With respect to the pattern matching process, hidden Markov models (HMMs) have proven to be an effective statistical approach to isolated tone recognition [14,15]. However, tone recognition in Thai connected speech using HMMs has never been attempted. We believe that a simple straightforward extension of an HMM isolated tone recognition algorithm is likely to produce unsatisfactory results for connected speech tone classification. This is partly due to the fact that connected speech tone recognition is a more difficult problem than isolated tone recognition. As illustrated in figure 1.3.2, there are differences in the F_0 realization of tones in an utterance when each individual word is spoken in isolation (see top panel) and when the whole utterance is naturally spoken in connected speech (see bottom panel). There appear to be interactions among several linguistic factors that affect the F_0 realization of tones in connected speech: syllable structure, tonal coarticulation, stress, and intonation.

Analogous to the problem of continuous speech phone recognition in which contextual variations between contiguous phones (i.e., phone coarticulation) must be taken into account, continuous speech tone recognition must also incorporate tonal coarticulation and other linguistic factors into the system. A simple modification of an HMM isolated tone recognizer to recognize tones in continuous speech requires constructing a maximum of 125 (5 previous X 5 current X 5 following tones in a three-tone analysis window) tone models in order to account for both perseverative and anticipatory tonal coarticulation. This model may not conducive to real-time applications even with a parallel implementation. Also, because of subtle changes in F_0 contours due to coarticulatory effects, the usual acoustic features, F_0 and ΔF_0, used in an HMM-based system may not adequately capture the acoustically discriminatory information among coarticulation patterns of tones. For these reasons and because tonal coarticulation appears to be rule-governed, we propose a novel algorithm to classify tones in connected speech using an *analysis-by-synthesis* model.

Analysis-by-Synthesis is an abstract model of the speech perception process proposed by Stevens [39]. The basic assumption of the model is that speech perception and production are closely tied. The major claim of the theory is that listeners perceive (analyze) speech by implicitly generating (synthesizing) speech from what they have heard and then comparing the synthesized speech with the auditory stimulus. According to the model, the perceptual process begins with an analysis of auditory features of the speech signal to yield an acoustic description in terms of auditory patterns. A hypothesis (or hypotheses) concerning the distinctive feature representation of the utterance is (are)

constructed. This information then becomes the input to a set of generative rules that synthesize candidate patterns. The candidate patterns are subsequently compared with the patterns of the original utterance. The results of this matching process are then sent to a control component that transfers the phonetic description to higher levels of linguistic analysis. This model represents one of many *bottom-up* approaches to speech perception. That is, the model does not incorporate the effects of lexical and other higher-level knowledge into the speech perception process; they are only considered during later stages of recognition.

We adopt this model in the development of a Thai connected speech tone classifier because the model is easily implemented in terms of incorporating linguistic constraints into the model, although there has been little empirical evidence to support its validity. As the name suggests, the model contains two major components: the analysis and the synthesis module. Roughly speaking, the function of the analysis module is to generate hypothesized tone sequences from the input F_0 contour. The synthesis module, in turn, generates predicted F_0 contours according to the hypothesized tone sequences. These predicted F_0 contours are basically reference templates to be used for pattern matching against the input contour. The synthesis module is based on our extension of Fujisaki's model for synthesizing F_0 contours to tone languages, and linguistic constraints are represented as synthesis rules in the form of the Fujisaki's model parameters. In this research, every factors affecting F_0 realization of tones in Thai connected speech, i.e., continuity effect, stress, tonal coarticulation and declination, have been accounted for.

3.3.2 The proposed tone classification algorithm

In this section, details of the proposed automatic tone classification algorithm based on the analysis-by-synthesis method are presented. The algorithm takes into account all factors affecting phonetic realization of Thai tones as previously mentioned. Also discussed are important considerations for the normalization procedures to achieve speaker-independence.

The general design of the algorithm involves steps as shown in Fig. 3.3.2. The first three blocks represent the pre-processing of the speech signal to extract relevant information or acoustic features for subsequent classification. These are steps necessary to produce relatively reliable, normalized F_0 contours. The last three blocks represent the

tone classification step based on the analysis-by-synthesis method. Each component of the system is described in detail below.

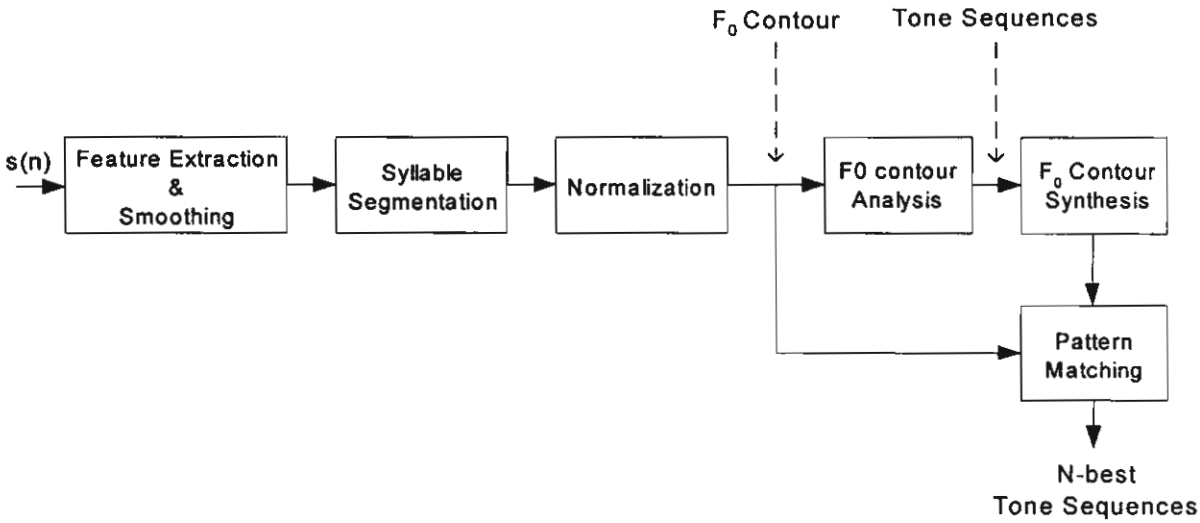

Fig. 3.3.2 The block diagram of the proposed tone classifier

Feature extraction and smoothing

Two suprasegmental features corresponding to acoustic correlates of prosody (F_0 and intensity) are extracted from the speech input. First, the raw F_0 contour is automatically extracted from the input speech signal using one of several methods to pitch extraction. Our implementation of the tone classifier relies on a CSL pitch extraction algorithm which employs a time domain approach to pitch analysis (modified autocorrelation with center clipping) with nonoverlapping variable frame length. For a particular speaker, frame length will be determined by his/her pitch range to ensure that there were at least two complete cycles within a frame. A typical frame length is 20 to 25 ms for male speakers, 15 to 20 ms for female speakers. To eliminate "drop-outs" during voiced speech segments, spurious pitch values in regions of unvoiced speech segments, and/or "double pulsing" effect, smoothing techniques, such as median filtering and linear interpolation, must be employed. In this experiment, the F_0 contours were smoothed using linear interpolation technique.

Secondly, the energy (intensity) measure will be used in placed of the amplitude measure of the speech signal since they are closely related. Energy calculation in decibels (dB) will be performed in a nonoverlapping frame-by-frame, pitch asynchronous manner using a Multi-speech algorithm that defines energy as the sum of

the square of absolute amplitude values within a frame. Frame length will be kept constant at 20 ms for all speakers. The raw energy value will be converted into dB by computing 20 times the log (base 10) of the square root of the ratio between the energy to the number of samples in the frame. A smoothing function was applied to the resulting energy contour.

The energy contours obtained above will be used to crudely identify syllables with CVS structure (i.e., syllables ending with stop consonant, /p/, /t/, and /k/). This is important in determining the rhythmic grouping of the input utterance. Since these coda consonants are glottalized, the syllable ends abruptly and the signal energy decreases very rapidly at the end of the syllable. This rapid energy drop results mainly from the articulatory requirement of the final stop consonant. A syllable ending with a stop consonant will cease abruptly even if the voiced portion preceding the stop consonant has been prolonged. To parameterize this characteristic, a smoothed short-time energy profile $E_S(j)$ is obtained for the voiced portion of the syllable using the above-described procedure. Let $j_{\max}$ denote the frame number in which maximum energy occurs and t_d be the time required for the energy to drop from 90% to 10% of $E_S(j_{\max})$. We can define an energy drop rate as the reciprocal of t_d. That is, $R_D = \frac{1}{t_d}$. It should be noted that the energy drop rate are highly correlated with the syllable duration. The shorter the duration, the faster the energy drops.

Syllable segmentation

Since tones are properties of syllables, it is logical to segment the smoothed and normalized F_0 contour into syllabic units. Syllable boundary information can be provided by an automatic syllable segmentation algorithm based on energy contours and spectral information, or by segmentation information from a phone recognizer unit. In this study, we have developed an automatic procedure for syllable segmentation. Automatic syllable segmentation is a crucial component that provides syllable boundary information necessary for our tone classification system. Traditionally, zero crossing rate and root-mean-square energy (RMSE) of the speech signal are the two most widely used features for locating syllable boundary. In this research, we propose a new segmentation algorithm based on a modified Teager's energy calculation [40]. We present details of the algorithm below.

The most common way of calculating the energy of a speech signal is the root mean square energy (RMSE), which is the square root of the average of the sum of the squares of the amplitude of the signal samples. Using a window of width W to segment the speech signal into frames, the RMSE of frame n, E_n, is given by:

$$E_n = \left[\frac{1}{W}\sum_{i=1}^{W} s_n^2(i)\right]^{\frac{1}{2}},$$

where $s_n(i)$ denote the i^{th} windowed speech sample in frame number n.

On the other hand, in modeling speech production, Teager developed a new algorithm for computing the energy of a signal. This algorithm has been presented by Kaiser as Teager's Energy Algorithm. Given a signal with the motion of an oscillatory body, its sample is defined as

$$x_i = A\cos(\Omega i + \phi),$$

where A is the amplitude of the oscillation, Ω is the digital frequency, and ϕ is the initial phase. In Teager's Algorithm, the instantaneous energy E_i of the sample x_i is as follows:

$$\begin{aligned} E_i &= x_i^2 - x_{i+1}x_{i-1} \\ &= A^2\sin^2(\Omega) \\ &\approx A^2\Omega^2 \end{aligned}$$

It is noted that the output of Teager's Algorithm is a function of the amplitude of the signal samples, as well as the oscillation frequency. This new energy measure is therefore capable of responding rapidly to the changes in both A and Ω. Thus, it has the ability to track rapid changes as well as the qualitatively different character of various signals.

The fact that the Teager energy algorithm reflects both the amplitude and frequency of a signal suggests that it may be a more suitable measure for different speech events than the RMSE, which reflects only the amplitude of the signal. From the point of view of speech production, the amount of energy used to produce noise-like fricatives should not be an order of magnitude less than that used to produce periodic voiced sounds. Yet, this is the typical difference we often get when using RMSE measure. Fricatives and plosives sounds have very low amplitude, but, unlike most vowels, these sounds have energy distributed in the frequency range above 5 kHz. As a result, Teager's energy measure should be more suitable for the calculation of the energy used in producing those fricatives and plosives.

To apply Teager's energy calculation to the problem of speech segmentation, we observe that the expression for the instantaneous energy can be related to the square of the samples of the derivative signal. This is equivalent to calculating the RMSE on the derivative of the speech sample x_i. The result is proportional to A^2 and Ω^2 as in Teager's energy calculation. As a result, we propose a new energy calculation based on a modification to Teager's calculation as follows:

1. Calculate the power spectrum of the speech signal;
2. Weight each sample in the power spectrum with the square of the frequency;
3. Take the square root of the sum of the weighted power spectrum.

Based on the above energy calculation, our syllable segmentation algorithm have been evaluated using the speech materials described in appendix C. To evaluate performance, we visually compare the estimated locations of syllable boundary using the different energy measures (both RMSE and Teager's). Zero crossing rate is also computed and used to aid our visual inspection of the correct boundaries. The detected boundaries are compared with those obtained from manual segmentation via audio playback of the speech signals selected between the detected boundaries.

Preliminary results are encouraging revealing several general properties of this new energy calculation. First, the new measure confirms a higher energy level for fricatives and plosives than that obtained form RMSE measure. Secondly, compared to RMSE, the new measure decreases the energy difference between voiced and voiceless sounds. Lastly, The new measure suppresses the energy level of background noise during silence intervals.

In addition to syllable boundary information obtained above, we also extract the durational patterns of every syllable in the utterance. Based on our automatic syllable segmentation algorithm above, syllable duration is computed. Note that syllable duration for our purpose is defined as the duration D of the voiced portion of a syllable only. This durational information will be used in discriminating between stress and unstressed syllables in the input utterance. For the purpose of computing the speaking rate, total duration marked by the beginning and end of the utterance is also calculated. The total duration of the target sentence will be measured from the onset of the consonant at the beginning of the sentence to the cessation of the coda consonant (closed syllable) or vowel (open syllable) of the last syllable at the end of the sentence. Speaking rate will then be computed by dividing the total sentence duration by the number of syllables in

that sentence. The speaking rate will be used in the normalization process, which will be described next.

Normalization

Normalization of the feature parameters is necessary because it will eliminate undesirable time and speaker variations of these parameters. In terms of pitch, for a multiple-speaker system, the normalization process is introduced to neutralize variability from one F_0 contour to the next. Sources of variability include speaker's physiological differences, the kinetics of vocal fold vibration, consonantal perturbations on F_0, and speaking rate. The raw F_0 contour is first converted into an equivalent-rectangular-bandwidth-rate (ERB) scale. This ERB normalization has an effect of neutralizing pitch ranges of different excursion size. To neutralize the declination effect in the F_0 contour, we subtract a time-varying mean F_0 value from the input F_0 contour. A time-varying mean F_0 value is computed by fitting an exponential curve to the overall contour as already discussed. Then, z-score normalization is employed to account for pitch range differences across speakers based on the precomputed mean and standard deviation from all utterances in the training set. This method has the effect of making the first- and second-order moments of the pitch distributions the same.

For the duration-related parameters D and R_D, normalization is needed. The speaking rate can be affected by emotional, stylistic and environmental factors, which may change from time to time. For example, the duration of a long syllable can be very short for fast speaking persons. The normalization factors are the precomputed means from all utterances in the training set.

