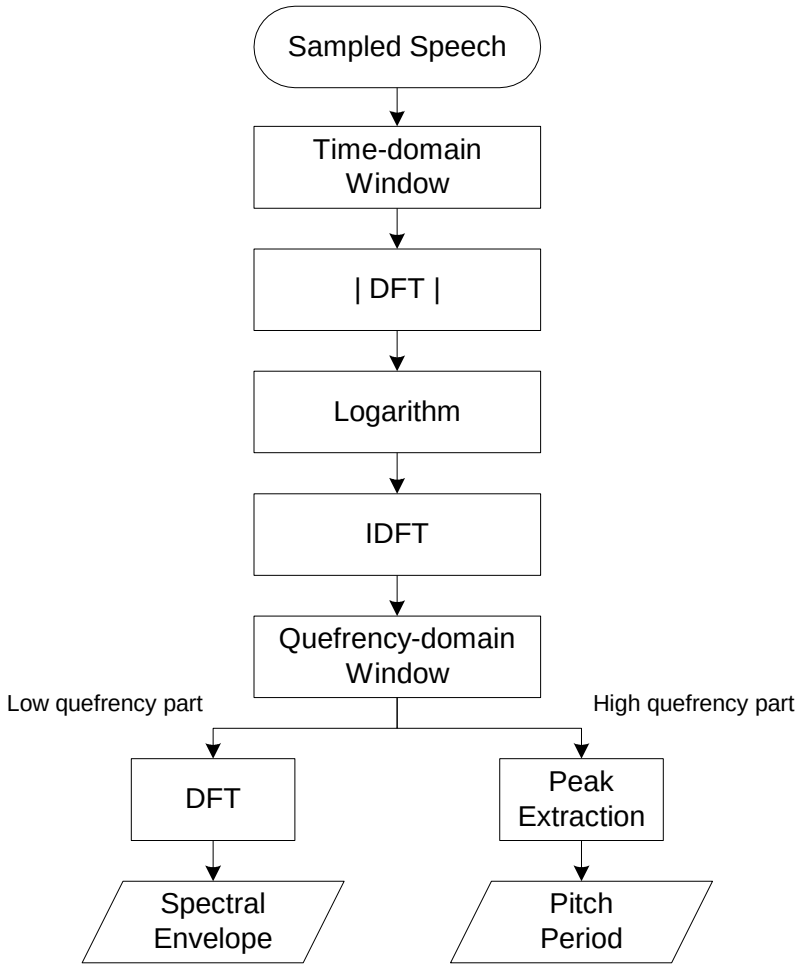


Fig 3-1 Cepstral analysis of speech signal

The analysis part of our TTS system is designed using cepstral analysis. The frame length is 25.6ms and the frame shifting time is 10ms. As the window function for speech analysis, a time–domain Hamming window is used with the length equal to frame length 25.6ms.

The cepstral coefficient or cepstrum is defined as the inverse Fourier transform of the short-time logarithmic amplitude spectrum [9]. The special feature of the cepstrum is that it allows for the separate representation of the spectral envelope and excitation. The resulting parameter of speech units include the number of frames and, for each frame, voiced/unvoiced (V/UV) decision, pitch period and cepstral coefficients $c(m)$, $0 \leq m \leq 29$.

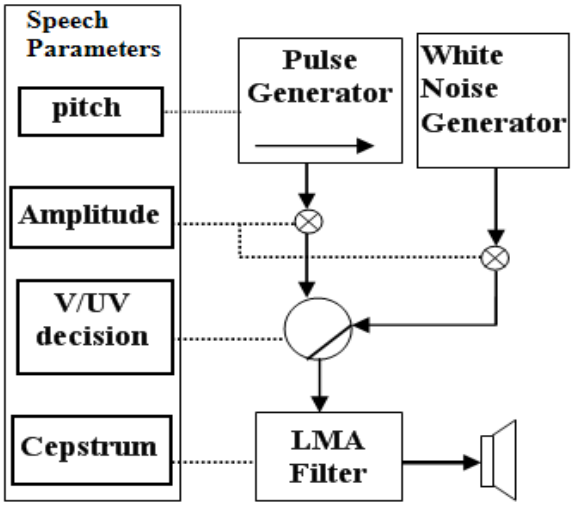


Fig.3-2 MyanmarTTS's speech synthesis sub-system

3.2 Speech Synthesis

The speech synthesis sub-system generates speech from pre-stored parameters. The source-filter model [10] is used as the speech production model. Fig. 3-2 shows the structure of the speech synthesis sub-system in MyanmarTTS. The synthetic sound is produced using the Log Magnitude Approximation (LMA) filter, which has been

introduced by Imai [2]. It presents the vocal tract characteristics. The spectral envelope is represented by the cepstral coefficients of 30 lower-order frequency elements. The LMA filter is a pole–zero filters that is able to represent efficiently the vocal tract features for all speech sounds.

We synthesized speech from pre-stored parameters with over 800 demi-syllables as speech units under the system rules. After synthesizing, the waveform was displayed and played. As a preliminary result, the quality of the synthetic sounds was intelligible.

3.3 Demisyllable database construction

Myanmar speech synthesis system is designed based on the idea of concation of speech unit, which are represented as cepstrum parameter.

3.3.1 Structure of demisyllables

The Myanmar syllable has the structure as shown in Fig.3-3. The syllable has the phonemic shape of C (G) V (N/ʔ) T, where an initial consonant C is obligatory, a glide consonant G is optional, a vowel V is obligatory, a final consonant-nasal N or stopped ʔ is optional, and tone T is obligatory, respectively. The minimum syllable is CVT. There are 6 possible syllable structures are as follows:

- CV /mè/ 'girl'
- CVC /mɛʔ/ 'crave'
- CGV /mjè/ 'earth'
- CGVC /mjɛʔ/ 'eye'
- CVVC /màuɴ/ (term of address for young men)
- CGVVC /mjáuɴ/ 'ditch'

Initial consonant	Glide consonant	Vowel	Final consonant	Tone
C	(G)	V	(N/ʔ)	T

Fig.3-3. Myanmar syllable structure

Initial consonant: Each of 33 consonants can be initial consonants.

Final consonant: The only consonants that can stand in the final consonants are /ʔ/ and /N/.

3.3.2 Demisyllable database

The demisyllable database is constructed for MyanmarTTS system and which is the first database for Myanmar speech synthesis system. Myanmar Consonant letters can be modified by one or more medial diacritics. The demisyllable database is constructed by combination of 33initial consonants and 12 vowels. The list of all combination CV units are shown in Appendix A, respectively. In order to prepare database Myanmar demi-syllables are collected and their sounds were prepared by recording on digital audio tape (DAT) at a 48 kHz sampling rate with a 16-bit resolution. After that, they were down-sampled to 10 kHz for analysis. Format of speech unit is Plus Code Modulation without compression. The result parameters are stored in database for synthesizing. All speech units are recorded with normal speaking rate**.**

Chapter 4
Text-to-speech system with Rule-based tone synthesis

4.1 MyanmarTTS System

The design of the MyanmarTTS system is shown in Fig. 4-1. This design is based on a general speech synthesis system [8]. The input text comprises Myanmar-script-equivalent characters. The output is the synthetic sound. In our system, the fundamental speech consisted of demi-syllables that were acquired by dividing a syllable with a cut-off point at 100ms of the vowel part for short vowel and a cut-off point of the middle of vowel part for long vowel. There are about 800 demi-syllables in Myanmar.

4.2 Text Analysis

The purpose of our text analysis was to extract the phonetic and prosodic features. We developed Myanmar pronunciation lists (database) using Myanmar-script-equivalent English alphabets. The input text is Myanmar-script-equivalent English alphabets as in Table 2-1. With the rule for mapping, the sequence of mapped characters is obtained from the list and then the information of the character is retrieved. We used English alphabets as the target of mapped sequences and made a mapping table for 4 tones as in Table 4-1.

For example, the sentence “သူမလှသည်” which means “She is beautiful.” is mapped into a sequence of syllables as “thu0ma1hla1thi0.”

Table 4-1.The rule for mapping

Marks	Number
Level tone	0
Falling tone	1
High falling tone	2
Checked tone	3

4.3 Speech Synthesis

The fundamental speech units of MyanmarTTS are demi-syllables which are developed by dividing a syllable with a cut-off point of 100ms of the vowel part for short vowel and a cut-off point of the middle of vowel part for long vowel.

4.4 System Rules

4.4.1 Syllable connection rules

A syllable is composed by demisyllables and a tone. Myanmar has not only monosyllable words but also disyllables words and tri-syllables words are included. In Myanmar, since most of syllables structure is CVT, the minimal minimum syllable, the connection rule is relatively simple. For two or three syllables words, the demisyllables are connected from the middle of vowel position of first syllable. Connection is implemented by interpolation of cepstral coefficient.

4.4.2 Interval

Myanmar has three kinds of interval marks. The interval rules are defined in database. The interval rules for MyanmarTTS are shown in table 4-2.

Table 4-2.The Interval rule

Interval mark	Symbol	Interval[ms]
Space		60
One stroke (Like a Comma)	၊	700
Two stroke (Like a Full stop)	။	1500

4.4.3 Tone

The four tones were analyzed to extract F_0 patterns. The data set was prepared as voiced sounds and meaningful words. We selected consonant-vowel (CV) form with voiced consonants /b/, /m/, and /l/ and three typical vowels /a/, /i/ and /u/. In total, 180 words (i.e., 3 consonants x 3 vowels x 4 tones x 5 speakers) were used for tone analysis. Myanmar tones have simple F_0 patterns and different lengths between short-tone and long-tone groups which are not only related to F_0 but also more specifically and importantly to length. A change in length constitutes a change in feature of tone. To obtain relative values among the tones, F0 and length of each tone was normalized Tone normalization procedure are presented in section 4.6.

4.4.4 Intonation

Intonation refers to the pattern of F_0 changes that extended over a segment of speech unit such as a phrase or sentence. In our system Fujisaki model for sentence intonation is adopted. An analysis of intonation is performed by considering pitch patterns in term of contour, for which pitch range height, and direction changes are generally characterized. The intonation is implemented by applying a simple declination line in log frequency domain.

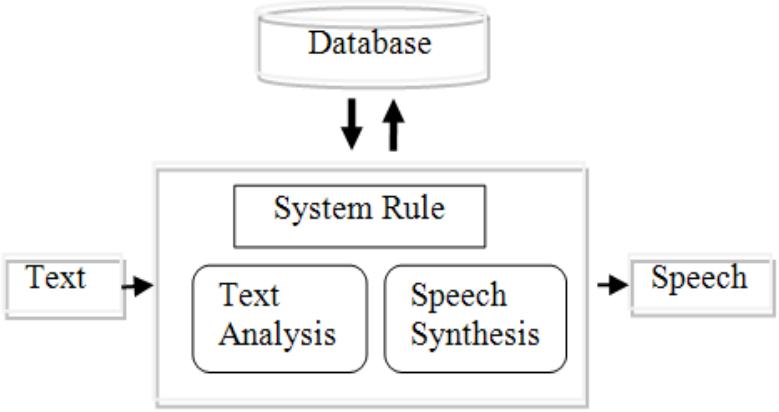


Fig. 4-1 Diagram of Myanmar TTS system.