F_0 contour analysis

This step is necessary to reduce the number of possible reference templates that have to be generated by the synthesis module, and thus, reduce the amount of time it takes to match against the input F_0 contour. The analysis procedure consists mainly of two steps. First, using the syllable durational patterns, a rhythm grouping among adjacent syllables is determined from the rules given in section 3.2.3 and repeated here for convenient. That is, the relative syllable duration for each type of rhythmic foot can be abstractly described together with the corresponding rule for matching the acoustic realization of a rhythmic foot with the abstract description as follows:

1. For a one-syllable rhythmic foot in an utterance-initial position,

| S | → | 3 | → | 2 |

2. For a one-syllable rhythmic foot in an utterance-final position with a non CVS structure, | S | → | 3 | → | 4 |

3. For a two-syllable rhythmic foot in which the salient syllable has a CVS structure, or the weak syllable is the first element of a compound that does not have a CVS structure; or both the salient syllable and the weak syllable are function words,

| S W | → | 2 : 1 | → | 2 : 2 |

4. For a three-syllable rhythmic foot in which the salient syllable has a CVS structure, or it is in an utterance-initial position, or it is a function word and the two weak syllables are two function words or a function word and a linker syllable.

| S W W | → | 1½ : ¾ ¾ | → | 1⅔ : 1⅔ 1⅔ |

Once the rhythmic grouping is determined, the second step involves the peak-and-valley analysis, i.e., the detection of local extrema of the given smoothed, normalized and segmented F_0 contour for that grouping. Local extrema (peaks and valleys) are detected by using first and second derivatives. The derivative at any point in the contour, except for the first two and last two points, is computed by calculating the linear regression coefficients of a group of five F_0 values consisting of the current point, and its preceding and following two points.

The locations of these extrema coupled with syllable boundary information and the energy drop rate are then used to identify all possible tone labels for the salient syllable in the rhythmic grouping based on some specified rules. For example, between two syllable boundaries, only the falling tone can occur if a maximum occurs, and only the rising or the high tone can occur if a minimum occurs. Also, if a maximum occurs at or in the vicinity of a syllable boundary, the preceding tone can either be a high or a rising tone. If a minimum occurs at or in the vicinity of a syllable boundary, the preceding tone can either be a mid or a low tone, or a sequence of two falling tones. For the rest of the weak or unstressed syllables within the given grouping, only three tonal labels (FH, M, and LR) are assigned depending on the overall temporal pitch variation. The FH label indicates an upward trend, the LR a downward tend, and M a level trend. These labels are derived based on the information obtained from the acoustic experiments described in chapter 2. They reflect the fact that unstressed syllables suffer tone neutralization, and the contrastive pattern among tones can be divided into roughly three tonal registers.

To deal with syllables with different duration, a time-aligned pitch profile is used [41]. The voiced portion of the syllable is divided evenly into 16 segments. For each segment, a pitch value is obtained from the given F_0 contours using a linear interpolation method. Thus, the pitch profile of each syllable has the same dimension of 16. Given a pitch profile $\{P(1), P(2), \ldots\ldots, P(i), \ldots, P(16)\}$, the overall temporal pitch variation within the profile can be measured using a pitching rising index, I_R, which is defined as

$$I_R = k \cdot \frac{\mathrm{Max}_{i=2}^{15}\{P(i)\} - \mathrm{Min}_{i=2}^{15}\{P(i)\}}{\mathrm{Max}_{i=2}^{15}\{P(i)\} - \mathrm{Min}_{i=2}^{15}\{P(i)\}}$$

$$\text{where } k = \begin{cases} 1 & \arg\mathrm{Max}_{i=2}^{15}\{P(i)\} > \arg\mathrm{Min}_{i=2}^{15}\{P(i)\} \\ -1 & \arg\mathrm{Max}_{i=2}^{15}\{P(i)\} > \arg\mathrm{Min}_{i=2}^{15}\{P(i)\} \end{cases}$$

It is noted that the first and the last segment of the pitch profile ($P(1)$ and $P(16)$) are not used in order to reduce possible errors in the pitch extraction process. The polarity of I_R indicates the overall temporal trend of pitch movement within the utterance and the magnitude of I_R represents the degree of such variation.

F_0 synthesis

Based on the extension of Fujisaki's model for synthesizing F_0 contours to Thai described in the previous section, the input tone sequences are used to generate predicted F_0 contours. These predicted F_0 contours are basically reference templates to be used for pattern matching against the input contour.

Pattern matching

The classification of input F_0 contours into likely sequences of tones is accomplished in this step by pattern matching against the predicted F_0 contours or reference templates generated by the F_0 model. Pattern matching techniques, such as a simple zero-lag crosscorrelation method or a one-stage dynamic programming search can be used. In both cases, some measure of goodness of fit must be established in order to rank the results so that N-best tone sequences can be obtained. For example, for the zero-lag crosscorrelation method, a correlation coefficient of 0.9 or higher could be used to indicate a relatively good fit. Thus, we can infer that a strong similarity exists between the

input and the predicted F_0 contours. For a one-stage dynamic programming search, a distance measure might be more appropriate. In this research, we used the zero-lag crosscorrelation method.

3.4.3 Performance evaluation

In order to train and evaluate our computer model, we need additional speech materials. Thirty-five target sentences of 11-15 syllables in length are chosen to closely represent continuous speech. Each target sentence consists of syllables with varying tone sequences. Additional requirement is that some of the sentences comprise voiced sounds throughout in order to increase the level of difficulty in performing the syllable segmentation procedure in our tone classification algorithm. Appendix C contains a list of the target sentences described above.

Test stimuli were different from the training stimuli used in training the Fujisaki's model in section 3.1.3. They were produced by a set of five speakers. Thus, there were a total of 175 utterances in the test set.

The classification test was performed on each of the 175 utterances from the test set to obtain the crosscorrelation coefficients between the input contour and each of the predicted contours. All in all, the algorithm misclassified 32 of 175 test utterances. Hence, the classification accuracy for this experiment is approximately 81.7%. In this experiment, the number of N-best output tone sequences is equal to six, i.e., N = 6. The number six was chosen arbitrarily. The reason for outputting N-best tone sequences as inputs to the word hypothesizer is because it is likely that the correct tone sequence could be recovered at that stage by using other linguistic constraints, such as tonal restrictions on the types of syllable structures, etc. The overall performance of the synthesis module was quite reliable in producing F_0 contours. Misclassification mainly occurs with unstressed syllables, especially linker syllables and function words. There are a total of 2,230 syllables in the test stimuli, and only 1822 were correctly classified. This might be due to the fact that unstressed syllables suffer not only from tone neutralization but also from the interaction with adjacent syllables in terms of tonal coarticulation. It is believed that this problem may worsen in the case of polysyllabic words containing linker syllables. However, this problem should not be solved at this stage, but at the stage of word hypothesization where pronunciation dictionary will help rule out ill-formed word.

3.4 Summary and Discussion

A mathematical model for generating F_0 contours for Thai and other tone languages was presented. The model is based on an extension of the Fujisaki's model of F_0 contours. Successfully incorporated into the model are linguistic factors affecting phonetic realization of Thai tones in continuous speech. They are continuity effect due to syllable structure, stress, tonal coarticulation, and declination.

Then, the prosody generation aspect of a text-to-speech system was described, and the above model was applied at the stage of prosody synthesis. The overall performance cannot be assessed at this time because our laboratory does not have a prototype of a Thai text-to-speech system available, and it is beyond the scope of this research. We plan to evaluate our model using the FESTIVAL system developed at Oregon Graduate Institute.

Furthermore, a bottom-up or data-driven approach to automatic classification of Thai tones in connected speech was described. The algorithm is based on the analysis-by-synthesis approach to speech perception, and it is simpler to implement than the left-to-right HMM-based system. Also, we believe that the computational cost of our model is much less than the HMM-based system because it uses fewer parameters.

The present implementation of the algorithm is a continuation of the work done by the principal investigator [17]. Several limitations, such as a lack of automatic segmentation of syllable boundaries, a need to incorporate stress effects into the synthesis module, and a small number of test sentences have been rectified. However, we still are not quite satisfied with the accuracy of the algorithm, yet the results indicate a step in the right direction toward implementing a connected speech tone recognition system. We believe that the overall performance of the algorithm can be improved through a better training of the model, a better pattern matching method, and a more robust F_0 contour analysis method.

4. INTEGRAING TONE CLASSIFICATION WITH A THAI PHONE RECOGNIZER

The question of how to design an automatic speech recognition system (ASR) for Thai has become increasingly more important as speech technology research in Thailand is vigorously pursued. Since Thai is a tone language, a simple emulation of an ASR system for western languages like English, which is not a tone language, is bound to be unsuccessful. Moreover, even though designers of an ASR system for tone languages like Chinese have reported experimental evidence suggesting a successful design, there is no guarantee that those design frameworks will be successfully translated to Thai. This is because the phonetics and phonology for those tone languages are quite different from the phonetics and phonology of Thai. For example, phonotactics (strategies for concatenating sounds) of Chinese is quite different from that of Thai. Mandarin Chinese syllables, for one thing, are sonorant-ending or live syllables only, whereas Thai permits obstruent-ending or dead syllables. Thus, the number of Thai syllables is larger than that of Mandarin Chinese. As far as tones are concerned, Mandarin Chinese has only four tones while Thai has five tones. In addition, although Thai is commonly thought to be a monosyllabic language like Chinese, many words are disyllabic and trisyllabic, especially those borrowed from foreign languages. Thus, it is obvious from the above discussion that past design frameworks from other languages can only serve as guidelines for a design of a Thai ASR system.

There are many design issues that need to be specified when designing a Thai ASR system. These issues involve many spoken language knowledge sources. Knowledge sources commonly used in speech understanding are shown in figure 4.1. There is good evidence that an implicit ordering among these knowledge sources exists. This order of precedence among spoken language knowledge sources indicates that one type of information must be available before it makes sense to progress to the next level.

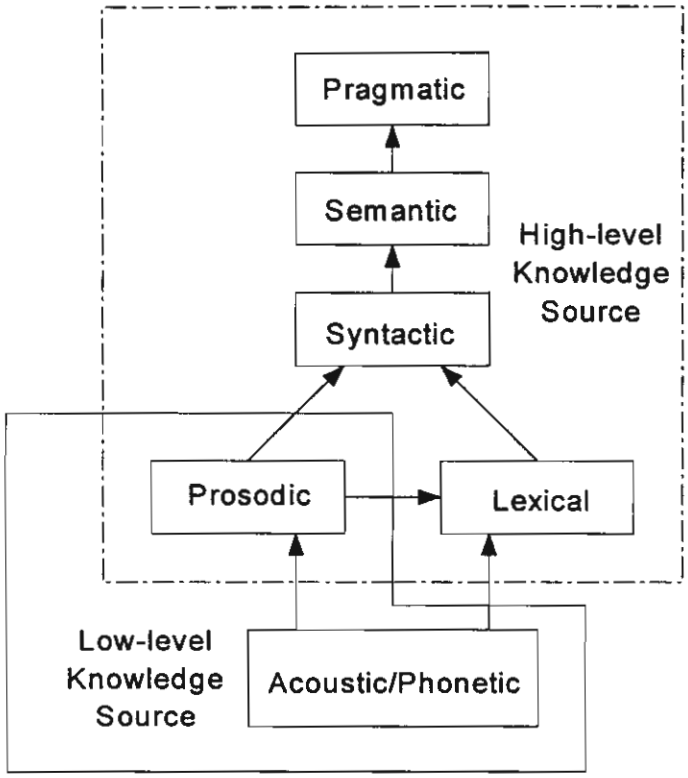

Fig. 4.1. The order of precedence among spoken language knowledge sources.

From the above figure, it should be noted that spoken language knowledge sources can be classified into two groups: low-level and high-level knowledge sources. Prosodic knowledge source is included in both groups because prosody can help a word recognizer rule out word candidates with unlikely stress and durational patterns, but it can also impact syntactic and semantic modules. Therefore, we depict the prosody module as both a high-level and a low-level knowledge source.

In general, speech recognition is aimed at simply recognizing speech (finding out what was said), but not understanding it (finding the meaning of what was said). And, from the above figure, it is clear that speech recognition only involves the use of low-level knowledge source. Thus, designing an ASR system is usually accomplished by finding the best acoustic/phonetic model and the prosody model. These two models should be designed to account for and interpret the acoustic information present in the speech input to the system. The acoustic model must capture the essence of the segmental makeup (sequence of consonants and vowels) of the input utterance while the prosodic model the essence of suprasegmental makeup (tone, stress, rhythm, and intonation). However, speech recognition performance can be improved through the use of some or all

of the knowledge sources from the high level, such as lexical, syntactic and prosodic information, etc. Speech models for other languages now include linguistic modeling to account for coarticulation and grammatical constraints to reduce the search space for the correct utterance.

In this research, we are particularly interested in the issue of tone classification as part of the prosodic model, and how to integrate this feature into the overall design of a Thai ASR system. This chapter outlines the proposed basic framework for a Thai speech recognition system and details the necessary steps for achieving a smooth integration of tone classification with acoustic model. Before describing our proposed system for Thai, background on phonetics and phonology of Thai is in order. This information is necessary for subsequent explanation of our chosen design.

4.1 Phonetics and Phonology of Thai

In this section, we present a brief survey of phonetics and phonology of standard Thai. Standard Thai, the focus of this research, is the dialect spoken in the capital and the central part of Thailand. It is considered the national language of Thailand and is used in broadcasting and in conducting official business and legal matters. It is also the medium of instruction in government schools throughout the country. Since a linguistic analysis of Standard Thai is not a primary goal of this thesis, this overview is by no means intended as an exhaustive linguistic description of Standard Thai. A brief description of the phonetics and phonology is given mainly as a linguistic framework for the subsequent acoustic investigation of acoustic and prosodic models in a Thai ASR system.

This survey is divided into two parts: the segmental features (consonants and vowels) and the suprasegmental features (tone, stress, rhythm, and intonation). The discussion also includes restrictions on the possible combinations of sounds within the frame of the syllable. These restrictions are referred to as phonotactic constraints.

4.1.1 The segmental phonemes

This section deals mainly with the systematic inventory of Thai consonants and vowels. It is a well-established body of knowledge although there are minor disagreements among various linguists. Twenty-one consonants and nine vowels (occurring as single vowels, geminates, and vocalic clusters) are presented.

Vowels

It is generally agreed that Thai has nine short or single vowels: /i/, /e/, /ɛ/, /ɯ/, /ə/, /a/, /u/, /o/, and /ɔ/, and each short vowel has a corresponding long vowel indicated by double letters: /ii/, /ee/, /ɛɛ/, /ɯɯ/, /əə/, /aa/, /uu/, /oo/, and /ɔɔ/. All of them are considered monophthongs. In addition, Thai also has three vocalic clusters or diphthongs: /ia/, /ɯa/, and /ua/. It is noted that the first member of diphthongs is always a high vowel, whereas the second member is always an /a/. Although long vowels are in contrast with their short counterparts, Thai linguists disagree about how to represent such contrasts phonemically [1,7, 42]. The disagreement revolves around the issue of whether or not long vowels should be represented as sequences of two short vowels as advocated by Abramson [1].

An acoustic characteristic that distinguishes among different vowels is the formant structure. Formant pattern is known to be the major physical correlate of vowel quality. Formants are resonant frequencies occurring as peaks in the vowel spectrum. They result from the filtering effect of the vocal tract, which produces amplitude peaks at certain frequencies by enhancing the harmonics at those frequencies while damping harmonics at other frequencies. Vowels have several formants, with the first three being the most important for speech perception. The eighteen Thai monophthongs, classified according to tongue heights and positions, are shown in Table 4.1.1. Also listed are their typical average formant frequencies taken from a detailed acoustic study of Thai vowels by Abramson [1].