4.5 Evaluation

4.5.1 Syllable's Intelligibility Test

All syllables in Myanmar have prosodic features that are an integral part of their pronunciation. In order to evaluate the intelligibility of MyanmarTTS system, syllables intelligibility test was carried out. To create a set of data we used JEIDA guide line (JEIDA-G-24-2000) as a reference. The test methodology is introduced by Imai, S and Abe,Y. [16].

Speech materials

The test consisted of 248 monosyllable words. The test stimuli are analysis-synthesis sounds with level tone which has a flat F_0 pattern. Meaningful monosyllables words were mostly used as well as some meaningless monosyllables words were used. Those monosyllable words are analysis-synthesis sounds with level tone and disregarding with other tones. The total number of sounds was 496 (i.e., 31 consonants x 8 vowels x 2 repetitions).

Method

Five native Myanmar listeners participated in listening tests. All listeners had normal hearing ability. In the test mono syllable words were presented randomly. The listening tests were performed in a sound proof room and the listeners used headphones (SONY MBR-XB700). Instructions for the test were prepared and the listener performed the practice test once. Each listener performed the test twice. The listener listened word by word and wrote the word they heard on the answer sheet. Each word was played one time within a 2 second interval with a pause time of 3 seconds to write the answer. The time taken for each test is 48 minutes

4.5.2 Result and discussion

Result

The average intelligibility score were 92.56%. We found that most of error came from nasal consonants word and dental consonant word. The results of intelligibility test are as shown in Fig. 4-2.and Fig.4-3. The score of intelligibility test for Japanese [12] is 91.2%. Comparing with Japanese intelligibility test, we think that our score is reasonable and enough for syllable's intelligibility.

Discussion

The results of correct answers for each syllable are shown in Fig.4-3. The score of correct answers for each syllable were reasonably high except the score of /θ/and /ŋ/.The score of /θ/ was worse than the other syllables, since /θ/ is very similar to /th/ and /t/, dental consonants. Besides the meaningless syllables combined with /θ/ were used in the listening test, which makes confusion among these dental consonants.

Also the score of /ŋ/ is worse than the other syllables, since /ŋ/ is very similar to /mya/, nasal consonant. And also the meaningless syllables with /ŋ/ were used in the listening test, which makes confusion among these nasal consonants. Besides, /ŋ/ is mostly used as final consonant for VC and the ending consonant of the sentence in

writing system. According to the above mentioned conditions, we think the intelligibility scores of /θ/and /η/ was worse than other syllables.

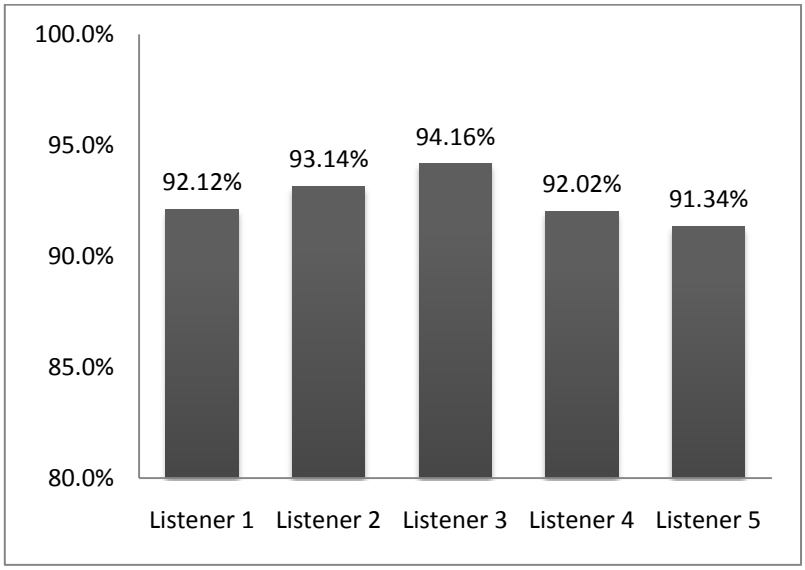


Fig.4-2 The results of intelligibility test for Myanmar syllables.

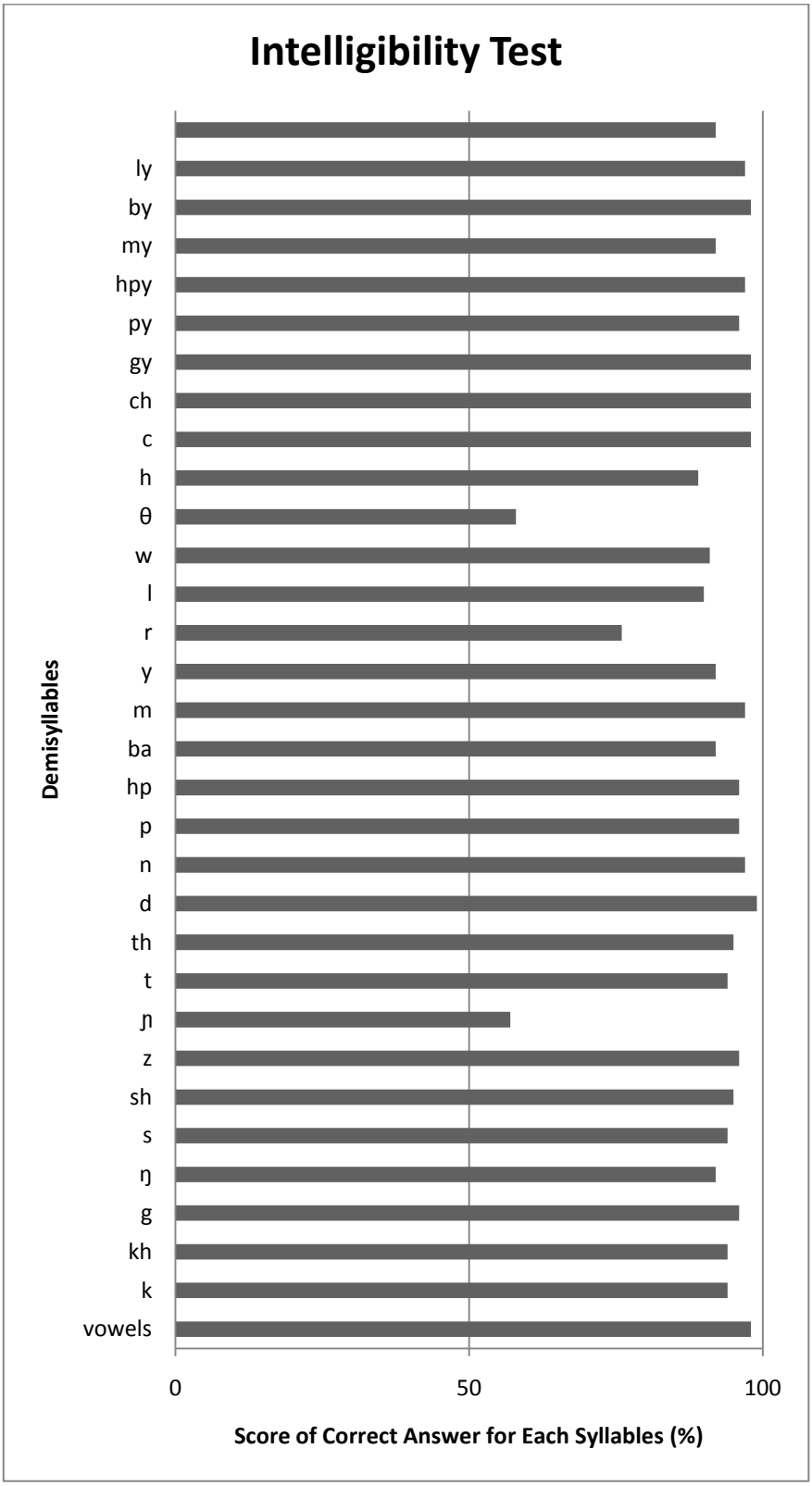


Fig. 4-3 The results of correct answers for each syllable**.**

4.6 Tone synthesis procedure

In this section tone synthesis rule and normalization of F_0 and duration are presented.

4.6.1 Tone Synthesis

The four tones were analyzed to extract F_0 patterns. The data set was prepared as voiced sounds and meaningful words. We selected consonant-vowel (CV) form with voiced consonants /b/, /m/, and /l/ and three typical vowels /a/, /i/ and /u/. In total, 180 words (i.e., 3 consonants x 3 vowels x 4 tones x 5 speakers) were used for tone analysis. After analyzing, four tones were distributed as shown in Fig. 4-4. We found that the four tone groups overlapped and were not clearly discriminated. Myanmar tones have simple F_0 patterns and different lengths between short-tone and long-tone groups. The Falling and the Checked tones belong to the short-tone group and the High falling and the Level tones belong to the long-tone group. The Checked tone has very short length and a glottal stop feature. Glottal stop in a speech synthesis system has been studied by Takara [11]. Consequently, we considered that the minimization of F_0 variations and lengths variations are required for tone discrimination.

Myanmar tones are not only related to F_0 but also more specifically and importantly to length. A change in length constitutes a change in feature of tone. For this reason, we normalized F0 and length to obtain relative values among the tones.

4.6.2 Normalization in F0 and Length

We selected F_0 from three frames at the center of syllable for each tone and for each speaker. The average F_0 was defined by means of a male speaker, a female speaker and both of male and female speakers to fix a standard value. Similarly the average length for each tone was defined. Subsequently, the F_0 and length of each tone for each speaker was normalized to an approximated standard value. In this view, our

normalization method is different and unique from other studies of tone synthesis research on such as Thai and Vietnamese [3, 4].

To minimize large differences of F_0 s and lengths among the speakers in a same tone, normalization is carried out as follows. The average of F_0 for each tone is given by

$$f_{ij} = \frac{1}{n}\sum_{k=1}^{n} f_{ij}^{(k)} \qquad (4\text{-}1)$$

where, n is number of F_0 frames at the center of syllables . f_{ij} is the average F_0 from the central n frames of syllable for i^{th} tone and j^{th} speaker. In this study, we set n = 3. The average F_0 of four tones is defined as A_j .

$$A_j = \frac{1}{t}\sum_{i=1}^{t} f_{ij} \qquad (4\text{-}2)$$

where, t is number of tones, which is four. The average of all speakers is defined as A_0.

$$A_0 = \frac{1}{s}\sum_{i=1}^{s} A_j \qquad (4\text{-}3)$$

where, s is number of speakers. In this research, s = 5 were used. We define the relative value R_{ij} as

$$R_{ij} = f_{ij} - A_j \qquad (4\text{-}4)$$

$f_{ij}^{\,0}$ is ruled-based F_0 for i^{th} tone, which is defined as

$$\boldsymbol{f_{ij}^{\,0}} = A_0 + \boldsymbol{R_{ij}} \qquad (4\text{-}5)$$

Similarly, the durations for each tone were normalized.