Consonants

In Table 4.1.2 are the consonantal phonemes in Thai which are classified according to the states of the glottis (voiced or voiceless), the manner of articulation (stop, non-stop), and the place of articulation (bilabial, dental, alveolar, palatal, velar, and glottal). Voiced and voiceless refer to the state of the glottis during a given articulation. Aspirated and unaspirated refer to the presence or absence of a period of voicelessness during and after the release of an articulation.

Table 4.1.1
Classification of the 18 Thai monophthongs according to tongue heights and positions along with their typical average formant frequencies.

Tongue Height	Formant Freq.	Tongue Advancement					
		Front		Central		Back	
		short	long	short	long	short	long
High		i	ii	ɯ	ɯɯ	u	uu
	F_1	360	300	300	300	360	300
	F_2	2100	2220	1380	1380	720	660
Mid		e	ee	ə	əə	o	oo
	F_1	540	480	540	540	480	480
	F_2	1980	1980	1200	1260	840	840
Low		ɛ	ɛɛ	a	aa	ɔ	ɔɔ
	F_1	780	720	720	780	660	660
	F_2	1800	1800	1380	1380	960	960

Table 4.1.2
The 21 consonantal phonemes in Thai classified according to the states of the glottis and the manner and place of articulation. The phoneme /w/ is entered under both 'bilabial' and 'velar' column to indicate its labio-velar place of articulation.

Manner of Articulation		Place of Articulation					
		Bilabial	Dental	Alveolar	Palatal	Velar	Glottal
Stop	Voiceless Unaspirated	p		t	c	k	ʔ
	Voiceless Aspirated	p^h		t^h	c^h	k^h	
	Voiced	b		d			
Non-stop	Fricative		f	s			h
	Nasal	m		n		ŋ	
	Lateral			l			
	Trill			r			
	Glides	w			j	w	

We list 21 consonantal phonemes although this number is not agreed upon by linguists. The disagreement is centered around the issue of whether or not to include the glottal stop, /ʔ/, as a phoneme. The argument for its exclusion from the phonological system can be summarized as follows. First, the occurrences of glottal stop in citation forms are predictable and phonologically conditioned. Secondly, its occurrence in connected speech is noticeably conditioned on the stylistic variation of speech. This issue will not be resolved until more research is conducted. However, we will include the glottal stop because its presence at the phonemic level enables us to more easily and systematically describe syllable structure in Thai.

4.1.2 From phoneme to phone

Presented above are the consonantal and vowel phonemes in Thai. They represent the abstract description or the phonology of all possible sound units in the language. However, when these sound units are spoken in connected speech, their acoustic manifestations will differ considerably from when they are uttered in isolation. This is due to the fact that speakers tend adjust their articulators in such a way that facilitate the ease of production, resulting in the phenomenon called phone coarticulation. This is similar to tonal coarticulation previously mentioned. Simply put, phone is the acoustic manifestation of phoneme, and maybe slightly different when spoken in continuous speech. Unfortunately for Thai, investigation along this line is scarced, and Thai linguists have not agreed upon a definite number of phones in Thai. Table 4.1.3 shows one possible phone enumeration in Thai using the syllable framework [43]. A syllable consists roughly of three sound units: initial consonant, nucleus or peak, and an optional coda or final consonant. It is noted that we also include foreign phones from English, as possible coda in our inventory. This simply reflects the pervasiveness of English in Thai society. More details on rules of syllabification in Thai will be presented in section 4.1.4.

In addition, regarding vowel length distinction (short vs. long), we have decided to keep them as separate and distinct phones instead of representing long vowel as a concatenation of two short vowels of the same category. Although it is notable that some of the short vowels and their long counterparts have somewhat different vowel qualities, which suggests that vowel spectrum (formant pattern) may be a cue to signal the length distinction in addition to relative duration (see Table 4.1.1). In fact, nearly 30 years after his first experiment, Abramson [44] reported that slight differences in formant pattern are

observed to be the secondary cue to the length distinction while relative duration is still the major cue. For all the vowel pairs, the distinction boundary is influenced by spectral differences with, perhaps, some effect of the timing of the context as well.

Table 4.1.3

Possible Thai phone inventory enumerated in the syllable context.
Phone coarticulation is accounted for as possible Diphone enumeration

Monophone		Thai	Borrowed	Total
Initial consonant	single	p, t, c, k, ʔ, pʰ, tʰ, cʰ, kʰ, b, d, f, s, h, m, n, ŋ, l, r, w, j	-	21
	cluster	pr, pʰr, tr, tʰr, kr, kʰr, pl, pʰl, kl, kʰl, kw, kʰw	br, bl, fr, fl, dr	17
Nucleus or Peak	monophthong	i, ii, e, ee, ɛ, ɛɛ, ɯ, ɯɯ, ə, əə, a, aa, u, uu, o, oo, ɔ, ɔɔ	-	18
	Diphthong	ia, ɯa, ua	-	3
Coda		p, t, k, ʔ, m, n, ŋ, j, w	f, l, s, cʰ	13
			Total	72

Diphone	C_iV	VC_f	VC_i	C_fC_i	Total
Total	912	288	456	912	2,568

4.1.3 The suprasegmental phonemes

This section deals with additional important features of speech sounds called the suprasegmental features, such as length, tone, intonation, and stress. As the word suprasegmental suggests, these features are thought of as 'riding on top of' other segmental features. They may apply either within a single phonetic segment or across numerous phonetic segments in an utterance.

Length

Speech sounds inherently have unequal duration. For example, voiceless affricates in Thai have longer duration than voiced stops. One of the most important uses

of length in Thai is the vowel length distinction used to signal lexical differences. Variation in length is used contrastively (e.g., /bàt/ 'card' vs. /bàat/ 'alms bowl'). Acoustically, long vowels have an average duration that is about twice as long as short vowels [45].

Pitch

Pitch is the psychological correlate of fundamental frequency (F_0) which depends on the rate of vibration of the vocal cords in phonation. Each opening and closing of the vocal cords causes a peak of air pressure in the sound wave, and F_0 is the number of repetitions or cycles of variation in air pressure per second. The unit of F_0 measurement is Hertz (Hz). Changes in pitch or the rate of vibration of the vocal cords can be produced by either stretching and tensing the vocal cords (the tenser the cords the higher the pitch) or by changing the air pressure below the vocal cords (the higher the subglottal air pressure the higher the pitch). However, the most important physiological factor that determines variation in pitch is the tension of the vocal cords. Voiced speech sounds, particularly vowels, may be produced at different pitch levels.

Many different kinds of information, either linguistic (grammatical information at the syllable, word, or sentence level) or non-linguistic (age, gender, etc.), can be conveyed by variations in pitch. Linguistically, Thai uses the variation in pitch called *tone* to convey lexical information about the meaning of a word. In other words, differences in lexical meanings at the syllable level are signaled by tones. Languages that make use of such variation in pitch belong to the class of tone languages. Thai also uses the variation in pitch called *intonation* to convey syntactic information at the phrase or sentence level. The intonation patterns are believed to be superimposed on the tones.

Tones, Intonation, and Stress

For the survey of tone, intonation, and stress features in Thai, see chapter 1.

4.1.4 Phonotactics

In every language, there are restrictions on the possible combinations of sound sequences in different positions in words, particularly at the beginning and end of a word. These restrictions can be formulated in terms of rules stating which sound sequences are

possible and which are not. Restrictions on possible combinations of sounds are known as phonotactic constraints.

Thai, as well as most other languages, employ strategies for concatenating sounds based on the notion of syllable. Although a syllable is relatively difficult to define, it can be roughly thought of as a unit comprising an onset (initial consonant sound), nucleus or peak (vowel sound), and/or a coda or arresting consonant (final consonant sound). Allowable syllable structures in Thai are relatively easy to enumerate compared to English. The following are rules pertaining to syllable structure in Thai.

Each of the 21 consonant sounds (see table 4.1.3) can be used as an onset. An onset is obligatory in Thai. An onset which contains two consonant sounds is called a consonant cluster. There are 12 possible Thai consonant clusters and 5 foreign consonant clusters. Note that the first member of a Thai consonant cluster is always a stop consonant (p, p^h, t, t^h, k, k^h) while the second member can only be r, l, w.

Each of the 18 monophthongs and three diphthongs (see table 4.1.1) can be used as a syllable nucleus or peak. A syllable with a short vowel is called a short syllable; a syllable with a long vowel a long syllable.

A syllable may or may not contain a coda. A syllable with a coda consonant is called a closed or checked syllable; otherwise, it is an open syllable. Only nine consonant sounds can be used as a coda (see table 4.1.3).

Syllables ending with one of the four stops are called obstruent-ending syllables or dead syllables. Only short syllables can end with / ʔ/. That is, the glottal stop never occurs after long vowels or diphthongs. The final glottal stop is always omitted in unstressed open syllables.

Open syllables and syllables ending with one of the nasals or glides are called sonorant-ending syllables or live syllables. Phonetically speaking, those syllables ending with /w/ and /j/ are considered open syllables with a phonetic diphthong or triphthong as a nucleus because of the vowel-like phonetic behaviors of /w/ and /j/. For /w/, the phonetic diphthongs or triphthongs have a final [u] sound; for /j/, a final [i] sound. A phonetic diphthong comprises a short or a long vowel and a final [u] or [i]; a phonetic triphthong, a vocalic cluster and a final [u] or [i]. However, only certain combinations of all possible phonetic diphthongs and triphthongs are allowed. The two sets of allowable combinations of phonetic diphthongs and triphthongs are [iu, iiu, eu, eeu, ɛu, ɛɛu,

ɘu, ɘɘu, au, aau, iau] and [ɘi, ɘɘi, ai, aai, ui, uui, ɔi, ɔɔi, oi, ooi, uai, ɯai]. The following combinations of phonetic diphthongs and triphthongs are prohibited: [ei, eei, ɛi, ɛɛi, ɯi, ɯɯi, iai, ɯu, ɯɯu, ou, oou, ɔu, ɔɔu, ɯau, uau].

With respect to tones, sonorant-ending syllables have no tonal restrictions. All five tones (mid, low, falling, high, and rising) occur on a sonorant-ending syllable. However, only three tones are possible for obstruent-ending syllables. Low and high tones occur on a short syllable ending with p, t, k, or ʔ with an occasional occurrence of a falling tone. Likewise, low and falling tones occur on a long syllable ending with p, t, or k. Occasionally, a high tone appears (mostly in borrowed English words).

From the above rules, 10 syllable structures are possible: CV, CCV, CVN, CCVN, CVVN, CCVVN, CVS, CCVS, CVVS, and CCVVS, where C represents initial consonants, CC consonant clusters, V short vowels, VV long vowels or diphthongs, N nasals, and S stops. Some syllables in Thai, however, are inadmissible due to the co-occurrence constraints. For example, labialized clusters /kw/, and /k^hw/ may not precede the rounded vowel u, o, ɔ, uu, oo, ɔɔ, ua. Likewise, the palatal /j/ may not precede the front vowel i, e, ɛ, ii, ee, ɛɛ, ia. Other co-occurrence constraints include those aforementioned phonetic diphthongs and triphthongs. In addition, syllables may be inadmissible due to the tonal restrictions mentioned above. Luksaneeyanawin [24] reported that out of the remaining admissible syllables, only 5,912 syllables exist in the lexicon of average adult Thai speakers, although the actual figure may vary from speaker to speaker. This figure of 5,912 syllables is based on the study of the distribution of all possible syllables at the morpho-phonological level by examining the vocabulary repertoire of many educated Thai speakers and by consulting several Thai dictionaries.

4.2 A Conceptual Model of a Thai ASR System

In this section, a conceptual model of a Thai ASR system is presented and described in detail including advantages and disadvantages of, and the rationale behind such a design. The model is based on a constraint-based system of integration as outlined in [46]. Fig. 4.2.1 illustrated our constraint-based system for developing a Thai ASR system. It is noted that the tone classification is included in the prosody processor, and the 'lexical decoder' module highlighted in boldface is the main focus of this research.

Roughly speaking, the system consists of two parts: low-level and high-level processing. Low-level processing usually consists of acoustic and prosodic models, which must be capable of accounting for the segmental and suprasegmental features in the speech input. High-level processing, on the other hand, involve the utilization of high-level knowledge source (lexical, syntactic, semantic, and/or pragmatics) at the appropriate level of linguistic unit, i.e., syllable, word, sentence, to correctly assemble what has been sent from the low-level processing units into meaningful text. Before we can describe the overall design of our system and why we chose such a design, a general discussion and reviews of current and past systems are given first.

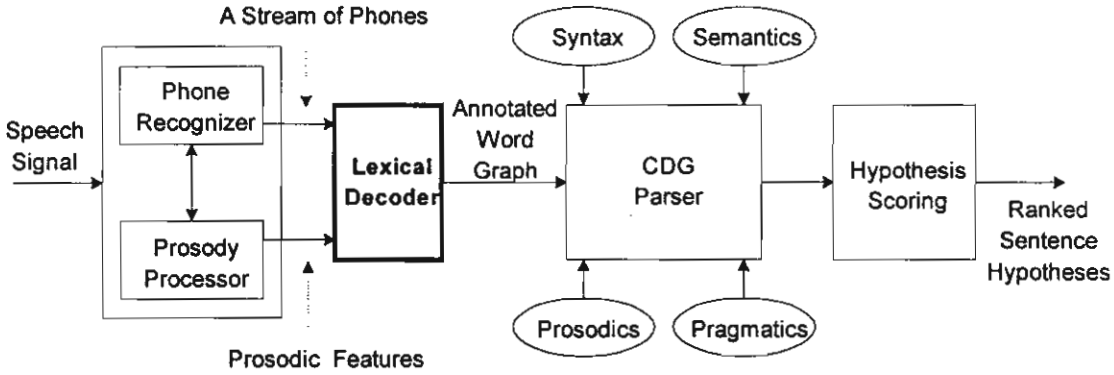

Fig. 4.2.1 A conceptual model of a Thai ASR system

4.2.1 General design criteria

There are many design considerations that need to be taken into account. Besides the difference in the phonetics and phonology of the language, a design of a Thai ASR system should results in a system that can be relatively easy to develop, simple to modify for scale-up purposes, and computationally tractable. Hence, the question of how to efficiently integrate every module together is becoming very important as Thai speech recognition technology matures. Since, a prototype of a Thai ASR system is, to the best of our knowledge, nonexistent, the answer to such a question will have to rely on past and current designs of the system for other languages. After all, speech in whatever language is nothing more than a structured stream of sounds. Similarities are bound to occur.

The level of integration for current and past approaches can be classified into one of three categories: tightly coupled, loosely coupled, or semi-coupled systems. A tightly coupled system is one that integrates all of the knowledge sources for speech into a highly

interdependent set of processes, which cannot be separated. If we apply software-engineering principles to tightly coupled systems, we observe the followings:

1. The language and acoustic models are not separable.
2. It is difficult to evaluate the impact of each of the knowledge sources.
3. For complex domains, the systems tend to be intractable. For example, tightly integrating acoustic/phonetic processing with syntactic processing can yield a system that is orders of magnitude slower than real time.
4. It is difficult to scale up tightly coupled systems to realistic tasks because the integration of knowledge sources makes the system much larger and more difficult to understand.

A loosely coupled system is one that isolates the knowledge sources into relatively independent modules, which communicate with each other. Again applying software-engineering principles, we observe the following properties of loosely coupled systems:

1. They require the system designer to determine the best way for the modules to communicate. This has proven to be a difficult problem [47, 48, 49].
2. They use level-appropriate information, which should make them more tractable. This avoids the combinatorial explosion caused by making acoustic decisions in the syntactic module, for example.
3. Since the knowledge sources are independent, they should scale up to larger problems better than tightly coupled systems.
4. It is easier to measure the impact of each of the knowledge sources (which is important given our current level of understanding).
5. The modularity should make them easier to understand, design, and debug. The individual modules can also be tested in a stand-alone fashion.
6. They can easily accommodate more than one task or language by replacing the individual modules.

Semi-coupled systems fall in between the previous two in that a knowledge source can be used to guide a lower level search in the system. In this capacity, the removal of the knowledge source from the system impacts the lower level search, and so the semi-coupled module is not completely independent. Semi-coupled systems tend to be intractable when they combine level of information from the low-level and high-level categories. Table 4.2.1 summarizes this discussion. We will next present a review of the pastthen detail our approach. We propose a loosely coupled system that uses a uniform approach to integrate the low-level and high-level knowledge sources.

Table 4.2.1.

Characteristics of spoken language systems

	Tightly-coupled	Semi-coupled	Loosely-coupled
Separability	One integrated model for all KS's	Some KS's can be removed but not isolated	Each KS is modeled by a stand-alone module
Inter-module Communication	NA	Typically a one-way communication	Designer specifies interaction
Easy to scale	No	No	Yes
Computation	Usually intractable for all but very small examples	Tractable for small problems or simple language models	May be tractable for large problems
Examples	Acoustic-level CYK parser, Inside-outside algorithm	LR parser with HMM phone verification, N-grams	Constraint-based system, Blackboard model as in Hearsay II

4.2.2 Review of related works

In a tightly coupled system, the language and acoustic models are applied simultaneously. CSELT's system, based on finite-state language models [50], falls in this category. It is interesting to note that their finite-state grammar model is applied as a post-processor in their paper; however, their stated goal is to incorporate it into the hidden Markov model (HMM). Although easily incorporated into HMM, a finite state grammar does not sufficiently constrain the utterances in a spoken language [51]. There are two ways to circumvent this shortcoming and still maintain tight coupling. The first is to modify the HMM to incorporate more powerful language models. For example, the inside-outside algorithm [52, 53] is an extension to HMM, which allows recursive embedding. Whereas a standard HMM can handle only regular grammars, the inside-outside algorithm can process a context-free grammar. The second is to utilize acoustic information in a syntactic processor. For instance, a CYK parser can be modified for acoustic input [54] by exhaustively finding all possible endpoints for every terminal. The modified CYK parser could be thought of as an extension to dynamic time warping speech recognition in much the same way that the inside-outside algorithm could be thought of as an extension to HMM speech recognition.

Although theoretically appealing, the modified CYK parser and the inside-outside algorithm are impractical due to their huge computational costs. Both are $O(n^3)$, where n is the number of input symbols. If these symbols are acoustic measurements (which are typically taken every 10 ms), then the system is intractable. Lari and Young [53] require 64 transputers for very small problems. The inside-outside algorithm requires a large training set in order to train both acoustic and language probabilities.

Tightly coupled systems are hampered by their degree of integration. The best acoustic models do not allow a detailed language model, and the best language models are not well suited for the low-level probabilistic pattern matching needed to accurately classify the acoustic patterns. Systems that work adequately for both acoustic and language processing are often intractable for all but simple examples.

A semi-coupled approach combined a language model with an acoustic model in such a way that they cannot be separated procedurally, even though some components can be removed from the system. In a top-down system, for example, the language model is invoked first at a particular decision point, and then the acoustic model is used to select the best of all candidates that are allowed by the language model. In [55, 56, 57, 58], Kita, Kawabata, and Saito use an LR parser to predict phones, which are then verified by a phone HMM. The phones that make up a word are specified by rules in the grammar. They use a stack splitting method to cope with ambiguity. The acoustic and language components are not entirely separable since the acoustic model receives its focus from the language model.

By far the most successful approach to integrating a language model with an acoustic model has been to embed an N-gram language model into an HMM [59, 60, 61]. The N-gram model assumes that the probability of the current word is a function of the previous N-1 words. This model can easily be integrated because of its simplicity and reduces the perplexity significantly compared to an HMM without a grammar. However, the approach has several disadvantages. Even for small N (i.e., 2 or 3), millions of words of text are required to estimate the N-grams for moderate to large vocabularies. Even so, many of the N-grams are undertrained and extensive smoothing is required. Another disadvantage of N-grams is their task dependence. Also, they do not provide a parse or a semantic representation for a sentence, both of which are useful for speech understanding. We classify this method as a semi-coupled approach since the N-gram model is typically used to guide lower level search and is not a post-processor. However, for N = 2, it may

be possible to incorporate the N-gram directly into HMM topology. In this case, the acoustic and language models are tightly coupled.

In a bottom-up system, the acoustical scores are found first and the language model is then applied to reduce the number of acoustic candidates. This duo of acoustical modeling followed by language modeling can be done at each decision point, or just once, where all acoustical information is extracted prior to utilizing any part of the language model. In the first case, the two models are semi-coupled; whereas in the second case, the language model is invoked as a post-processor and is loosely coupled. For a loosely coupled bottom-up system to work correctly, all relevant acoustic information must be preserved by the acoustic processor. To be tractable, most superfluous information must be discarded.

Several modern systems utilize the language model as a post-processing step, and so these systems are loosely coupled. In Bates [62], the authors first find the N-best [63, 64, 65] sentences with an HMM, and then apply syntactic and semantic rules using a chart parser. CMU's Phoenix uses frame-based parsing and semantic phrase grammar [66] on single sentences. Although individually processing each sentence hypothesis provided by a speech recognizer simplifies the task of the language model, it is inefficient because many sentence hypotheses can be generated with a high degree of similarity. MIT's voyager uses LR parsing [67] as a post-processing step with N-best input. N-best input is simple to process, but at the cost of much repeated work. Seneff's robust parser [68, 69] operates on the most likely sentences and is a post-processor for the speech recognizer.

The most dramatic example of the loosely coupled systems is the blackboard model employed by Hearsay-II [47, 48, 49]. The blackboard model represents each knowledge source as an independent process gathering useful information from and dispensing new information to the blackboard. The blackboard consists of a uniform multi-level network, permitting generation and linkage between alternative hypotheses at all levels. Despite the complete modularity of this approach, it has not been as successful as current approaches that use bigram and trigram models. This is partially because acoustic processing has improved with the advent of new techniques. One reason this approach has not come back into favor may be that the blackboard approach is too loosely couple. When a system is divided into many independent, cooperating processes (as in parallel processing), it is often more difficult to understand, coordinate, and debug the complex interactions among the modules.

4.2.3. Our Approach

In this research, we argue that loose coupling is more appropriate given the current state-of-the-art computing power and given that it allows one to measure more precisely which components of the language model are the most important. We have divided our system into three loosely coupled modules.

The first module consists of two components. The first component corresponds to the low-level knowledge source consisting of acoustic/phonetic and prosodic models. The acoustic-processing component consists of a hidden Markov model (HMM), which outputs likely phone candidates to the next module. At present, HMMs has been proven to be an effective approach to the problem of statistical pattern matching, especially phone recognition. Note that we have chosen phone as our unit of recognition because there are only 72 phones resulting in a construction of 72 phone models to be pattern-matched against the input model. Next, the prosody-processor consists of tone and stress classifiers, which capture the prosodic information and pass them on to the next module as well. It also interacts with the acoustic-processing component to share common and related information such as duration. Note that a tone recognition process should not be tightly integrated into the acoustic processor. This is because the unit of recognition for both phones and tones are quite different. Phone is a segmental feature whereas tone is a suprasegmental feature of a syllable consisting of several phones. Thus, phone spans fewer frames of speech and thus, a need for the second component, the lexical decoder. In light of this difference in length, a conceivable solution is to use syllable as a unit of speech recognition instead of phone. However, as mentioned in section 4.1.3, there are 5,912 admissible syllables in Thai. This number is considered too large for the implementation of an HMM given the computational efficiency of today's computers. That is, we are required to construct 5,912 template models to be matched against the input model.

The second component of the first module, and the focus of this research, is a lexical decoder. Its function is to efficiently and correctly combine together the information passed from the first stage into a sequence of words. This involves the process of syllabification, word hypothesization, and the construction of annotated word graph. A word graph is a directed acyclic graph which provides a very compact and expressive way of capturing all possible sentence hypotheses given the ambiguity inherent in the task of recognizing the words in a continuously spoken sentence. Our goal

for the speech recognizer is to produce a word graph with as few word nodes as possible without eliminating the target sentence hypothesis from the word graph. This is similar to passing a list of N-best sentences to the language model. However, an N-best list of sentence hypotheses limits the information passed between the acoustic model and the language model. In contrast, a word graph of word candidates is typically more compact and more expressive than a list of the most likely sentences. To compare N-best and word graph representations, Harper, et. al. [70, 71], have constructed word graphs from sets of sentence hypotheses. The word graphs provided an 83% reduction in storage, and in all cases they encoded more possible sentence hypotheses than were in the original list of hypotheses. In some cases, the target sentence did not appear in the N-best list but did appear in the word graph. Prosodic information can also be stored in the word graph for higher-level processing.

Clearly, pruning the word graph is important, but in some cases higher-level knowledge is more accurate at pruning the word graph. The pruning that is done by the word recognizer can be done based on extremely low acoustic scores (i.e., a poor match to word candidate), a very simple embedded language model (e.g., N-gram), and word-level prosodic information. The more pruning that can be done before or during word graph construction, the less work the language model has to do. Since the language model typically uses some sort of parsing algorithm, the running time for the algorithm will be at least $O(n^3)$, where n is the number of word nodes in the graph.

The second module involves the language modeling aspect of the system, corresponding to the utilization of high-level knowledge source in ruling out unlikely sentence hypotheses. The module consists of a constraint-based processing component, which is based on a extension to Constraint Dependency Grammar (CDG) parsing as defined by Maruyama [72,73]. This component employs constraint propagation to prune word graphs. This system is capable of propagating a wide variety of constraints, including lexical, syntactic, semantic, prosodic, and pragmatic constraints.

Because the overall system is loosely coupled and the language model is based on a constraint dependency-parsing algorithm, this approach is a very attractive choice for Thai. The first advantage of this approach is that the parser uses a word graph augmented with parse-related information. For both written and spoken Thai sentences, a word graph provides a very compact and less-redundant data structure for simultaneously parsing multiple sentence hypotheses generated by a word segmentation algorithm. There is a lack of delimiters (blanks) between a sequence of words in written Thai and a lexical

decoder cannot segment speech into words in only one way. Secondly, a dependency grammar approach to syntactic analysis is better suited for analyzing Thai than Context-free grammars (CFGs). CFGs are not well suited for parsing Thai sentences because of the absence of inflectional and derivational affixes in Thai, the inconsistent ordering relations within and across phrasal categories, and the discontinuities within sentence constituents.

In terms of power and flexibility, this approach has several advantages, especially the way the system can overcome many of the problems associated with loosely-coupled systems. First of all, instead of using production rules as in CFG, the parser rules out ungrammatical sentences by propagating constraints. Constraint propagation provides a uniform method for applying high-level knowledge sources to prune the word graph. This is different from a blackboard approach in that the designer does not need to create a set of functionally different modules and worry about their interface with other modules. The use of constraints allows a wide variety of information sources, i.e., lexical, syntactic, semantic, pragmatic (contextual), prosodic, and acoustic information, to be represented in a uniform way.

Secondly, the system is more flexible than those that use a CFG parsing approach. The control over which set of constraints to apply is extremely flexible. As a result, unlike a CFG parser, which cannot invoke additional production rules to further prune a set of ambiguous parses for a sentence, the presence of ambiguity in CDG parsing can trigger the propagation of additional constraints to further refine the parses. Also, tight coupling of prosodic and semantic rules with CFG syntactic rules typically increases the size and complexity of the grammar and reduces its understandability. In CDG, syntactic, semantic, and prosodic constraints can be developed independently; the presence of semantic and/or prosodic constraints does not affect (increase) the number of syntactic constraints.

Finally, the set of languages accepted by a CDG grammar is a superset of the set of languages that can be accepted by CFGs. In fact, Maruyama [72, 73], was able to construct CDG grammars with two roles (degree = 2) and two variable constraints (arity = 2) which accept the same language as an arbitrary CFG converted to Griebach normal form. Also, to parse free-order languages like Latin, CFGs require that additional rules containing the permutations of the right-hand side of a production be explicitly included in the grammar [74]. A free-order language can easily be handled by a CDG parser because order between constituents is not a requirement of the grammatical formalism.

Thus, a CDG parser is capable of efficiently analyzing free-order languages because it does not have to test for all possible word orders. The features and advantages of CDG described above make this approach attractive for Thai.

The third module in our system represents our method for merging the influence of the previous two modules in order to select the best sentence hypothesis. That is, it constructs and ranks sentence hypotheses from the pruned word graph of the second module in order to select the best sentence candidates. By annotating the word graph with likelihood information from the first module and then pruning it with constraints representing higher level knowledge from the second module, we are using the appropriate information from both modules to select the best sentence candidates. In the next section, we describe our novel approach in designing the lexical decoder, which integrates the phone recognizer with the tone classifier.

4.3 A Novel Three-stage Lexical Decoder

Although a phone recognizer is necessary as a testbed for methods of acoustic modeling, word recognition is the ultimate goal. In this section, a phone recognizer and a tone classifier serve as the foundation for a three-stage lexical decoder. The phone recognizer outputs the most likely sequence of phone candidates found with an HMM whereas, at the same time, a tone classifier outputs a sequence of tone labels. Figure 4.3.1 shows a phone HMM configuration. It is notable that increasing in popularity is a neural network phone recognizer. Then, the information is passed on to the lexical decoder consisting of a three-stage process: syllabification, word hypothesization, and word-graph construction.

In the first stage, a syllabification process is accomplished through the use of another HMM, which models each syllable as a concatenation of phone labels. The phone labels from the first HMM in the phone recognizer serve as the observations in the second HMM. Figure 4.3.2 shows a syllable HMM with output distributions magnified for the syllable 'ฉา'. Note that the number of states per syllable is set to the number of phones given in the pronunciation dictionary (default pronunciation). However, the node is optional (i.e., it can be bypassed with a null transition). Also, all states have a self-transition (not depicted) so that many labels can be modeled by a single state. The initial

output distribution for each state is determined by smoothing the default pronunciation using the confusion matrix compiled during phone recognition in the phone recognizer.