In Fig.4-5, we can see the result which plotted. Figure 4-5 shows distribution of four tones for both males and females after normalization in F_0 and duration, which are clearly grouped and discriminated. These results mean that the duration and F_0 are equally important for discriminating the tones. From this figure, we confirmed that

normalization is an effective method of discriminating the four tones. In Fig.4-5, The normalized F_0 and duration of four tones for the words “ba”, “mi”, and “lu” are distributed. These distributions include the F_0 and duration for both males and females. To define general tone rules for male and female we calculated the average of two parameters, F_0 and duration, from the results of Fig.4-5. The tone rules are implemented with F_0 linear patterns as shown in Fig.4-6.We applied the equations (4) and (5) in Fig.4-6. The relative values of two parameters, F_0 and duration are used with some simplifications. We set the average F_0 value of the Level tone as F_b for the tone rule. F_b for males is 110Hz and that for females is 160Hz. We carried out listening tests to evaluate intelligibilities of tones for synthetic speech of syllables and to evaluate the effect of normalization by using these rules.

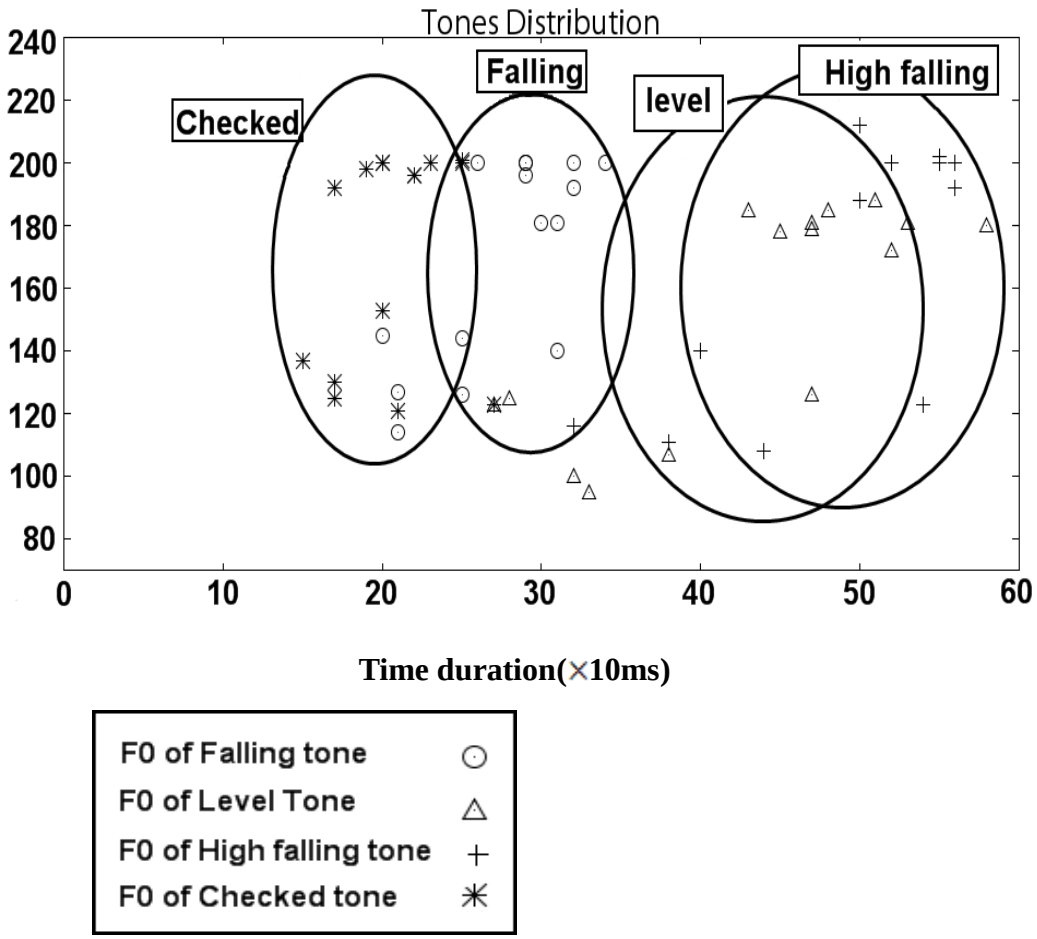


Fig.4-4. Tones distribution of the analysis synthesis words “ba”, “mi”, and “lu” uttered by three female speakers and two male speakers

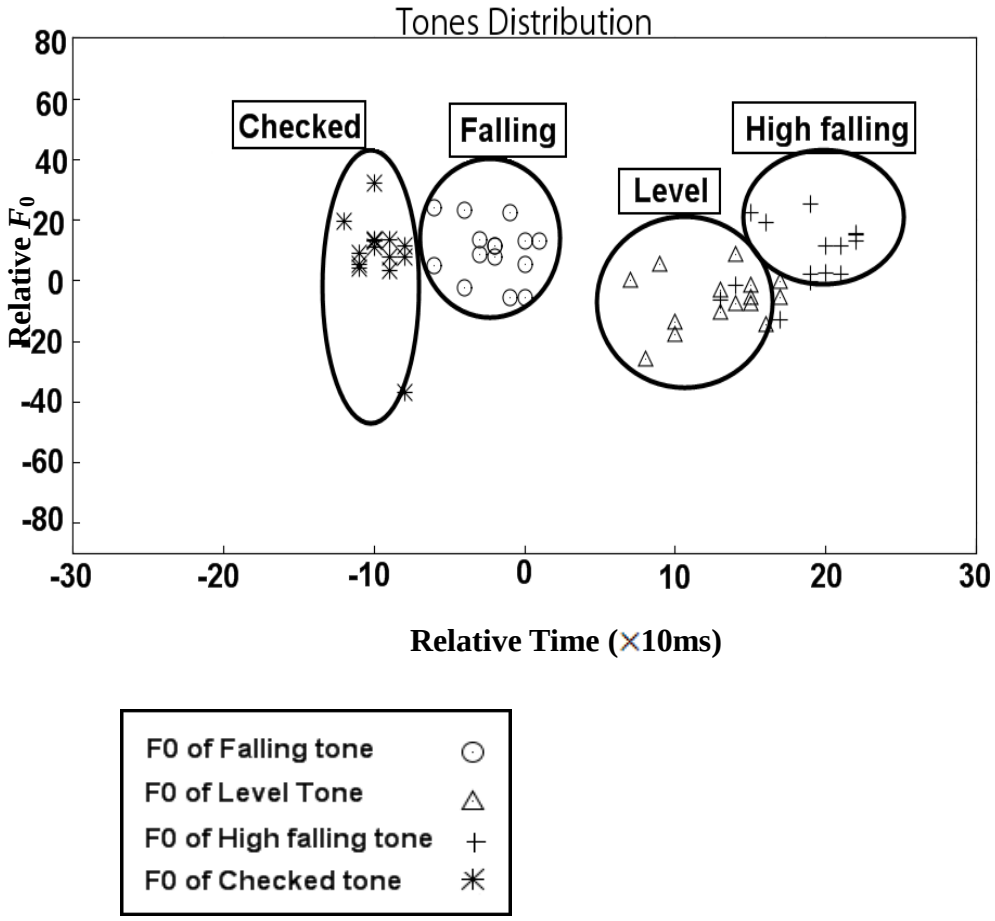
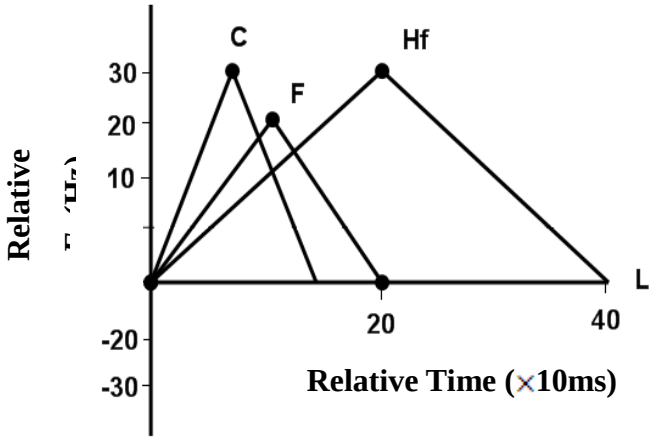


Fig. 4-5 Tone distribution of the words "ba", "mi", and "lu" uttered by three female speakers and two male speakers after normalization in F_0 and time (length).



L: Level tone, F: Falling tone, Hf: High falling tone, C: Checked tone

Fig.4-6. Diagram of tone rule for males

4.7 Evaluation by listening test of tones' intelligibility

The intelligibility of tones was evaluated through three types of listening test by three native listeners. We incorporated male and female speech.

Type 1: Rule-based sounds

1-1 : Synthetic sounds: Rule-based male sounds

1-2 : Synthetic sounds: Rule-based female sounds

Type 2: Analysis-Synthesis sounds

Type 3: Original sounds

All synthetic sounds used cepstra from speech units with level tone. The word set comprises three typical vowels "a", "i" and "u" with voiced consonants "b", "m" and "l" then syllable words were "ba", "mi" and "lu". Forty datums were prepared, given ten tokens for each tone and all sounds were meaningful words. The total number of sounds for each type was 240 (i.e., 3 words x 4 tones x 10 tokens x 2 genders).

All types of listening tests were done for each gender and for each word. In the all tests, each sound was played once at random and the listeners had to choose a word out of four possible words (tones) within 2 seconds. The listening tests were performed in a sound proof room. Among the three listeners, two were Yangon (Capital of Myanmar) natives and one was from the middle of Myanmar. They had normal hearing ability.

4.8 Results and discussion

Results of these tests are shown in Table 4-3 and Fig.4-7. Table 4-3 shows confusion matrices of the tests. Figure 4-7 shows the correct rate of listening test. The result of our tone synthesis system and effectiveness of normalization are discussed as follows:

- Type 1, proposed method achieves the reasonably high correct rate 95.8% for male speakers and 97.8% for female speakers. Type 2, analysis synthesis sounds, elicits a high correct rate 98.8% and Type 3, original sounds, elicits the highest correct rate 100%. These results indicate that the proposed method has a high intelligibility score for listening test.
- The proposed method is similar to the F_0 linear pattern of the VieTTS system [4] with the same analysis-synthesis method. In VieTTS system, the result for the linear pattern is about 85% for male, whereas the result of our system for male is 95.8%, even though our listening tests were done using the speech sounds of multiple speakers and different genders.
- Compared with these results, our linear pattern of tone rule is more effective than VieTTS's corresponding one because we adopted the normalization method for multiple speakers with different genders and for multiple syllables.

- From the confusion matrices of Type 1, we can see the errors occurred mostly between the High falling tone and the Level tone and between the Falling tone and the Checked tone. In Type 1, we used the same length for the Level tone and the High falling tone. Furthermore, the spectra of the Level tone are used for all tones, whereas the original lengths of speech unit are used for Type 2 and Type 3. The results of confusion matrices show that the lengths of tones are as important as F_0 for tone synthesis.
- Consequently, our proposed tone synthesis rule with linear pattern is sufficient and effective for Myanmar tone synthesis for both male and female speech.

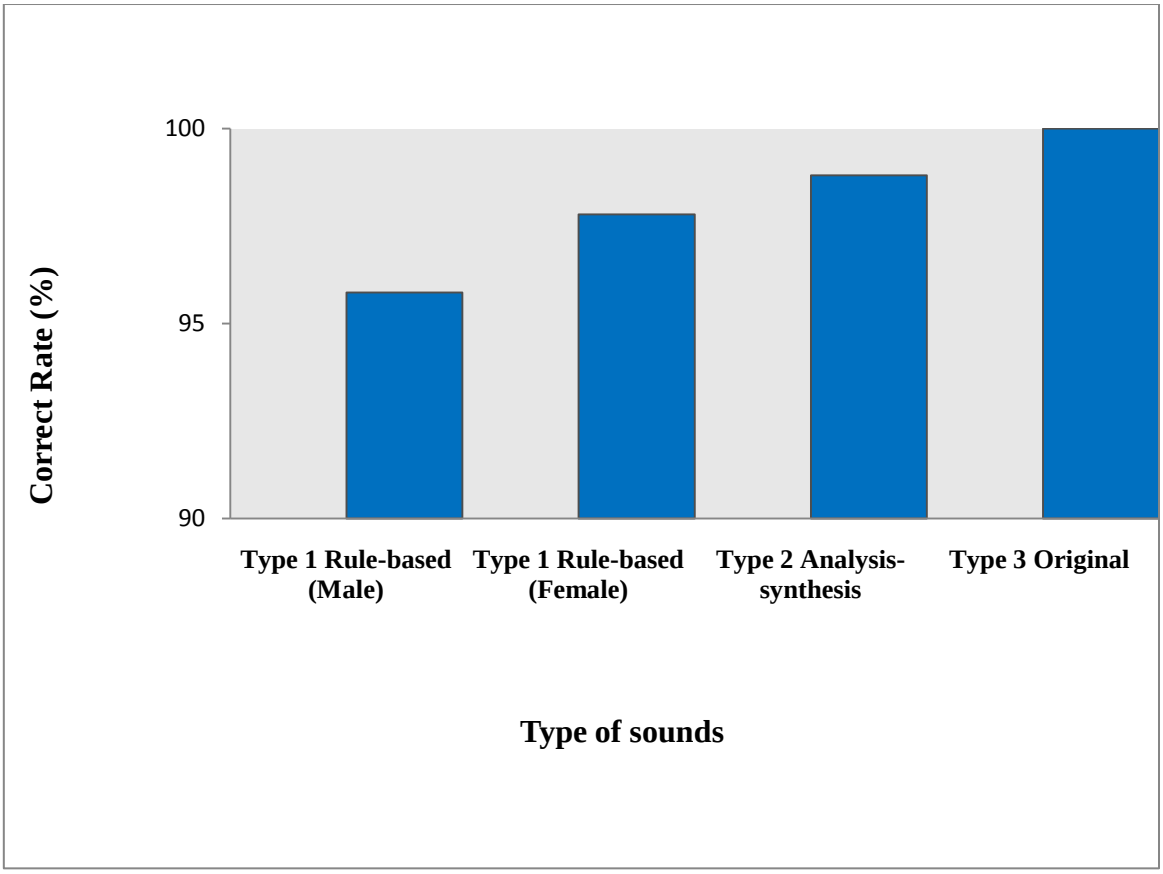


Fig. 4-7 The results of correct rate of perception of synthesized tone

Table 4.-3 Confusion matrices of tone synthesis (Unit: %)

Type 1 Rule-based					
Male		L	F	Hf	C
	L	96.8	0	2.6	0.6
	F	0.6	95.6	1.6	2.2
	Hf	8.5	0.5	91	0
	C	0.2	0	0	99.8
Female		L	F	Hf	C
	L	96.4	0	3.6	0
	F	0	97.5	0	2.5
	Hf	0.3	0	99.7	0
	C	0	0.3	0	99.7
Type 2 Analysis-Synthesis					
		L	F	Hf	C
	L	100	0	0	0
	F	0	95.6	0	4.4
	Hf	0	0	100	0
	C	0	0	0	100
Type 3 Original sounds					
		L	F	Hf	C
	L	100	0	0	0
	F	0	100	0	0
	Hf	0	0	100	0
	C	0	0	0	100

Chapter 5
Experimental Study on Optimization of Fundamental Frequency and Length of syllables

An optimization method to define the parameters; F_0 and syllable's length for tone synthesis is introduced. We implemented tone rules of linear pattern based on two parameters, the optimized F_0 at the center of syllable and the optimized syllable's length. The effectiveness of the proposed method is confirmed by distribution of tones and the intelligibility scores of listening test. Although the high intelligibility of synthesized tone draws reasonably high correct rates in former research, the proposed method achieve the better results. Furthermore, in the proposed method, the optimized parameters can be separated into male and female groups. The introduced proposed method is applicable for other tone synthesis rule of other tonal languages.

5.1 Tone Synthesis

The four Myanmar tones are analyzed to extract F_0 patterns. The data set is prepared as voiced sounds and meaningful words. We select consonant-vowel (CV) form with voiced consonants /b/, /m/, /l/ and three typical vowels /a/, /i/ and /u/. In total, 180 words (= 3 consonants x 3 vowels x 4 tones x 5 speakers) are used for tone analysis. After analyzing, four tones are distributed as shown in Fig.5-1. We find that the four tone

groups overlapped and are not clearly discriminated. In our former research, we normalized F_0 and length to obtain relative values among the tones. The normalized parameters of tones using one syllable word were plotted in the distribution [3]. In this research the normalized parameters by former normalization method using three syllables are shown in Fig.5-2 and Fig.5-3.

5.2 Optimization Method

Lagrange's optimization method [13-14] is used for normalization. In this study we use 36 words of F_0 patterns by utterance of five native speakers. The words include three typical vowels "a", "i" and "u" with voiced consonants "b", "m" and "i". We select F_0 from three frames at the center of syllable word for each tone. The average F0 values are selected from the middle frames of F_0 contours.

To minimize large differences of F_0 and differences of lengths among the speakers by means of tones, optimization method is carried out. The average of F0 contours for each tone is given by

$$f_{ij}=1/n\sum_{k=1}^{n} f_{ij}^{k} \qquad (5\text{-}1)$$

where n is number of F_0 contour. f_{ij} is F_0 at the center of syllable of i^{th} tone and j^{th} speaker. Similarly, the average of tones is defined as A_j and the average of all speakers is defined as A.

To normalize $f_{ij,}$ Lagrange's optimization technique is utilized in this paper. For convenience, we define $U_{ij}{}^{0}$ and R_{ij} such as

$$R_{ij} = A\text{-}A_j \qquad (5\text{-}2)$$

$$U_{ij}{}^{0} = f_{ij}{}^{0} - f_{ij} \tag{5-3}$$

where, $f_{ij}{}^{0}$ are normalized values of $f_{ij.}$

Then, in our problem, concentration of $f_{ij}{}^{0}$ around *A* i is accomplished by minimizing

$$W(f_{ij}^{0}) = \sum_{j=1}^{s} (A - f_{ij}^{0})^{2} \tag{5-4}$$

under the constraints

$$U_{ij}{}^{0} = \alpha_{ij} R_{ij} \tag{5-5}$$

where, αij are scale numbers and s is numbers of speaker.

Thus, normalized., $f_{ij}{}^{0}$ are given by minimizing Lagrange's function L (. $f_{ij}{}^{0}$)

$$\mathrm{L}\left(f_{ij}^{0}\right) = \mathrm{W}\left(f_{ij}^{0}\right) + \sum_{j=1}^{s} \lambda_j \left(U_{ij} - \alpha_{ij} R_{ij}\right) \tag{5-6}$$

For Eq. (5-6), we have

$$\frac{\partial \mathrm{L}}{\partial f_{ij}^{0}} = 2\left(f_{ij}^{0} - A\right) + \lambda_j = 0 \tag{5-7}$$

$$\frac{\partial \mathrm{L}}{\partial \lambda_j} = U_{ij} - \alpha_{ij} R_{ij} = 0 \tag{5-8}$$

Solving Eqs. (5-7), (5-8) gives

$$f_{\mathrm{ij}}{}^{0} = \boldsymbol{f_{ij}} + \boldsymbol{\alpha_{ij}} R_{ij} \tag{5-9}$$

$$\lambda_{\mathrm{j}} = 2(A - f_{ij} - \alpha_{ij} R_{ij}) \tag{5-10}$$

According to Eqs.(5-2) and (5-3)**,** equation (5-5) indicates that if α_{ij} =1, f_{ij} around A_j, i.e., f_{ij} - *Aj* is shifted to $\boldsymbol{f_{ij}^{0}}$ around A, i.e., $\boldsymbol{fij^{0}}$ –*A*, while α_{ij} = 0, i.e., $f_{ij}{}^{0}$ = *fij* which doesn't

give normalization. When male and female speakers intermix, average A behaves as a center of *Aj* for male and *Aj* for female.

On the other hand, the minimum value of L is derived as follows:

$$L_{min} = \sum_{j=1}^{s} (A - f_{ij} - \alpha_{ij}^{0} R_{ij})^2 \qquad (5\text{-}11)$$

which leads

$$\alpha_{ij}^{0} = (A - f_{ij}) / R_{ij} \qquad (5\text{-}12)$$

because $L_{min} \geq 0$.

$$(A - f_{ij}) / R_{ij} > 0 \qquad (5\text{-}13)$$

Hence, fij and Aj are always the same side of A.

Then, we have the relation

$$0 \leq \alpha ij \leq \alpha_{ij}^{0} \qquad (5\text{-}14)$$

From Eqs.(5-3) and (5-5),we get general equation

$$fij0 = fij + \alpha ij Ri \qquad (5\text{-}15)$$

For the sake of convenience, we may simply choose αij in this paper, such that

$$\alpha ij = \alpha = 1/2 \qquad (5\text{-}16)$$

In this way fij is normalized. The normalized value fij0 is given by,

$$fij0 = fij + \alpha Rij \qquad (5\text{-}17)$$

The optimized results are plotted in Fig. 5-5.and Fig.5-6. These figures show the distribution of four tones with optimized F_0 and optimized lengths, which are clearly discriminated in tone groups. From these figures we confirm that proposed method is an effective method to define the parameters for speech synthesis rule. Furthermore, as an advantage in the proposed method, the male and female can be distinguished.