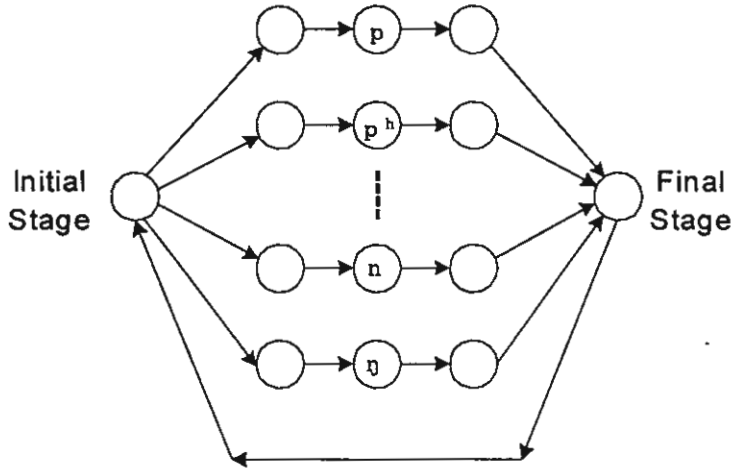

Fig. 4.3.1 A configuration of a phone HMM.

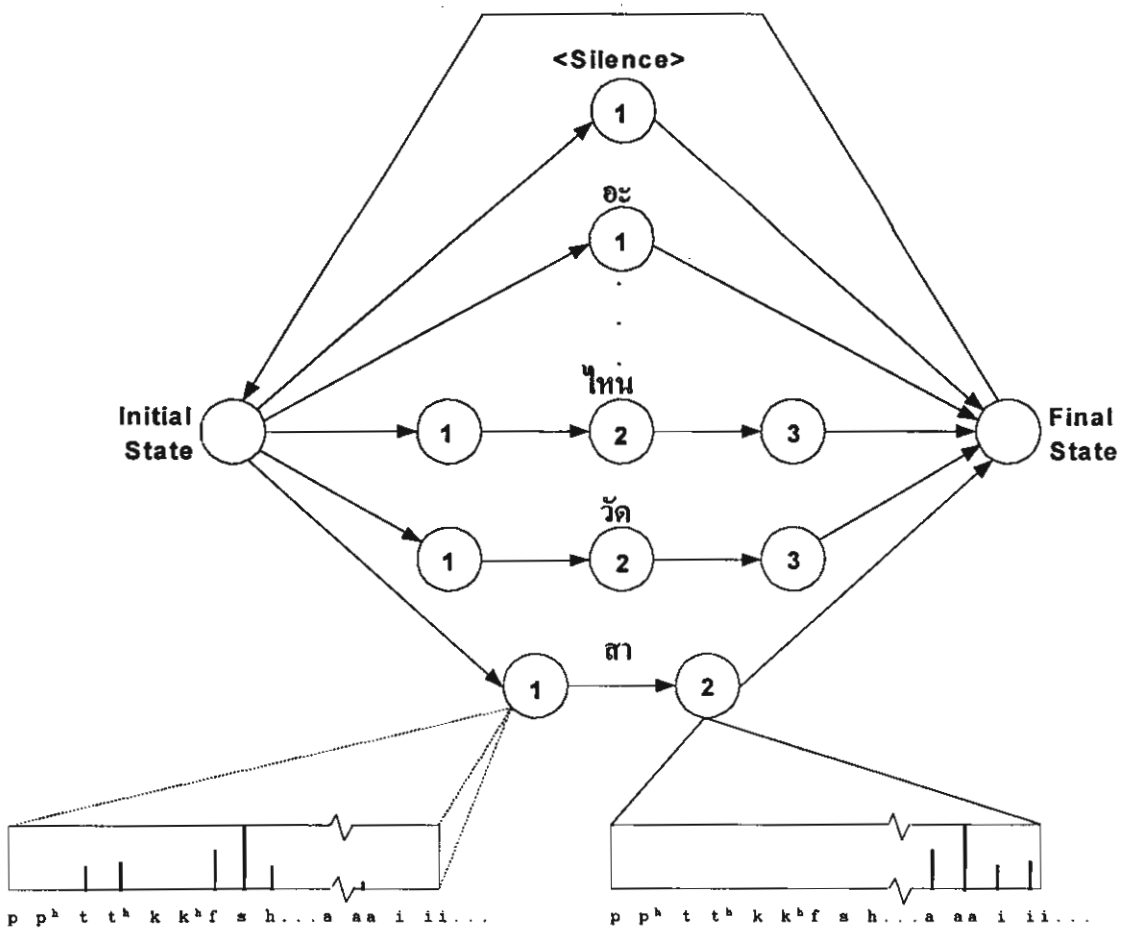

Fig. 4.3.2 Syllable HMM with output distributions magnified for the syllable 'สา'.

Next, the syllable and tone information are now ready to be integrated using dynamic programming to match the hypothesized syllable string with each of the N-best tone sequences. The algorithm utilizes dips in the energy contours and duration information in the matching process. This step is crucial because the ability to assign the right tone to the appropriate syllable require the best alignment of the two types of information. This misalignment problem is to be expected when phone and tone are recognized separately and the segmentation of the input utterance into syllables may become quite different in the two recognizers. The problem gets even worse when insertion/deletion occurs in one of the recognizers or both.

In the second stage, a word network models each vocabulary word as a concatenation of tone-assigned syllable labels from the first stage via a lexical access process. This process is called a word hypothesization process, and, as a result, a word lattice is constructed. The underlying structure for the hypothesizer network is provided by the use of another HMM. The most likely word string can be found using a Viterbi-search. For every possible starting time i and ending time j, the most likely words will be chosen based on Viterbi probability of the subsequence $O_i, O_{i+1}, \ldots, O_j$. The probability of a word occurring from i to j will be approximated as the probability of the word staring at i multiplied by the probability of the word ending at time j. The output of this hypothesizer network is a large recognition lattice containing acoustic and grammar likelihoods for each word node. In addition, a single tone-assigned syllable can generally appear as a monosyllabic word or as part of a polysyllabic word. Therefore, the resulting word lattice can be very large and complicated especially when an N-best tone-assigned syllable sequence is used and N is large. A partially shown sample of a word lattice constructed from a test sentence "**สารวัตรอยู่ไหน**" is illustrated in fig. 4.3.3.

It should be noted that the problem of obtaining a good word lattice is not easily answered. It is difficult to define a single good measure for word lattices because there is a tradeoff between the size of the lattice and correctness. Obviously, the smallest possible lattice that contains the correct path is desired. However, such a lattice can be prohibitively large. In some cases, a small number of omissions can be tolerated if the lattice is small enough to work with and it contains the intended meaning.

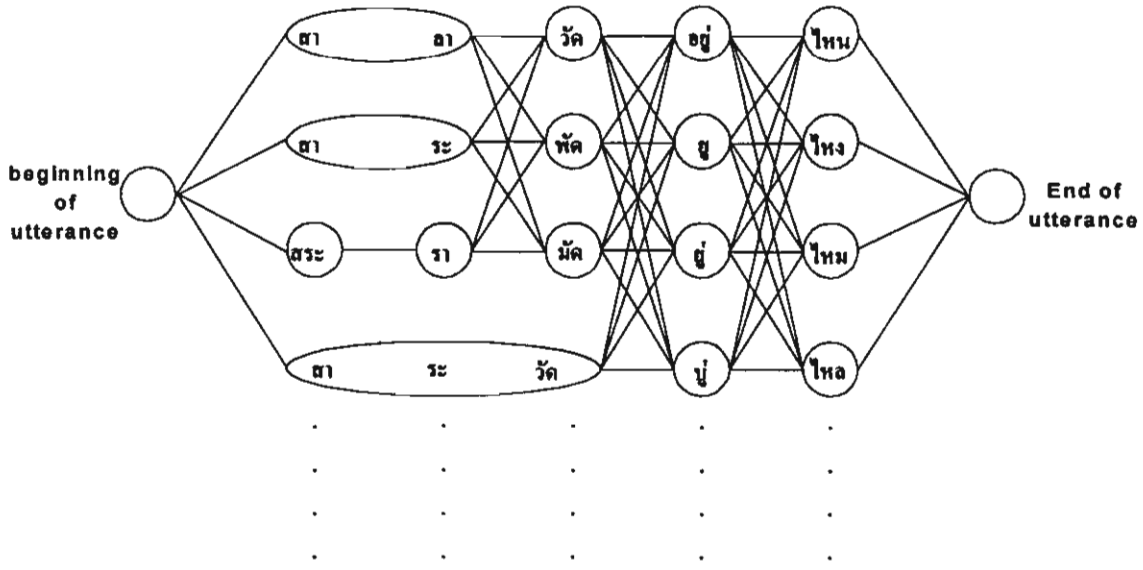

Fig. 4.3.3 A partially shown sample of a word lattice, in which each circle represents a monosyllabic word while each ellipse a polysyllabic word

As previously mentioned, the most successful automatic speech recognition systems are those that utilize higher-level knowledge source such as syntax and semantics, in addition to acoustic and lexical knowledge. Hidden Markov modeling has been one of the most successful strategies for acoustic pattern matching, but this method is generally difficult to integrate with adequate language models. Approaches that jointly model the grammar and the acoustic signal have been applied to small problems successfully. Widespread use of these strategies for larger problems has been limited due to computational costs, insufficient training data, or an inadequate language model.

By separating the language model from the acoustic model, it is possible to use a more accurate language model without increasing computational costs or the amount of training data required. Decoupling these knowledge sources is possible only if the language model is conditionally independent of the acoustic model given some intermediate knowledge source. One of the most promising intermediate representations is a probabilistic word lattice described above. Nevertheless, we have chosen to transform the word lattice into a word graph, annotated with probabilities that allow the highest probabilities sentences to be examined in order of decreasing probability. This word graph representation can accommodate our chosen language model based on constraint dependency grammar, which is powerful and suitable for describing Thai. Next, we describe the construction of our word graph from the word lattice obtained above.

The third stage of our lexical decode involves the construction of a word graph to be used in interfacing a CDG parser with a lexical decoder in our system. As previously mentioned, a word graph is a directed acyclic graph representing the possible word paths through the utterance. Nodes in the word graph represent the words and connecting arcs represent word transitions. Construction of the word graph is accomplished by post-processing the hypothesization lattice from the second stage. Since the lattice contains full alignment information such as start and end times for each word node, it can include many identical or nearly-identical path that vary only with regard to time alignment. The language processing component does not need this alignment information and is slowed by the redundant information. Therefore, word graph generation includes an algorithm to eliminate identical sub-graphs from the lattice, resulting in graph that represents all possible word-level paths without eliminating or adding any path possibilities. Note that each word graph has a distinct starting node and a distinct ending node. Fig. 4.3.4 shows a sample word graph for four sentence hypotheses obtained from the lattice in fig. 4.3.3.

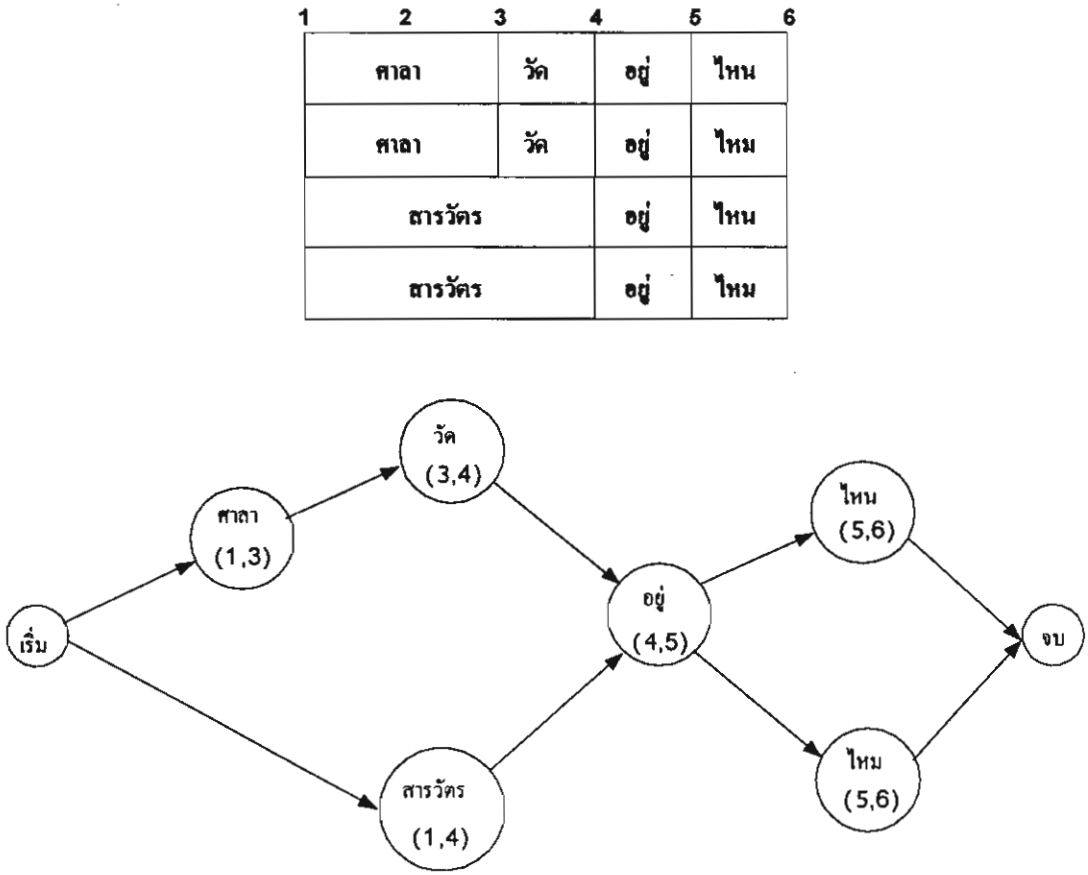

Fig. 4.3.4 A sample word graph generated from four sentence hypotheses obtained from the word lattice in fig. 4.3.3.

4.4 Summary and Discussion

The lexical decoder, the primary focus of this chapter, has been describes in detail above. The module consists of a three-stage process, namely, syllabification, word hypothesization, and word graph construction. In this section, we describe the advantages and disadvantages of our design of the lexical decoder.

First of all, decoupling the acoustic and language processors adds flexibility. A variety of language models can be tried with a single acoustic model. In this research, a constraint-based parser can utilize higher-level knowledge source such as syntax, semantics, prosody, and pragmatics efficiently using constraint propagation.

Secondly, we have chosen to decouple the acoustic processor from the word matcher. Separating the phone recognizer from the word recognizer is attractive for several reasons. For one thing, the phone level pattern matching is essentially shared across word contexts, resulting in much lower computational requirements. Moreover, this approach helps facilitate the development of a robust recognizer that is effective across a number of databases (or recording conditions) since the word classifier is decoupled from the acoustic processor. In addition, new words can be added to the vocabulary more easily. However, decoupling the phone and word recognizers makes sense only if most of the relevant acoustic information is passed on by the phone recognizer. With 25 % phone recognition errors or worse, the word recognizer must utilize higher-level knowledge source (such as lexical) and model the lower level errors in order to overcome the poor performance of the acoustic level pattern matcher. This is exactly what takes place in a standard HMM (i.e., the best performance by far for a phone recognizer is approximately 72% accuracy). The lexical constraints are enforced by the topology and the phone errors are modeled by the difference in likelihood.

Thirdly, an HMM word matcher has the advantage of modeling phone errors probabilistically. However, the disadvantage is due to the imperfection of the lexical knowledge. There are often many ways to pronounce a single word, and only a few pronunciations are modeled. With the HMM word matcher approach, detailed lexical modeling is accomplished by automatically learning pronunciations during training.

Fourthly, we have chosen to model our syllabifier using a separate HMM from the phone recognizer. There are several advantages of using two successive HMM stages rather than on large HMM. A frame level acoustic HMM weights all frames equally so those longer phones have a disproportionately large influence on the Viterbi probability.

The first HMM is also more sensitive to longer phones, but this does not prevent short phones from being recognized because all phones are connected in parallel. On the other hand, the second HMM processes phone labels so that long phones and short phones are equally weighted. Furthermore, the segmented output of the first HMM provides an easy way to incorporate prosody. For instance, average pitch can be measured over a phone segment, and delta pitch can be used as a feature of the second HMM. The final advantage is that the two-level HMM requires less computation since the frame level pattern matching is essentially shared across a large number of contexts. Because of its lower computational requirements, the two-stage HMM is applicable to the problem of fast matching to reduce the search space of a more detailed pattern matcher.

Finally, the main disadvantage of using two HMM stages is that some acoustic information is being discarded. The most likely word sequence may contain some very unlikely phones. Although deletions, insertions, and substitutions can be modeled by the second HMM, errors may be impossible to overcome if the implicit segmentation produced by the first HMM is poor. However, most phone errors are substitutions within a category (e.g. vowels) so that the implicit segmentation is likely to be acceptable.

It should be noted that the above description of our chosen model for a Thai ASR system and how we integrate the tone classifier with the word recognizer is based on a preliminary development of the model. The implementation was performed using HTK Version 3.1 by Entropic [75, 76]. The models are trained using the speech materials used in the implementation of an automatic tone classification in chapter 3 (see appendix C). A lack of standard acoustic-phonetic speech databases for Thai and time limitation prevent us from a full-blown implementation of the system. We estimated at least another year for the completion of such endeavor. Hence, accuracy testing and performance evaluation could not be assessed as planned. But, we do hope that we have offered enough insight into the design of a Thai ASR system so as to provide a stepping stone for further investigations into the subject for us and for other researchers in the field.

5. CONCLUSIONS

The goal of this chapter is to put the work described in the previous chapters into perspective. The summary of the research is first presented. Next, the limitations and drawback of the approach are given, as well as recommendations for future research. Finally, a list of outputs and contributions of this study to research in the field of speech and natural language processing of Thai are discussed.

5.1 Summary

This report has presented research directed toward the development of a Thai ASR system. The research concentrated mainly on the issues of tone classification and its impact on the design of a speaker-independent ASR system for Thai. It should be emphasized that the goal of this research is not to build a prototype of such a system. Instead, we strive to identify the best possible design of the system given the current state-of-the-art of technology in terms of computing power, signal processing methods, and modeling techniques. Equally important for such a design is the basic knowledge regarding the linguistic description of continuous Thai speech, both at an abstract level and at the acoustic-phonetic level. This bulk of knowledge is absolutely essential and crucial toward a design of this magnitude in order to avoid the trial-and-error method of selecting a procedure that often leads to 'hit-or-miss' results. In other words, we support the position that an empirical study of the acoustic realization of the event of interest should be carried out before performing a computer simulation.

The research begins with the conduction of two acoustic experiments to study 1) the effects of stress on F_0 contours of Thai tones in connected speech and 2) to assess and quantify the interactions between stress and tonal coarticulation affecting the F_0 realization of Thai tones. In particular, we were interested in the acoustic manifestations

in terms of height and shape of the F_0 contours of five Thai tones and the pattern of contrast among them. The findings with respect to the first experiment indicates that under the influence of stress alone, F_0 contours of stress syllables more closely approximate the F_0 contours of syllables spoken in isolation than that of the unstressed syllables. Furthermore, the contrasting patterns of F_0 contours of unstressed syllables become less clear as a result of tone neutralization, which is a process by which F_0 contour loses its original shape in response to the influence of stress. However, a new pattern emerges and F_0 contours of the five Thai tones can be categorized into three tonal registers in the tone space, the low, the mid, and the high register. With respect to the second experiment, the findings seem to suggest that F_0 contour of unstressed syllable not only suffers tone neutralization as in the first experiment, but also absorbs more easily the lingering effect of perseverative tonal coarticulation from the previous tone. Regarding the pattern of contrast, F_0 contours of the five Thai tones migrate toward the middle of the tonal space resulting in the narrowing down of the movement of the dynamic tones in particular.

Based on findings from the acoustic experiments above, we have successfully modified our analysis-by-synthesis method of automatic tone classification. Details of the implementation were presented in chapter three of this report. The highlight of our design is the ability to incorporate all of the linguistic factors affecting F_0 realization of Thai tones in connected speech into the model. Moreover, we have also demonstrated the application of our mathematical model of F_0 contours generation in a Thai text-to-speech system. This model should help improve the intelligibility and naturalness of synthesized Thai speech.

In chapter four of the report, we have presented the conceptual model of a Thai ASR system. We paid particular attention to the design of a lexical decoder module of the system, which is the main focus of this research. The purpose of a lexical decoder is to 1) combine the process of tone classification with a Thai phone recognizer, 2) identify the best possible word sequences from the speech input, and 3) present its outputs in a format or representation that is conducive to further processing by the language processing component of the system. We have proposed a novel three-stage lexical decoder utilizing successive HMMs. It is unfortunate that we were unable to fully assess the performance of our lexical decoder due to many limitations. For one thing, a lack of standard acoustic-phonetic speech databases prevents us from implementing a better phone recognizer. Secondly, the model is far more complicated than originally anticipated resulting in huge

requirements of both time and computing power. Given the allotted time of six months, we can address certain aspects of the model only. As a result, performance evaluation in terms of recognition accuracy could not be performed as planned.

5.2 Future Research

The work presented in this report represents only a first step toward achieving the goal of automatic recognition of Thai connected speech. Our attempts can only be classified as a crude design given the amount of resources available at our disposal. Our design is obviously in desperate needs of major refinements. However, we are quite certain that we have achieve our goal, at some level or another, of identifying the best possible design of a Thai ASR system. We hope that our work will inspire other researchers in the field to come up with a better design. There are many issues that require further investigation, some of which we discuss next.

With respect to the problem of automatic tone classification, our approach is classified as a rule-based method. To improve the classification accuracy, new and better rules that can capture the complex interactions among several linguistic factors must be devised. On the other hand, a statistical pattern-matching approach, such as HMM or neural network-based classifier, might offer an answer to this complex problem. However, this undertaking may require enormous amount of training speech data, which, at the moment, is few and far in between.

Concerning the best and most efficient design of a Thai ASR system, the solutions to this difficult task are unlikely to emerge some time in the near future. Nonetheless, we offer the following suggestions.

First of all, the speech research community is in dire needs of several standard speech databases to be used for training and testing speech models. This task requires concerted efforts among researchers in the field.

Secondly, several improvements to the acoustic modeling techniques must be tried in order to achieve higher recognition accuracy. In general, higher recognition accuracy can be achieved by any of the following methods. These include 1) determining a set of acoustic features that better represent the speech signal over short intervals (i.e., frames), 2) improved modeling of the relationship between successive frames, which may depend considerably on the history and perhaps future of the speech signal, and/or 3) finding better measures to compare various aspects of the utterance to those of the model. In

addition, recognition accuracy can also be improved by correctly modeling higher-level knowledge sources, such as prosody, syntax and semantics, into a mathematically tractable model.

Finally, higher recognition accuracy should not be the only goal of such a design. The speech recognition system should be easy to expand or scale up to larger problem as well. This may require the construction of speech tools to facilitate the implementation of newly developed modeling techniques. Thus, it is necessary to develop a faster method of experimentation in terms of parallel processing utilizing a large number of workstations. Parallel computing should help reduce training time of computationally intensive models from weeks to several hours.

5.3 List of Outputs

5.3.1 International publications

1) Potisuk, S. "F_0 Realization of Thai Tones in Connected Speech" Manuscript under revision and waiting to be submitted to *Phonetica*.

Note: The manuscript was originally sent to Professor Dr. Jack Gandour, Department of Audiology and Speech Sciences of Purdue University for initial review. Since the author's main expertise is not in theoretical linguistics, Dr. Gandour's comments and suggestions has proven to be valuable and should help increase the likelihood of the paper getting accepted for publication. The followings are key points in the paper suggested for revision:

a) The entire paper should be split into two.
b) Statistical analysis should be redone. Student-Neuman-Kools post-hoc planned comparison test should be added regarding pattern of contrast.
c) Relevant discussion and interpretation of results in terms of theoretical linguistic significance should be added.

Based on the above comments and suggestions, we have revised the entire manuscript by splitting it into two papers, one for each acoustic experiment. The first one, entitled "The Effects of Stress on F_0 Contours of Thai Tones in Connected Speech," is included with this report. The second one, entitled "Acoustic Characteristics of Unstressed Syllables Under the Influence of Perseverative Tonal coarticulation in Thai," is being prepared awaiting results from new statistical analysis.

2) Potisuk, S. "Automatic Classification of Thai Tones in Continuous Speech" Manuscript is expected to be submitted for publication in *IEEE Transactions of Speech and Audio Processing* or *Computer, Speech, and Language Journal.*

5.3.2 National publications

Potisuk, S. "Prosody Generation in a Thai Text-to-speech System" Manuscript submitted for publication to *The Sixth National Computer Science and Engineering Conference* (NCSEC 2002). To be held during 29-31 October 2002 in Pattaya.

5.3.3 Copyrighted materials

-

5.3.4 Manuscripts in preparation

Potisuk S. "Syllable Segmentation of Thai Speech Using a Modified Teager's Energy Algorithm"

Potisuk S. "A Novel Method for Integrating Tone Classification with a Thai Phone Recognizer: A Lexical Decoder Design"

Potisuk, S. "Acoustic Characteristics of Unstressed Syllables Under the Influence of Perseverative Tonal coarticulation in Thai"

LIST OF REFERENCES

LIST OF REFERENCES

[1] A. S. Abramson, "The vowels and tones of standard Thai: acoustical measurements and experiments," *International Journal of American Linguistics*, vol.28-2, Part III (Publication No.20), 1962. [Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics].

[2] J. T. Gandour and R. Harshman, "Cross-language differences in tone perception: A multidimensional scaling investigation," *Language and Speech*, vol. 21, pp. 1-33, 1978.

[3] J. T. Gandour, "Tone perception in Far Eastern languages," *Journal of Phonetics,* vol. 11, pp. 149-175, 1983.

[4] S. H. Chen, S. Chang, and S. M. Lee, "A statistical model based fundamental frequency synthesizer for Mandarin speech," *Journal of The Acoustical Society of America*, vol. 92, pp. 114-120, 1992.

[5] J. T. Gandour, "Consonant types and tones in Siamese," *Journal of Phonetics*, vol. 2, pp. 337-350, 1974.

[6] R. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, vol. 11, pp. 51-62, 1983.

[7] S. Luksaneeyanawin, *Intonation in Thai,* Ph.D. dissertation, University of Edinburgh, 1983.

[8] S. Potisuk, J. T. Gandour, and M. P. Harper, "Acoustic correlate of stress in Thai," *Phonetica*, vol. 53, pp. 200-220. 1996.

[9] S. Potisuk, J. T. Gandour, and M. P. Harper, "Contextual Variations in Trisyllabic Sequences of Thai Tones," *Phonetica*, vol. 54, pp. 22-42. 1997.

[10] B. Connell and D. R. Ladd, "Aspect of pitch realization in Yoruba," *Phonology*, vol. 7, pp. 1-29, 1990.

[11] A. S. Abramson and K. Svastikula, "Intersections of tone and intonation in Thai," Haskins Laboratories, New Heaven, CT, Status Report on Speech Research SR-74/75, pp. 143-154, 1983.

[12] S. X.-N. Shen, *The Prosody of Mandarin Chinese*. Berkeley, CA: University of California Press, 1989.

[13] X. Chen, C. Cai, P. Guo, and S. Ying, "A hidden Markov model applied to Chinese four-tone recognition," in *1987 International Conference on Acoustics, Speech and Signal Processing*, Vol. II, May 1987, pp. 787-800.

[14] L., Liu, W. Yang, H. Wang, and Y. Chang. Tone recognition of polysyllabic words in Mandarin Speech," *Computer Speech and Language*, vol. 3, pp. 253-264, 1989.

[15] R. Wu, J. A. Orr, and S-K. Hsu, "Recognition of four tones in Chinese speech by parametric estimation of frequency trajectories," in *1989 2nd Biennial Acoustics, Speech and Signal Processing Central New England Miniconference*, 1989.

[16] T. Lee, P. C. Ching, L. W. Chan, Y. H. Cheng, and Brian Mak, "Tone recognition of isolated Cantonese syllables," *IEEE Trans. Speech and Audio Processing*, vol. 3-3, pp. 204-209, May 1995.

[17] S. Potisuk, M. P. Harper, and J. T. Gandour, "Classification of Thai tone sequences in syllable-segmented speech using the analysis-by-synthesis method," *IEEE Transaction on Speech and Audio Processing*, Vol.7, No.1, January 1999, pp.95-102 .

[18] H. Fujisaki, "Dynamic characteristics of voice fundamental frequency in speech and singing," in *The production of Speech*, P. F. MacNeilage, Ed. New York: Springer-Verlag, 1983, pp. 39-55.

[19] L. S. Lee, C.Y. Tseng, K. J. Chen, I. J. Hung, M. Y. Lee, L. F. Chien, Y. Lee, R. Lyu, H. M. Wang, Y. C. Wu, T. S. Lin, H. Y. Gu, C. P. Nee, C. Y. Liao, Y. J. Yang, Y. C. Chang, and R. C. Yang, "Golden Mandarin (II) - An improved single-chip real-time Mandarin dictation machine for Chinese Language with very large vocabulary," *1993 Proceedings of the International Conference on Acoustic, Speech, and Signal Processing*, pp. 503-506. May, 1993.