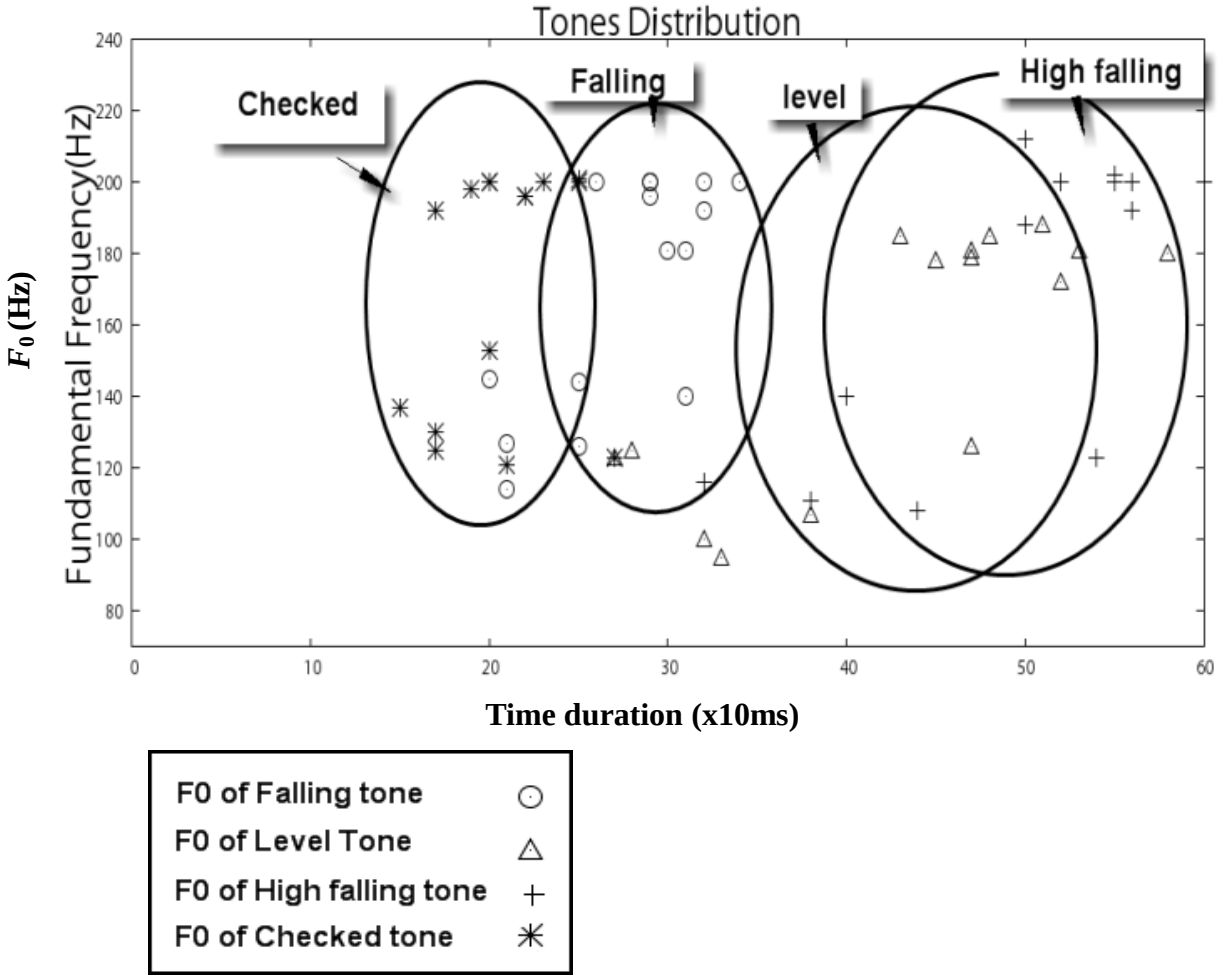


Fig.5-1 Tones distribution of analysis-synthesis sounds by three female speakers and two male speakers before optimization

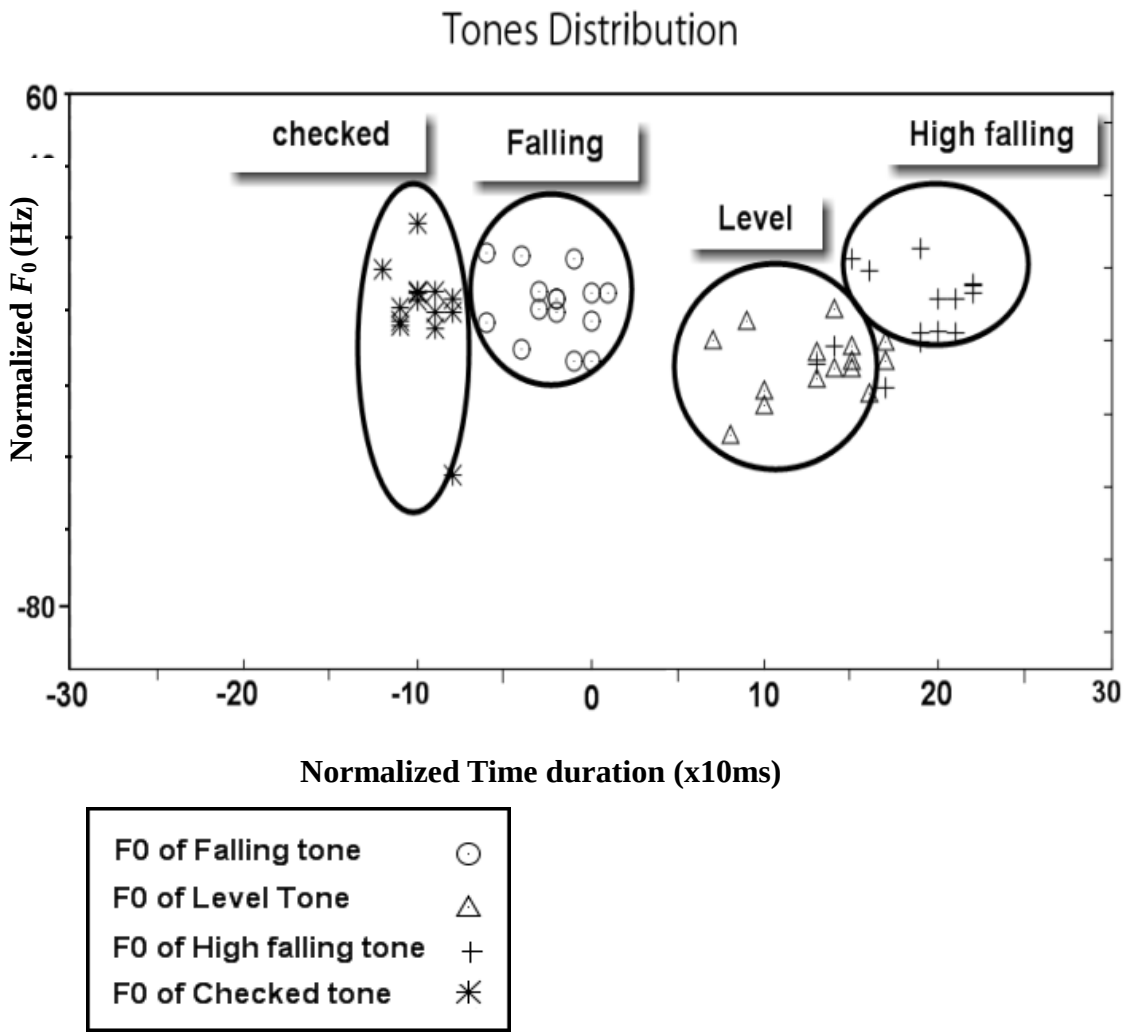


Fig.5-2. Tones distribution of analysis-synthesis sounds by three female speakers and two male speakers with normalized F_0 and normalized time (length)

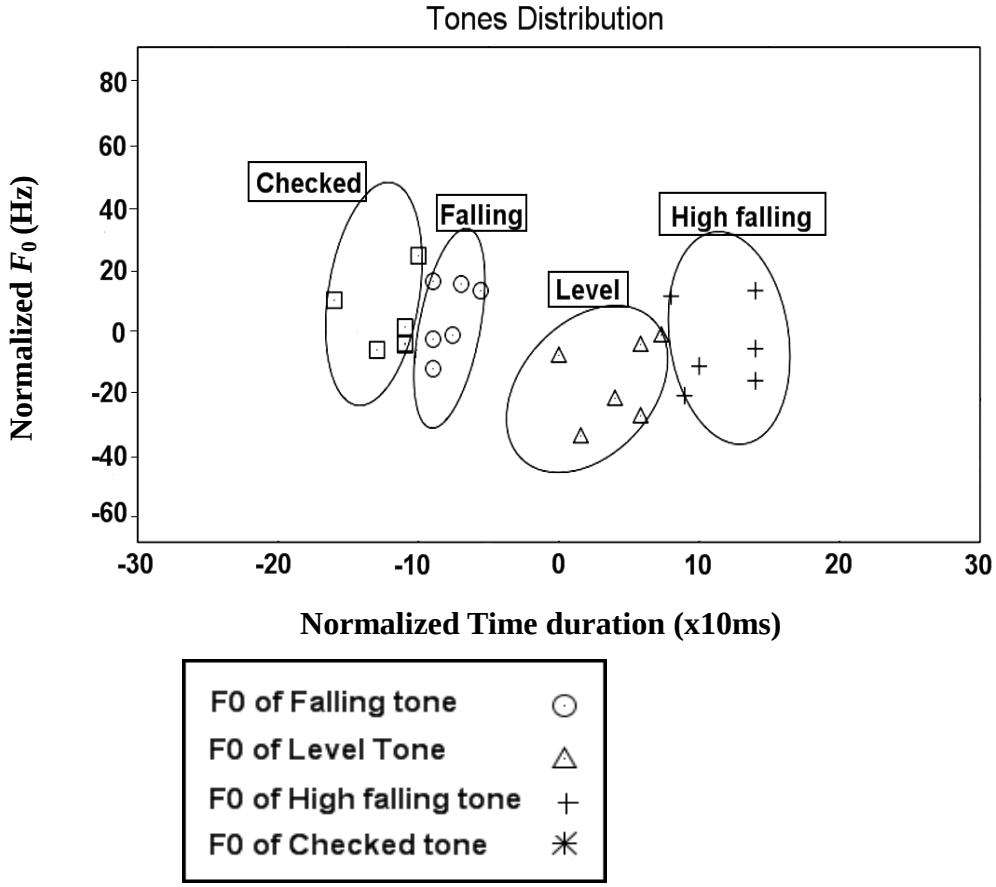


Fig.5-3. Tones distribution of analysis-synthesis sounds uttered by two male speakers with normalized F_0 normalized time (length)

5.3 Tone synthesis rule with linear F_0 pattern

Myanmar tones are unique in their simplistic pattern not only related to F_0 but also more specifically and importantly in terms of length. Myanmar tones have different lengths between short-tone and long-tone groups. In accordance, after optimization we define tone rule employing two parameters; F_0 at the center of syllables and syllable's length as opposed to focusing on length alone. Tone rules are constructed with linear F_0 patterns.

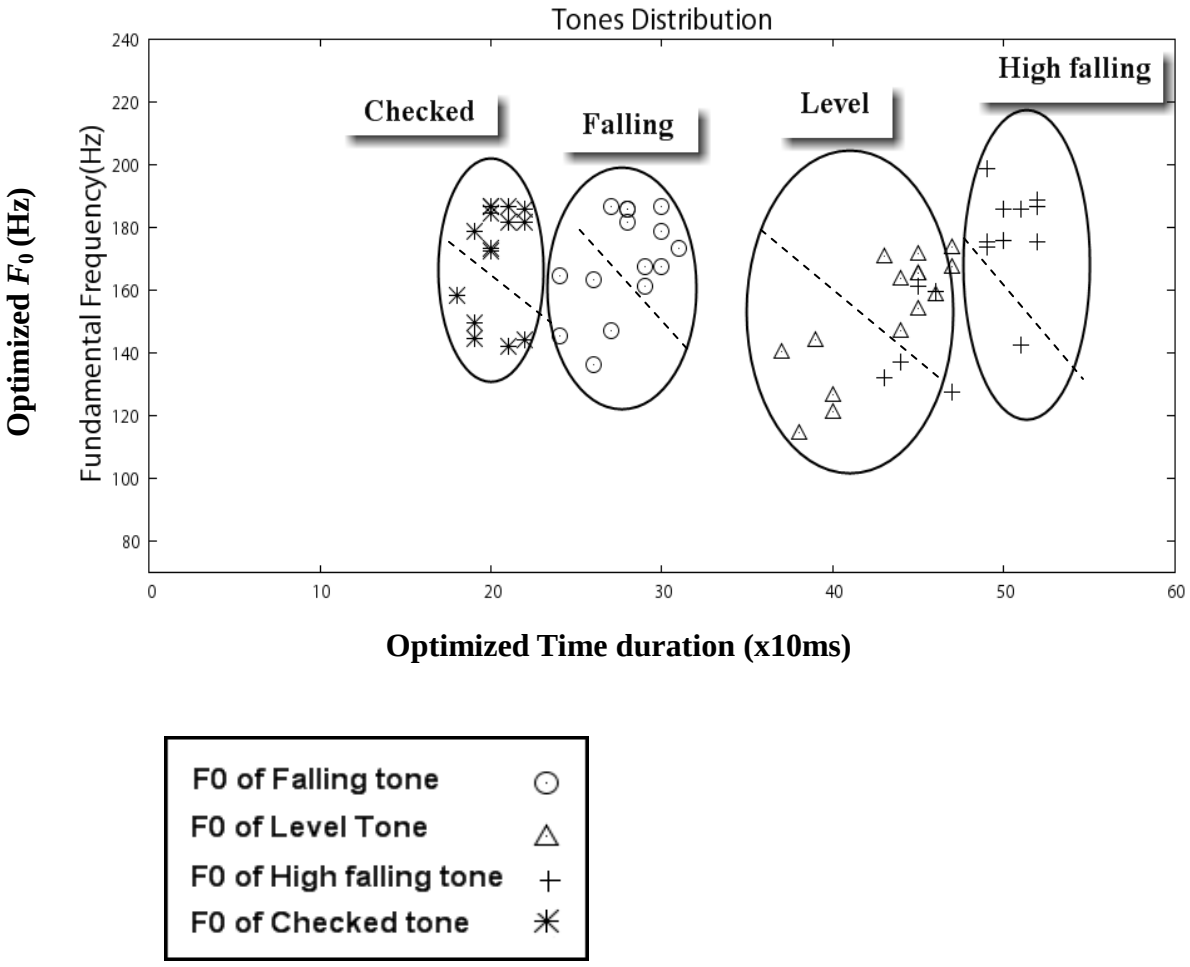


Fig.5 -4. Tones distribution by three female speakers and two male speakers with optimized F_0, and optimized length

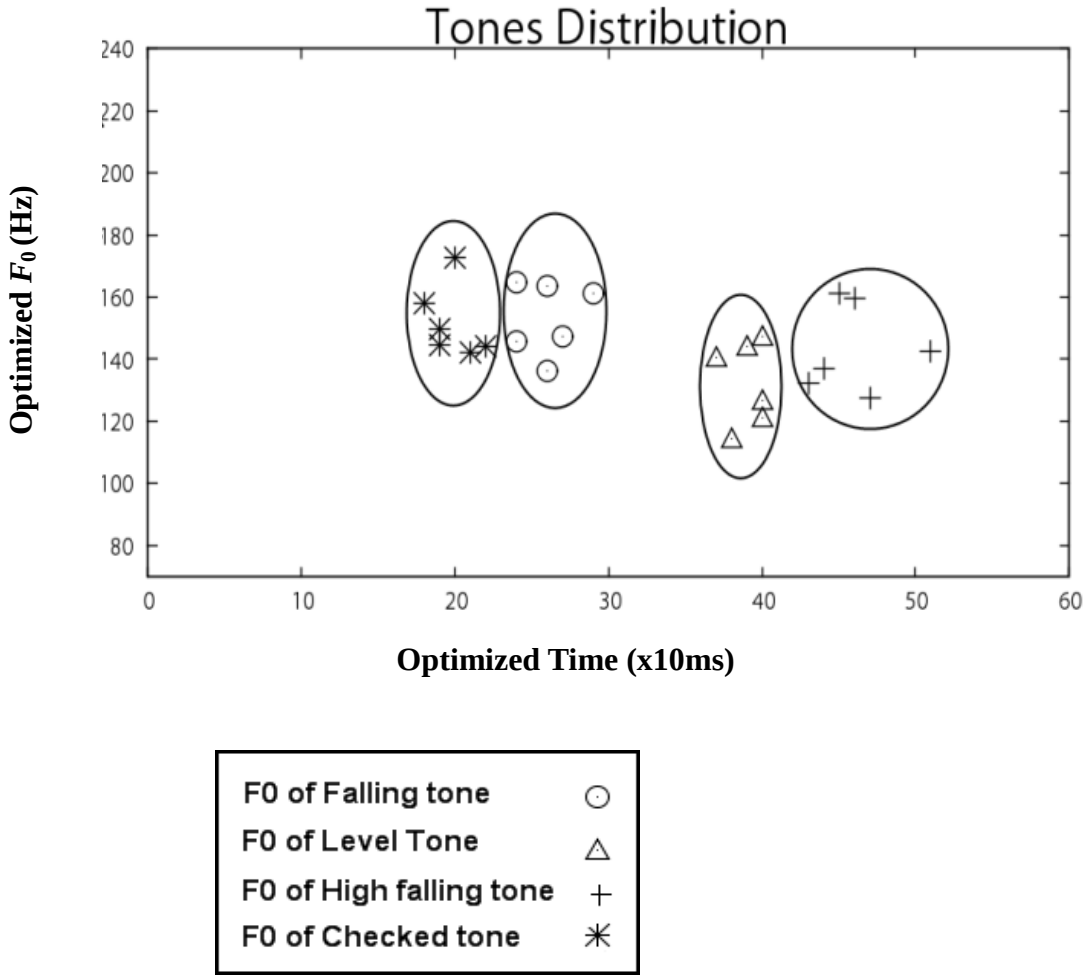


Fig.5-5. Tones distribution by two male speakers with optimized F_0, and optimized length

When we calculated the average frame length and average F_0 to make tone rules for male and female, we apply the concept of the center of gravity. As an example, Fig. 5-6 shows the calculation design of average F_0 and length using center of gravity. The tone rules are implemented based on optimized F_0 and optimized length of each tone as shown in Fig. 5-7.

We consider F_0 distribution as the mass distribution. We calculate average F_0 and length by using the concept of center of gravity x as follows:

$$x = (\sum_{i=1}^{n} x_i m_i)/M \qquad (5\text{-}18)$$

$M = m_1 + m_2 + m_3 + \ldots.. + m_n$

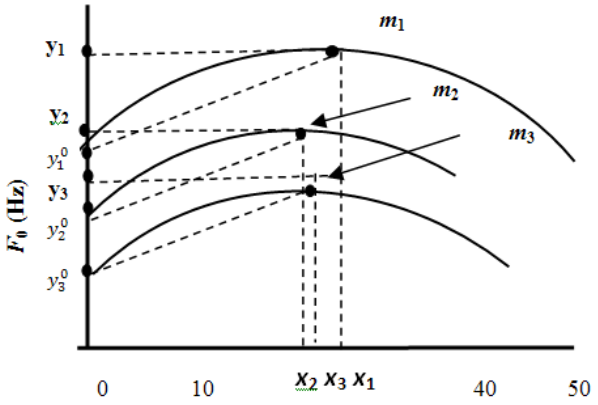


Fig.5-6 The calculation design of average F_0 and length

where m_i represents the weight of personal quality of F_0 of i^{th} speaker and x is average length of F_0 contour. Specifically, weight of personal quality of F_0 is different among the different speakers. As an example for three speakers,m_1,m_2 and m_3 are different values. In our experiments, all speakers are native and they have clear utterances and hearing ability. Therefore in this paper we consider their speech units have the same reliability. Then we have,

$m_1=m_2=m_3= m$ (*Example*: for three speakers)

From Eq. (16) average F_0 value at the center of contour y is calculated as

$$y = \frac{m(y_1 + y_2 + y_3)}{3m} = \frac{(y_1 + y_2 + y_3)}{3} \qquad (5\text{-}19)$$

Similarly the average length of time co-ordinate x is calculated as

$$x = \frac{(x_1 + x_2 + x_3)}{3} \qquad (5\text{-}20)$$

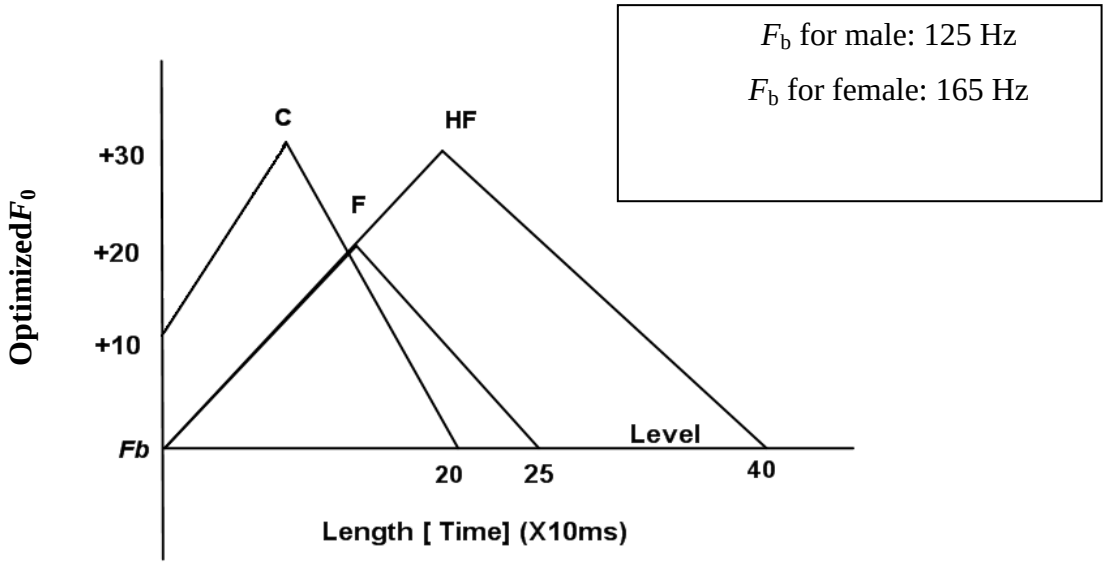


Fig. 5-7 The diagram of tone rule

L*: *Level tone, F*: *Falling tone, Hf*: *High falling tone, C: Checked tone,

Using these rules, we carried out the listening tests to evaluate intelligibilities of synthetic speech of syllables and to evaluate the effect of proposed method.