[20] Y. Gao, H. W. Hon, Z. Lin, G. Loudon, S. Yogananthan, B. Yaun, "Tangarine: A large vocabulary Mandarin dictation system," *1995 Proceedings of the International Conference on Acoustic, Speech, and Signal Processing*, Detroit, Michigan. pp. 77-80. May, 1995.

[21] P. Bee, "Restricted phonology in certain Thai linker-syllables," in *Studies in Tai Linguistics in honor of William J. Gedney*, J. Harris and J. Chamberlain, Eds. Bangkok: Central Institute of English Language, 1975, pp. 17-32.

[22] S. Hiranburana, *The role of accent in Thai grammar.* Ph.D. dissertation, School of Oriental and African studies, University of London, 1971.

[23] S. Hiranburana, "Changes in the pitch contours of unaccented syllables in spoken Thai," in *Tai phonetics and phonology*, J. Harris and R. Noss, Eds. Bangkok: Central Institute of English Language, 1972, pp. 23-27.

[24] J. Gandour, "On the representation of tone in Siamese," in *Studies in Tai Linguistics in honor of William J. Gedney*, J. Harris & J. Chamberlain (Eds.). Bangkok: Central Institute of English Language, 1975, pp.170-195.

[25] P. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation," *Journal of Acoustical Society of America*, vol. 90-6, pp. 2956-2970, 1991.

[26] P. Rose, "Considerations in the normalisation of the fundamental frequency of linguistic tone," *Speech Communication,* vol. 6, pp. 343-351, 1987.

[27] J. T. Gandour, S. Potisuk, and S. Dechongkit, "Inter- and intraspeaker variability in fundamental frequency of Thai tones," *Speech Communication,* vol. 10, pp. 355-372, 1991.

[28] P. Ladefoged, *A Course in Phonetics* (3rd ed.). New York: Harcourt Brace Jovanovich, 1993.

[29] B. Möbius, M. Pätzold, and W. Hess, "Analysis and synthesis of German F_0 contours by means of Fujisaki's model," in *Speech Communication*, vol. 13, pp. 53-61, 1993.

[30] G. Bailly, "Integration of rhythmic and syntactic constraints in a model of generation of French prosody," *Speech Communication*, vol. 8, pp. 137-146, 1989.

[31] D. Hermes and J. Van Gestel, "The frequency scale of speech intonation," *Journal of the Acoustical Society of America*, vol. 90, pp. 97-102, 1991.

[32] S. Potisuk, J. T. Gandour, and M. P. Harper, "Vowel and stress in Thai," *Acta Linguistica Hafniensia*, vol. 30, pp. 39-62. 1998.

[33] E. O. Selkirk, *Phonology and Syntax: The Relation Between Sound and Structure*. MIT press, 1984.

[34] M. Nespor and I. Vogel, *Prosodic Phonology*, J. Koster and H. V. Riemsdijk, Eds. Dordrecht-Holland, 1986.

[35] P. E. Vongvipanond, "Linguistic problems in computer processing of the Thai language," in *1993 Proceedings of the Symposium on Natural Language Processing in Thailand.* Chulalongkorn University, Bangkok, Thailand, 1993, pp. 519-545.

[36] S. Potisuk and M. P. Harper, "CDG: An alternative formalism for parsing written and spoken Thai." *Proceedings of the Fourth International Symposium on Language and Linguistics: Pan-Asiatic Linguistics*, Vol. 4, pp. 1177-1196, 1996.

[37] J. P. Gee and F. Grosjean, "Performance structures: A psycholinguistic and linguistic appraisal," *Cognitive Psychology*, vol. 15, pp. 411-458, 1983.

[38] T. Luangthongkum, *Rhythm in standard Thai,* Ph.D. dissertation, University of Edinburgh, 1977.

[39] G. H. Yeni-Komshian, "Speech Perception," in *Psycholinguistics*, J. B. Gleason and N. B. Ratner, Eds. Fort Worth: Harcourt Brace, 1993, pp. 89-131.

[40] J. F. Kaiser, "On a simple algorithm to calculate the energy of the signal," *1990 Proceedings of the International Conference on Acoustic, Speech, and Signal Processing*, pp. 381-384. April, 1990.

[41] A. Komatsu, A. Ichikawa, K.Nakata, Y. Asakawa, and H. Matsuzaka, "Phoneme recognition in continuous speech," in *1982 Proceedings of the International Conference on Acoustic, Speech, and Signal Processing*, pp. 883-886. May, 1982.

[42] E. J. Henderson, "The phonology of loanwords in some South-East Asian languages," *Transactions of the Philological Society*, pp. 131-158, 1951.

[43] P. Tuaychareon, *Applied Phonetics*. Thammasart University press, 1990.

[44] A. S. Abramson and N. Ren, "Distinctive vowel length: Duration vs. Spectrum in Thai," *Journal of Phonetics*, vol. 18, pp. 79-92, 1990.

[45] K. Svastikula, *A Perceptual and Acoustic Study of the Effects of Speech Rate on Distinctive Vowel Length in Thai*, Ph.D. dissertation, University of Connecticut, 1986.

[46] S. Potisuk, *Prosodic Disambiguation in Automatic Speech Understanding of Thai*. Unpublished Ph.D. Dissertation, 1995

[47] L. D. Erman and V. R. Lesser, "The Hearsay-II speech understanding system: A tutorial," in *Trends in Speech Recognition*, W. A. Lea, Ed. Apple Valley, MN: Speech Science Publications, 1986, pp. 361-381.

[48] F. Hayes-Roth, "Focus of attention in the Hearsay-II speech understanding system," Department of Computer Science, Carnegie-Mellon University, PA, Tech. Rep., 1977.

[49] V. R. Lesser, R. D. Fennell, L. D. Erman, and D. R. Reddy, "Organization of the Hearsay-II speech understanding system," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-23, pp. 237-257, July 1991.

[50] E. P. Giachin, "Automatic training of stochastic finite-state language models for speech understanding," in *1992 International Conference on Acoustics, Speech and Signal Processing*, March 1992.

[51] F. Pereira and D. Roe, "Empirical properties of finite state approximations for phrase structure grammars," in *1992 International Conference on Spoken Language Understanding*, Vol. I, October 1992, pp. 261-264.

[52] J. Kupiec, "Hidden Markov estimation for unrestricted stochastic context-free grammars," in *1992 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1992, pp. 177-180.

[53] K. Lari and S. J. Young, "Applications of stochastic context-free grammars using the inside-outside algorithm," *Computer Speech and Language*, vol. 5-3, pp. 237-257, July 1991.

[54] H. Ney, "Dynamic programming parsing for context-free grammars in continuous speech recognition," *IEEE Trans. Signal Processing*, vol. 39-2, Feb. 1991.

[55] T. Hanazawa, K. Kita, S. Nakamura, T. Kawabata, and K. Shikano, "ATR HMM-LR continuous speech recognition system," in *1990 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1990, pp. 53-56.

[56] K. Kita, T. Kawabata, and T. Hanazawa, "HMM continuous speech recognition using stochastic language models," in *1990 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1990, pp. 581-584.

[57] K. Kita, T. Kawabata, and H. Saito, "HMM continuous speech recognition using predictive LR parsing," in *1989 International Conference on Acoustics, Speech and Signal Processing*, Vol. II, May 1989, pp. 703-706.

[58] K. Kita and W. H. Ward, "Incorporating LR parsing into SPHINX," in *1991 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1991, pp. 269-272.

[59] L. R. Bahl, F. Jelinek, and R. L. Mercer, "A maximum likelihood approach to continuous speech recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 179-190, 1983.

[60] D. B. Paul, "An efficient A* stack decoder algorithm for continuous speech recognition with a stochastic language model," in *1992 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, March 1992.

[61] P. Placeway, R. Schwartz, P. Fung, and L. Nguyen, "The estimation of powerful language models from small and large corpora," in *1993 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, April 1993.

[62] M. Bates, R. Bobrow, P. Fung, R. Ingria, F.Kubala, J. Makhoul, L. Nguyen, R. Schwartz, and D. Stallard, "The BBN/HARC spoken language understanding system," in *1993 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, April 1993, pp. 111-114.

[63] R. Schwartz and S. Austin, "A comparison of several approximate algorithms for finding multiple N-best sentence hypotheses," in *1991 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1991.

[64] R. Schwartz, S. Austin, F. Kubala, J. Makhoul, L. Nguyen, and P. Placeway, "New uses for the N-best sentence hypotheses within the Byblos speech recognition system," in *1993 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, April 1993.

[65] R. Schwartz and Y-L. Chow, "The N-best algorithm: An efficient and exact procedure for finding the N most likely sentence hypotheses," in *1990 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, April 1990.

[66] W. Ward, "Understanding spontaneous speech: The Phoenix system," in *1991 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1991.

[67] V. Zue, J. Glass, D. Goodine, H. Leung, M. Phillips, J. Polifroni, and S. Seneff, "Integration of speech recognition and natural language processing in the MIT Voyager system," in *1991 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, May 1991.

[68] S. Seneff, "Robust parsing for spoken language systems," in *1992 International Conference on Acoustics, Speech and Signal Processing*, Vol. I, March 1992.

[69] S. Seneff, "TINA: A natural language system for spoken language applications," *American Journal of Computational Linguistics*, vol. 18, pp. 61-86, 1992.

[70] C. B. Zoltowski, M. P. Harper, L. H. Jamieson, and R. Helzerman, "PARSEC: A constraint-based framework for spoken language understanding," in *International Conference on Spoken Language Processing*, Oct. 1992, pp. 249-252.

[71] M. P. Harper, L. H. Jamieson, C. B. Zoltowski, and R. A. Helzerman, "Semantics and constraint parsing of word graphs," in *1992 International Conference on Acoustics, Speech and Signal Processing*, Vol. II, April 1992, pp. 63-66.

[72] H. Maruyama, "Constraint dependency grammar," IBM, Tokyo, Japan, Tech. Rep. RT0044, 1990.

[73] H. Maruyama, "Constraint dependency grammar and its weak generative capacity," *Computer Software*, 1990.

[74] M. D. Moshier and W. C. Rounds, "On the succinctness properties of unordered context-free grammars," in *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 1987, pp. 112-115.

[75] Entropic Cambridge Research Laboratory Ltd. HTK version 3.0, 2000.

[76] S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book*. Entropic Cambridge Research Laboratory Ltd. HTK version 3.0, 2000.

APPENDICES

APPENDIX A

STIMULI FOR EXPERIMENT 1

1. a) ถ้าจะไปกินข้าวด้วยกันวันนี้ ฉันไม่เลี้ยงหรอกนะ คนจนจะแย่อยู่แล้ว

   / **k^{h}on # con** ca jêɛ jùu léɛw /

   'If you want to go to dinner with me today, you will have to pay for your own meal. I'm broke.'

   b) รัฐบาลประกาศขึ้นราคาน้ำมัน ไม่เห็นใจชาวบ้านเลย คนจนจะแย่อยู่แล้ว

   / **k^{h}oncon** ca jêɛ jùu lɛɛ́w /

   'The government announces a price hike on gasoline. They don't seem to care about working people. The poor are suffering tremendously.'

2. a) คุณปรุงรสอาหารได้อร่อยทุกอย่างเลย แต่ขอติสักนิด แกงจืดมากไปหน่อยนึง

   / **kɛɛŋ # cɯ̀ɯt** mâak paj nɔ̀ɔj nɯŋ /

   'Every dish you made tasted really good. Only the curry was a little bit too bland.'

   b) ปริมาณอาหารสำหรับงานเลี้ยงกำลังดี ยกเว้นอย่างเดียว แกงจืดมากไปหน่อยนึง

   / **kɛɛŋcɯ̀ɯt** mâak paj nɔ̀ɔj nɯŋ /

   'The amount of food for the party is about right. But, there is a little bit too much soup.'

3. a) ตาอายุมากแล้ว อย่าให้ตาชั่งของเลย ตาชั่งไม่ค่อยจะตรง

   / **taa # c^{h}âŋ** mâj k^{h}ɔ̂j ca troŋ /

   'Grandfather is quite old. Don't ask him to do any weighing job. Grandfather has difficulties weighing accurately.'

   b) รู้แล้วว่าทำไมหมู่นี้ดูเหมือนน้ำหนักขึ้นมากกว่าปกติ ตาชั่งไม่ค่อยจะตรง

   / **taac^{h}âŋ** mâj k^{h}ɔ̂j ca troŋ /

   'I've just found out why lately it seems I have gained a lot of weight. The scale is not quite accurate.'

4. a) ช่วงหน้าแล้งผู้คนอพยพออกจากเมืองไปกันเกือบหมด เมืองร้างอยู่เป็นหย่อม ๆ

   / **mɯaŋ # ráaŋ** jùu pen jɔ̀ɔm jɔ̀ɔm /

   'During the dry season, a lot of people migrate out of town. Deserted areas are scattered across the town.'

   b) ตอนไปเที่ยวทางตะวันตกเห็นเมืองร้างเต็มไปหมด เมืองร้างอยู่เป็นหย่อม ๆ

   / **mɯaŋráaŋ** jùu pen jɔ̀ɔm jɔ̀ɔm /

   'When I was vacationing in the West. I saw a lot of ghosttowns. Ghosttowns are scattered across the land.'

5. a) เมื่อคืนฝนตกหนัก บ่อปลาน้ำท่วม **ปลาไหลไปหมดแล้วละ**

   / **plaa # lăj** paj mòt léɛw lá? /
   'It rained cats and dogs last night. The fish pool overflowed. The fish have all escaped.'

   b) คนแย่งซื้อปลาที่ตลาดกันใหญ่ เหลือปลาไม่กี่ชนิดเลย **ปลาไหลไปหมดแล้วละ**

   / **plaalăj** paj mòt léɛw lá? /
   'People rush to buy fish at the market. Not many kinds of fish are left. The eels are all gone.'

6. a) ข้อสอบเติมคำที่ให้ต่อทำน่ะตรวจเสร็จแล้ว **ต่อเติมไม่ได้ความเลย**

   / **tɔ̀ɔtəəm** mâj dâjk^{h}waam ləəj /
   'I have finished grading that fill-in-the-blank exam Tor was tested on. Tor had done a poor job completing it.'

   b) บริษัทที่จ้างมาทำบ้านน่ะห่วยจริง ๆ **ต่อเติมไม่ได้ความเลย**

   / **tɔ̀ɔtəəm** mâj dâjk^{h}waam ləəj /
   'The company we hired to remodel our house was no good. They have done a poor job with those additions.'

7. a) คุณปรุงรสอาหารได้อร่อยทุกอย่างเลย แต่ขอติสักนิด **ผัดเผ็ดมากไปหน่อยนึง**

   / **p^{h}àt # p^{h}e□t** mâak paj nɔ̀ɔj nɯŋ /
   'Every dish you made tasted really good. Only stir-fry was a little bit too hot.'

   b) ปริมาณอาหารสำหรับงานเลี้ยงกำลังดี ยกเว้นอย่างเดียว **ผัดเผ็ดมากไปหน่อยนึง**

   / **p^{h}àtp^{h}e□t** mâak paj nɔ̀ɔj nɯŋ /
   'The amount of food for the party is about right. But, there is a little bit too much spicy stir-fry.'