5.4 Results and discussion

Results of these tests are shown Fig. 5-8. These results have been obtained by using listening test. The result of our tone synthesis system and effectiveness of optimization are discussed as follows:

- Proposed method elicits the highest correct rate 99.68% for male speakers and 98.75% for female speakers.
- From these results we can confirm that optimized F_0 and length are conducted natural synthetic speech. Since we defined the scale factors of relative values properly, the optimized values are obtained.
- In VieTTS system[9], the result for linear pattern is about 85% for male, whereas the result of our system for male is 95.8%, even though our listening tests were done using the speech sounds of multiple speakers and different genders. Consequently, we can show that our linear pattern for tone rule is more effective than VieTTS's corresponding one since we applied the optimization method by means of multiple speakers and multiple phonemes.
- As a discussion concerning with above mentioned comparison, we consider that the optimization gives the effective values for both male and female, since we defined the scale factors of relative values correctly.
- Consequently, the introduced optimization method is effective and applicable for other speech synthesis rule for other tonal languages.

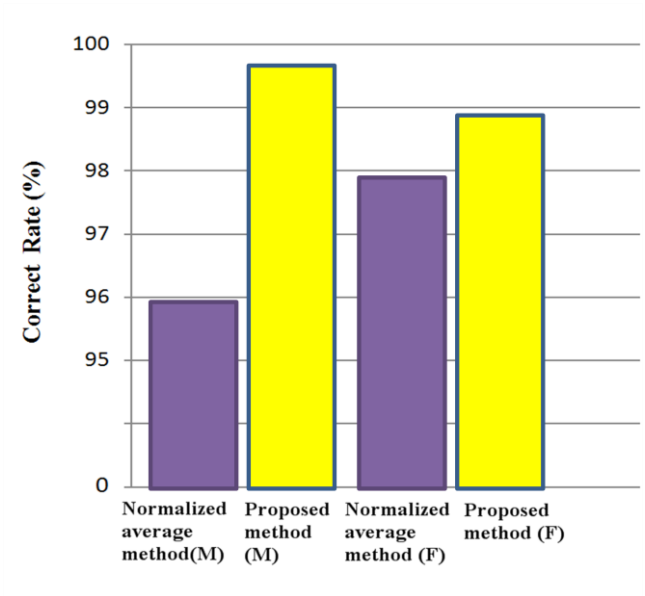


Fig.5-8. The results of correct rate of perception of synthesized tone

Chapter 6
Conclusion

In this chapter, we summarize the work that already expressed in the previous chapter and to present the limit and drawbacks of the approach, as well as perspective of the future work.

6.1 Summary

In this research we designed the first MyanmarTTS system with rule-based speech synthesis using cepstral parameter. The speech unit that is used in this system is demisyllable. The rule is implemented with linear F_0 pattern. Large variability exists in the F_0 and length uttered by different speakers in different syllables. Hence, for tone synthesis, normalization of the F_0 and length are important and necessary to discriminate tones. The effectiveness of this normalization method was confirmed in the distribution of the F_0 value and length. The high intelligibility of synthesized tone was confirmed through listening tests of synthesized words

In this dissertation, we presented with 6 chapters. The first chapter is Introduction. In this chapter, background, objective and chapter organization were explained. In chapter 2,

Myanmar language overview was introduced. In chapter 3, Myanmar speech analysis and synthesis system was presented.

In chapter 4, we described about MyanmarTTS system with rule-based speech synthesis. We implemented tone rules of linear pattern based on two parameters, the average value of F_0 at the center of syllable and syllable's length. We confirmed that our proposed tone synthesis rule is sufficient and effective for Myanmar tone synthesis for the speech of both male and female. Myanmar tones are unique in their simplistic pattern not

only related to F_0 but also more specifically and importantly to length. In view of this, normalization by means of average F_0 and average length are useful to access the reliable values for tone discrimination and tone synthesis. We showed that length is strongly related with tones as same as F_0 for Myanmar. The effectiveness of normalization is confirmed by distribution of tones and the intelligibility scores of listening test. The introduced normalization method is applicable for other tone synthesis rule of other tonal languages.

Chapter 5, focused on the study of tone synthesis. An optimization method to define the parameters; F_0 and syllable's length for tone synthesis is introduced. We implemented tone rules of linear pattern based on two parameters, the optimized F_0 at the center of syllable and the optimized syllable's length. The effectiveness of the proposed method is confirmed by distribution of tones and the intelligibility scores of listening test. Although the high intelligibility of synthesized tone draws reasonably high correct rates in former research, the proposed method achieve the better results. Furthermore, in the proposed method, the optimized parameters can be separated into male and female groups. The introduced proposed method is applicable for other tone synthesis rule of other tonal languages.

6.2 Future work

This research is the first step for Myanmar speech synthesis. There are, however, issues which still require investigations, such as the improvement of speech naturalness, modification of the variation of F_0 and length in continuous speech. An interesting theme that also remains is the adaptation of F_0 using spectra among speakers of different genders.

Acknowledgements

First of all I would like to express my gratitude to my advisor and also my supervisor, Professor. Tomio Takara, for his support, supervisions, patience and understanding.
Secondly, I am deeply grateful to my referees Professor Hayao Miyagi and Professor Mohammad Reza Asharif, for their helpful support and suggestions.

I sincerely thanks to all the professor of The Engineering Faculty for their encouragements, guidance and sharing their knowledge to me. I also want to express my grateful to all the staff of the office in the Engineering Faculty for their kindness and support to me.

I am grateful to all the staff of International student centre for their warmly support from the beginning of my study life in Okinawa, in especially Associated professor Kinjyo Kaori, for her invaluable kinds to me.

I also want to thank the Government (MONBUKAGAKUSHO) and the people of Japan, Scholarships Foundations and WFWP (Okinawa) who provided financial support for my study in Japan. I am grateful to Rotary Yoneyama Scholarship Foundation, Naha Rotary Club, for their trust and support to me during my study.

I also want to express my thanks to Professor Ishiguro Eiji, my supervisor and Nishioka sensei for their guidance and kind helps during my research period under the program of MONBUKAGAKUSHO in the Faculty of Education.

I would like to thanks Myanmar students, who help to me to record speech sound and to perform listening test. I want to express my heartfelt gratitude to my senior, Prof.Dr. Mar Mar Win(Myanmar). my friend Dr. Nina Lawrence and all of friends .

I'm grateful to all my laboratory members (2007 to 2011) for their friendship, especially Hiyane San, Miyagi San and .Nakamura Shogo San who provided and supported their knowledge to me when the beginning of my research on our lab.

Finally I would like to express my deepest gratitude to my Dad, my Mom my loving daughter and her Dad, all family members and all of my friends as well, who always encourage me to achieve my goals and cheer me up, for their trust, patience and support.
This Doctoral thesis is done at Graduate School of Engineering and Science, University of the Ryukyus, academic years 2010 – 2011.

Kyawt Yin Win
(ကျော်ယဉ်ဝင်း)
27th February, 2011

Bibliography

[1] S. Imai, "Log magnitude approximation (LMA) Filter," *Trans. IECE Jpn.*, **J63-A**, 886-893 (1980).

[2] S. Furui, "*Digital Speech Processing, Synthesis, and Recognition*," Second Edition, (Marcel Dekker, Inc., New York, 2001). pp. 30-31.

[3] P. Seresangtakul and T. Takara, "Analysis of pitch contour of Thai tone using Fujisaki's model," *Proc.ICASSP* '02, Vol **1**, 505-5089 (2002).

[4] Myanmar Language Committee, "Myanmar Grammar", Myanmar Language Committee, Ministry of Education, Myanmar, 2005.

[5] U. Thein Tun, "Some acoustic properties of tones in Burmese" In D. Bradley (ed), *Papers in South- East Asian Linguistics8: Tonation* Canberra: Australian National University, 77-116 (1982).

[6] J. K. Wheatley, "Burmese" In B. Comrie (ed), *The World's Major Languages* (Oxford University Press), New York,.834-845 (1987).

[7] T. Takara and T. Kochi, "General speech synthesis system for Japanese Ryukyu dialect," *Proc. 7th Westpac,* pp. 173-176 (2000).

[8] A. M. Noll, "Cestrum Pitch Determination", *J. Acoust. Soc. Am.*, **41**, No.2, 293-309 (1967).

[9] T. T. Do and T. Takara, "Vietnamese text-to-speech system with precise tone generation," *Acoust. Sci. & Tech.*, **25**, 5, 347-353 (2004).

[10] T. Takara "Experimental study on perception of the glottal explosive of the Japanese Ryukyu dialect," *Proc. Euro Speech '95*, pp. 953-956 (1995).

[11] Huang, J., Holt, L.L.: General Perceptual Contributions to Lexical tone normalization. *J. ACoust.Soc.Am.*125(6), June (2009).

[12] Zhang, S., Huang, T., Xu, B.,: Tone Modeling for Contious Mandarin Speech Recognition. INTERNATIONAL JOURNEL OF SPEECH TECHNOLOGY 7. 115-128 (2004).

[14] Xia, Y., Wang, J, A General Methology for Desiging Globally Convergent Optimization Neural Networks. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL.9.NO. 6, NOVEMBER, (1998)

[15] Deng, L., Shaughnessy, D. O.: Speech Processing A dynamic and Optimization-Oriented Approach. Marcel Dekker, Inc. New York (2003)

[16] S.Imai and Yoshiharu. ABE "Experimental Speech Synthesis from CV-syllable Cepstral Parameters," *Trans. IECE Jpn*., vol. **J64-D**, pp. 861-868, 1981.

[17] Ministry of Education, Myanmar, Elementary Myanmar, 2002.

Appendix A

Speech data of combination of CV units

		a		aa		i	
1	က	ka	ကာ	kaa	ကိ	ki	ကီ
2	ခ	kha	ခါ	khaa	ခိ	khi	ခီ
3	ဂ	ga	ဂါ	gaa	ဂိ	gi	ဂီ
4	င	nga	ငါ	ngaa	ငိ	ngi	ငီ
5	စ	sa	စါ	saa	စိ	si	စီ
6	ဆ	sha	ဆာ	shaa	ဆိ	shi	ဆီ
7	ဇ	za	ဇာ	zaa	ဇိ	zi	ဇီ
8	ည	nya	ညာ	nyaa	ညိ	nyi	ညီ
9	တ	ta	တာ	taa	တိ	ti	တီ
10	ထ	hta	ထာ	htaa	ထိ	hti	ထီ
11	ဒ၊ဓ	da	ဒါ၊ ဓါ	daa	ဒိ၊ ဓိ	di	ဒီ၊ ဓီ
12	န၊ ဏ	na	နာ၊ ဏာ	naa	နိ၊ ဏိ	ni	နီ၊ ဏီ
13	ပ	pa	ပါ	paa	ပိ	pi	ပီ
14	ဖ	hpa	ဖါ	hpaa	ဖိ	hpi	ဖီ
15	ဗ၊ဘ	ba	ဗာ၊ ဘာ	baa	ဗိ၊ ဘိ	bi	ဗီ၊ဘီ
16	မ	ma	မာ	maa	မိ	mi	မီ
17	ယ	ya	ယာ	yaa	ယိ	yi	ယီ
18	ရ	ra	ရာ	raa	ရိ	ri	ရီ
19	လ	la	လာ	laa	လိ	li	လီ
20	ဝ	wa	ဝါ	waa	ဝိ	wi	ဝီ
21	သ	tha	သာ	thaa	သိ	thai	သီ
22	ဟ	ha	ဟာ	haa	ဟိ	hi	ဟီ
23	အ	a	အာ	aa	အိ	i	အီ
24	ကျ၊ ကြ	ca	ကျာ၊ ကြာ	caa	ကျိ၊ ကြိ	ci	ကျီ၊ကြီ
25	ချ၊ ခြ	cha	ချာ၊ ခြာ	chaa	ချိ၊ ခြိ	chi	ချီ၊ ခြီ
26	ဂျ၊ ဂြ	gya	ဂျာ၊ ဂြာ	gyaa	ဂျိ၊ ဂြိ	gyi	ဂျီ၊ ဂြီ
27	ပျ၊ ပြ	pya	ပျာ၊ ပြာ	pyaa	ပျိ၊ ပြိ	pyi	ပျီ၊ ပြီ
28	ဖျ၊ ဖြ	hpya	ဖျာ၊ ဖြာ	hpyaa	ဖျိ၊ ဖြိ	hpyi	ဖျီ၊ ဖြီ
29	ဗျ၊ ဗြ	mya	ဗျာ၊ ဗြာ	myaa	ဗျိ၊ ဗြိ	myi	ဗျီ၊ ဗြီ
30	မျ၊ မြ	bya	မျာ၊ မြာ	byaa	မျိ၊ မြိ	byi	မျီ၊ မြီ
31	လျ	lya	လျာ	lyaa	လျိ	lyi	လျီ
	u.		uu		e		ဧ
ကု	ku.	ကူ	kuu	ကေ	kei	ကဲ	ke
ခု	khu.	ခူ	khuu	ခေ	khei	ခဲ	khe
ဂု	gu.	ဂူ	guu	ဂေ	gei	ဂဲ	ge

ဍု	ngu.	ဍူ	nguu	ဍေ	ngei	ဍဲ	nge
စု	su.	စူ	suu	စေ	sei	စဲ	se
ဆု	shu	ဆူ	shuu	ဆေ	shei	ဆဲ	she
ဇု	zu.	ဇူ	zuu	ဇေ	zei	ဇဲ	ze
ညု	nyu.	ညူ	nyuu	ညေ	nuei	ညဲ	nye
တု	tu.	တူ	tuu	တေ	tei	တဲ	te
ထု	htu	ထူ	htuu	ထေ	htei	ထဲ	the
ဒု၊ ဓု	du.	ဒူ၊ ဓူ	duu	ဒေ၊ ဓေ	dei	ဒဲ၊ ဓဲ	de
နု၊ ဏု	nu.	နူ၊ ဏူ	nuu	နေ၊ ဏေ	nei	နဲ၊ ဏဲ	ne
ပု	pu.	ပူ	puu	ပေ	pei	ပဲ	pe
ဖု	hpu.	ဖူ	hpuu	ဖေ	hpei	ဖဲ	hpe
ဗု၊ ဘု	bu.	ဗူ၊ ဘူ	buu	ဗေ၊ ဘေ	bei	ဗဲ၊ ဘဲ	be
မု	mu.	မူ	muu	မေ	mei	မဲ	me
ယု	yu.	ယူ	yuu	ယေ	yei	ယဲ	ye
ရု	ru.	ရူ	ruu	ရေ	rei	ရဲ	re
လု	lu.	လူ	lu	လေ	lei	လဲ	le
ဝု	wu.	ဝူ	wu	ဝေ	wei	ဝဲ	we
သု	thu.	သူ	thuu	သေ	thei	သဲ	the
ဟု	hu.	ဟူ	huu	ဟေ	hei	ဟဲ	he
အု	u	အူ	uu	အေ	ei	အဲ	e
ကျု၊ ကြု	cu	ကျူ၊ ကြူ	cuu	ကျေ၊ ကြေ	cei	ကျဲ၊ ကြဲ	ce
ချု၊ ခြု	chu.	ချူ၊ ခြူ	chuu	ချေ၊ ခြေ	chei	ချဲ၊ ခြဲ	che
ဂျု၊ ဂြု	gyu.	ဂျူ၊ ဂြူ	gyuu	ဂျေ၊ ဂြေ	gyei	ဂျဲ၊ ဂြဲ	gye
ပျု၊ ပြု	pyu.	ပျူ၊ ပြူ	pyuu	ပျေ၊ ပြေ	pyei	ပျဲ၊ ပြဲ	pye
ဖျု၊ ဖြု	hpyu.	ဖျူ၊ ဖြူ	hpyuu	ဖျေ၊ ဖြေ	hpuei	ဖျဲ၊ ဖြဲ	hpye
မျု၊ မြု	myu.	မျူ၊ မြူ	myuu	မျေ၊ မြေ	myei	မျဲ၊ မြဲ	mye
ဗျု၊ ဗြု	byu.	ဗျူ၊ ဗြူ	byuu	ဗျေ၊ ဗြေ	byei	ဗျဲ၊ ဗြဲ	bye
လျု	lyu.	လျူ	lyuu	လျေ	lyei	လျဲ	lye

	au		ɔ		an(ã)		o
ကော	kau	ကော်	kaw	ကံ	kan	ကို	ko
ခေါ	khau	ခေါ်	khaw	ခံ	khan	ခို	kho
ဂေါ	gau	ဂေါ်	gaw	ဂံ	gan	ဂို	go
ငေါ	ngau	ငေါ်	ngaw	ငံ	ngan	ငို	ngo
စော	sau	စော်	saw	စံ	san	စို	so
ဆော	shau	ဆော်	shaw	ဆံ	shan	ဆို	sho
ဇော	zau	ဇော်	zaw	ဇံ	zan	ဇို	zo
ညော	nyau	ညော်	nyaw	ညံ	nyan	ညို	nyo
တော	tau	တော်	taw	တံ	tan	တို	to
ထော	htau	ထော်	htaw	ထံ	htan	ထို	hto
ဒေါ၊ ဓော	dau	ဒေါ်၊ ဓော်	daw	ဒံ၊ ဓံ	dan	ဒို၊ ဓို	do
နော၊ဏော	nau	နော်၊ဏော်	naw	နံ၊ ဏံ	nan	နို၊ ဏို	no
ပေါ	pau	ပေါ်	paw	ပံ	pan	ပို	po
ဖေါ	hpau	ဖေါ်	hpaw	ဖံ	hpan	ဖို	hpo
ဗော၊ဘော	bau	ဗော်၊ဘော်	baw	ဗံ ၊ဘံ	baw	ဗို ၊ဘို	bo
မော	mau	မော်	maw	မံ	man	မို	mo
ယော	yau	ယော်	yaw	ယံ	yan	ယို	yo
ရော	rau	ရော်	raw	ရံ	ran	ရို	ro
လော	lau	လော်	law	လံ	lan	လို	lo
ဝေါ	wau	ဝေါ်	waw	ဝံ	wan	ဝို	wo
သော	thau	သော်	thaw	သံ	than	သို	tho
ဟော	hau	ဟော်	haw	ဟံ	han	ဟို	ho
အော	au	အော်	aw	အံ	an	အို	o
ကျော ၊ ကြော	cau	ကျော် ၊ ကြော်	caw	ကျံ ၊ ကြံ	can	ကျို ၊ ကြို	co
ချော ၊ ခြော	chau	ချော် ၊ ခြော်	chaw	ချံ ၊ ခြံ	chan	ချို ၊ ခြို	cho
ဂျော ၊ ဂြော	gyau	ဂျော် ၊ ဂြော်	gyaw	ဂျံ ၊ ဂြံ	gyan	ဂျို ၊ ဂြို	gyo
ပျော ၊ ပြော	pyau	ပျော် ၊ ပြော်	pyaw	ပျံ ၊ ပြံ	pyan	ပျို ၊ ပြို	pyo
ဖျော ၊ ဖြော	hpyau	ဖျော် ၊ ဖြော်	hpyaw	ဖျံ ၊ ဖြံ	hpyan	ဖျို ၊ ဖြို	hpyo
မျော ၊ မြော	myau	မျော် ၊ မြော်	myaw	မျံ ၊ မြံ	myan	မျို ၊ မြို	myo
ဗျော ၊ ဗြော	byau	ဗျော် ၊ ဗြော်	byaw	ဗျံ ၊ ဗြံ	byan	ဗျို ၊ ဗြို	byo
လျော	lyau	လျော်	lyaw	လျံ	lyan	လျို	lyo

Appendix B

List of words used in intelligibility test for Myanmar tone

	Level	Falling	High falling	Checked
1.	ba	bà	bá	baʔ
2.	bi	bì	bí	biʔ
3.	bu	bu`	bu´	buʔ
4.	ma	mà	má	maʔ
5.	mi	mì	mi´	miʔ
6.	mu	mu`	mu	muʔ
7.	la	là	lá	laʔ
8.	li	lì	li´	liʔ
9.	lu	lu`	lu´	luʔ