8. a) สวนครัวชักเป็นรูปเป็นร่างขึ้นมาแล้ว ครั้งสุดท้ายที่ดู **ถั่วงอกจวนจะหมดแล้ว**

   / **t^{h}ùaŋɔ̂ɔk** cuan ca mòt léɛw /
   'Our vegetable garden is shaping up very nicely. The last time I checked, almost all of the beans have sprouted.'

   b) ถ้าคุณจะไปตลาด ฝากซื้อของด้วยนะ **ถั่วงอกจวนจะหมดแล้ว**

   / **t^{h}ùaŋɔ̂ɔk** cuan ca mòt léɛw /
   'If you plan to go to the market, pick something up for me. The bean sprouts are almost all gone.'

9. a) วันนี้ไปเที่ยวสวนสัตว์มา น่าเอ็นดูจริง ๆ **สัตว์เลี้ยงลูกอยู่ในกรง**

   / **sàtliáŋ** lûuk jùu naj kroŋ /
   'I went to the zoo today. It was really cute the way the animal cared for her young in the cage.'

   b) ไปนอนเถอะลูก ไม่ต้องเป็นห่วง **สัตว์เลี้ยงลูกอยู่ในกรง**

   / **sàtliáŋ** lûuk jùu naj kroŋ /
   'Son, go to bed. Don't worry. Your pet is in the cage.'

10. a) ต่อไปขโมยคงไม่กล้ามาอีก เราเพิ่งปักหลักกั้นรั้วใหม่ หลักแหลมดีจริง ๆ เลย

/ **làk lɛ̌ɛm** dii ciŋ ciŋ ləəj /

'From now on, no thief will dare break in. We've just put up new poles and fences. The poles are quite sharp.'

b) คุณเอาตัวรอดมาจากวิกฤตการณ์ได้ด้วยปัญญาแท้ ๆ หลักแหลมดีจริง ๆ เลย

/ **làklɛ̌ɛm** dii ciŋ ciŋ ləəj /

'You were able to come out of the crisis unscathed only because of your wit. You are very clever.'

11. a) รีบแต่งตัวเข้า เสื้อกางเกงรีดให้เรียบร้อยแล้ว เสื้อคลุมอยู่หลังเก้าอี้

/ **suîak^{h}lum** jùu lăŋ kâw?îi /

'Hurry up and get dressed. I've already ironed your shirt and pants. The shirt is on the back of the chair.'

b) อากาศข้างนอกหนาวมากนะ อย่าลืมหยิบเสื้อคลุมไปด้วย เสื้อคลุมอยู่หลังเก้าอี้

/ **suîak^{h}lum** jùu lăŋ kâw?îi /

'The weather is really cold outside. Don't forget to bring your coat. Your coat is on the back of the chair.'

12. a) มานี่หน่อยซิลูก ดูนะ ถ้าต้องต้องการใช้น้ำร้อน ลูกบิดอันทางขวามือ

/ **lûuk # bìt** ?an t^{h}aaŋ k^{h}wăa mɯɯ /

'Come here, son. Look. If you want hot water, turn the knob on the right.'

b) ตรงทางเข้ามีประตูสองประตูติดกัน ลืมบอกไป ลูกบิดอันทางขวามือ

/ **lûukbìt** ?an t^{h}aaŋ k^{h}wăa mɯɯ /

'At the entrance, there are two doors side by side. I forgot to tell you, It's the knob on the right.'

13. a) เกิดอุบัติเหตุขึ้นกับลูกวันนี้ ลูกกลิ้งตกมาจากโต๊ะ

/ **lûuk # klîŋ** tòk maa càak tó? /

'A bad accident happened to our child today. Our child fell off the table.'

b) ไม่ได้ตั้งใจวางมันไว้ที่พื้นหรอก ลูกกลิ้งตกมาจากโต๊ะ

/ **lûukklîŋ** tòk maa càak tó? /

'I didn't mean to put it on the floor. The roller fell off the table.'

14. a) ถ้ามีลูกจะเอาไปฝากไว้ให้แม่ดูแล แม่เลี้ยงเด็กสมบูรณ์ดี

/ **mɛ̂ɛ # liáŋ** dèk sŏmbuun dii /

'I want mother to take care of my kid. Mother does an excellent job of raising children.'

b) เห็นครอบครัวที่เพิ่งย้ายมาใหม่หรือยัง แม่เลี้ยงเด็กสมบูรณ์ดี

/ **mɛ̂ɛliáŋ** dèk sŏmbuun dii /

'Have you seen the family that has just moved in? The child's stepmother is a bit chubby.'

15. a) ในบรรดาญาติพี่น้อง ป้าสวยที่สุด ป้าขาวดีจังเลยนะ

/ **pâa # k^{h}ăaw** dii caŋ leej náʔ /
'Among all the relatives, my aunt is the prettiest. Aunt has a very fair complexion.'

b) วันนี้ไปบ้านคุณลุงมา เจอป้าขาวด้วย ป้าขาวดีจังเลยนะ

/ **pâak^{h}ăaw** dii caŋ leej náʔ /
'We went to visit uncle today. Aunt Khao was home, too. Aunt Khao is very kind.'

16. a) คุณเชื่อไหมว่าเกิดอะไรขึ้นวันนี้ ช้างพังบ้านคุณลุงหาญ

/ **c^{h}áaŋ # p^{h}aŋ** bâan k^{h}unluŋ hăan /
'You wouldn't believe what happened today. The elephant destroyed Uncle Hahn's house.'

b) เห็นช้างตัวนั้นไหม ช้างพังบ้านคุณลุงหาญ

/ **c^{h}áaŋp^{h}aŋ** bâan k^{h}unluŋ hăan /
'Do you see that elephant? That female elephant belongs to Uncle Hahn.'

17. a) เราไปถึงสนามม้าช้าไปหน่อย ตอนที่ไปถึงน่ะ ม้าแข่งอยู่ในสนามแล้ว

/ **máa # k^{h}èŋ** jùu naj sanăam léɛw /
'We arrive at the track too late. When we arrived, the horses had already begun racing.'

b) ฉันมีหน้าที่พาม้าแข่งไปกินหญ้าในสนาม แต่ไม่รู้ยังไง ม้าแข่งอยู่ในสนามแล้ว

/ **máak^{h}èŋ** jùu naj sanăam léɛw /
'I was supposed to take the racehorses to the field to graze. But to my surprise, the racehorses were already on the field.'

18. a) ฉันจะออกไปข้างนอกสักครู่ วานดูในครัวให้หน่อย น้ำต้มอยู่บนเตานะ

/ **náam # tôm** jùu bon taw náʔ /
'I', going to step outside for a moment. Keep an eye on the kitchen, will you? The water is boiling on the stove.'

b) ถึงเวลาต้องทานยาแล้ว หยิบน้ำให้แก้วนึงซิ น้ำต้มอยู่บนเตานะ

/ **náamtôm** jùu bon taw náʔ /
'It's time for me to take medicine. Please bring me a glass of water. Boiled water is on the stove.'

19. a) เดินบนสนามหญ้าตอนเพิ่งรดน้ำเสร็จ รองเท้าจะเปียกนะ น้ำค้างอยู่บนยอดหญ้า

/ **náam # k^{h}áaŋ** jùu bon jɔ̂ɔt jâa /
'Your shoes will get wet when walking on the lawn after it has been watered. Droplets of water still rest on the blades of grass.'

b) เดินบนสนามหญ้าตอนเช้า ๆ ระวังรองเท้าจะเปียกนะ น้ำค้างอยู่บนยอดหญ้า

/ **náamk^{h}áaŋ** jùu bon jɔ̂ɔt jâa /
'Be careful when walking on the lawn in the morning. Your shoes will get wet. The morning dew is still on the blades of grass.'

20. a) น้ำแก้วนี้น่าดื่มจัง น้ำหอมกลิ่นดอกกุหลาบ

/ **náam # hɔ̌ɔm** klìn dɔ̀ɔkkulàap /
'I want to drink this glass of water. The water has the fragrance of roses.'

b) หล่อนดีใจมากเมื่อเปิดดูห่อของขวัญที่ได้รับ น้ำหอมกลิ่นดอกกุหลาบ

/ **náamhɔ̌ɔm** klìn dɔ̀ɔkkulàap /
'She was ecstatic after opening her present. It was the tea rose perfume.'

21. a) ทำไมอาการของเขายังไม่ดีขึ้นเลย หมอดูไม่ถูกแน่ ๆ

/ **mɔ̌ɔ # duu** mâj tʰùuk nêɛ nêɛ /
'How come his condition has not improved at all? The doctor must surely have made a misdiagnosis.'

b) มันจะไม่เป็นจริงตามคำทำนายหรอก หมอดูไม่ถูกแน่ ๆ

/ **mɔ̌ɔduu** mâj tʰùuk nêɛ nêɛ /
'It will never happen as predicted. The fortuneteller is definitely wrong.'

22. a) รู้ไหมว่าเจออะไรตอนล้มตัวลงนอน เข็มกลัดอยู่บนที่นอน

/ **kʰěm # klàt** jùu bon tʰîinɔɔn /
'Do you know what I found when I lay down? The needle was on the mattress.'

b) ติดเครื่องหมายบนเครื่องแบบให้หน่อยซิ ใช้เข็มกลัดนะ เข็มกลัดอยู่บนที่นอน

/ **kʰěmklàt** jùu bon tʰîinɔɔn /
'Please put the insignia on my uniform. Use safety-pins. The safety-pins are on the bed'

23. a) ไม่ค่อยชอบกางเกงที่สั่งตัดเลย ขาสั้นน้อยกว่าที่สั่ง

/ **kʰǎa # sân** nɔ́ɔj kwàa tʰîi sàŋ /
'I don't like that tailored pants at all. The legs are not as short as I ordered.'

b) เสื้อผ้าที่เราสั่งมาขายเพิ่งจะมาถึง รู้สึกจะไม่ครบ ขาสั้นน้อยกว่าที่สั่ง

/ **kʰǎasân** nɔ́ɔj kwàa tʰîi sàŋ /
'Those clothes we ordered have just arrived. Something is missing. The number of shorts is less than we ordered.'

24. a) เพิ่งนึกได้เมื่อเช้านี้เอง ของค้างอยู่ที่ห้องเขา

/ **kʰɔ̌ɔŋ # kʰáaŋ** jùu tʰîi hɔ̂ŋ kʰáw /
'I've just realized it this morning. I left my stuff in his room.'

b) กับข้าวเหลือจากเมื่อวานตั้งเยอะ ไปเอามาหน่อย ของค้างอยู่ที่ห้องเขา

/ **kʰɔ̌ɔŋkʰáaŋ** jùu tʰîi hɔ̂ŋ kʰáw /
'There was plenty of food left from yesterday. Go get it. The leftovers are in his room.'

25. a) วันนี้ไปเที่ยวสวนสัตว์มา ยังประหลาดใจไม่หายเลย เสือขาวผิดปกติ

/ **suŭa # k^{h}ăaw** p^{h}ìt pòkatì? /

'I went to the zoo today. I'm still amazed at the tiger which has unusual white skin.'

b) ตำรวจจับเสือขาวได้แล้ว ทุกคนประหลาดใจเมื่อเห็นสภาพของเขา เสือขาวผิดปกติ

/ **suŭak^{h}ăaw** p^{h}ìt pòkatì? /

'Police has successfully apprehended notorious Khao. People were amazed when they saw him. That notorious Khao is deformed.'

APPENDIX B

STIMULI FOR EXPERIMENT 2

1. a) เมืองอยู่ไกล คนไม่ค่อยมาเมืองเลย
   / kʰon mâj kʰɔ̂ːj **maa mɯaŋ** ləəj /
   'The city is too far away. People don't usually come to the city.'
   b) เมืองเลยอยู่ไกล คนไม่ค่อยมาเมืองเลย
   / kʰon mâj kʰɔ̂ːj **maa mɯaŋ**ləəj /
   'The city of Loei is too far away. People don't usually come to Loei.'

2. a) เมืองไม่น่าอยู่ คนไม่ค่อยอยู่เมืองเลย
   / kʰon mâj kʰɔ̂ːj **jùu mɯaŋ** ləəj /
   'The city is not conducive to living. People don't usually live in the city.'
   b) เมืองเลยไม่น่าอยู่ คนไม่ค่อยอยู่เมืองเลย
   / kʰon mâj kʰɔ̂ːj **jùu mɯaŋ**ləəj /
   'The city of Loei is not conducive to living. People don't usually live in Loei.'

3. a) ในเมืองอันตราย คนไม่ค่อยเข้าเมืองเลย
   / kʰon mâj kʰɔ̂ːj **kʰâw mɯaŋ** ləəj /
   'It is dangerous to be in the city. People don't usually come into the city.'
   b) เมืองเลยอันตราย คนไม่ค่อยเข้าเมืองเลย
   / kʰon mâj kʰɔ̂ːj **kʰâw mɯaŋ**ləəj /
   'It is dangerous to be in the city of Loei. People don't usually come into Loei.'

4. a) ถึงเมืองจะแห้งแล้ง คนไม่ค่อยทิ้งเมืองเลย
   / kʰon mâj kʰɔ̂ːj **tʰiŋ́ mɯaŋ** ləəj /
   'Even though the city is hot and dry . People don't usually abandon the city.'
   b) ถึงเมืองเลยจะแห้งแล้ง คนไม่ค่อยทิ้งเมืองเลย
   / kʰon mâj kʰɔ̂ːj **tʰiŋ́ mɯaŋ**ləəj /
   'Even though the city of Loei is hot and dry. People don't usually abandon Loei.'

5. a) เมืองสกปรกจัง คนไม่ค่อยสนเมืองเลย
   / kʰon mâj kʰɔ̂ːj **sŏn mɯaŋ** ləəj /
   'The city is quite dirty. People don't care much about the city.'
   b) เมืองเลยสกปรกจัง คนไม่ค่อยสนเมืองเลย
   / kʰon mâj kʰɔ̂ːj **sŏn mɯaŋ**ləəj /
   'The city of Loei is quite dirty. People don't care much about Loei.'

6. a) คนจนจะมีปัญญาไปเริ่มกิจการอะไรได้ เขาว่าคนจนขาดทุน

   / k^{h}áw wâa k^{h}**oncon k^{h}àat** t^{h}un /

   'The poor are quite at a disadvantage when it comes to starting a business. It's often said that the poor lack funds.'

   b) อัตราภาษีเก่าไม่ยุติธรรมเลย เขาว่าคนจนขาดทุน

   / k^{h}áw wâa k^{h}**oncon k^{h}àatt^{h}**un /

   'The tax rate is really unfair. It's often said that the poor are at a loss.'

7. a) พวกลูกจ้างหยุดงานเพราะกลัวไม่ได้รับเงินค่าแรง เขาว่าคนจ่ายขาดทุน

   / k^{h}áw wâa k^{h}**oncàaj k^{h}àat** t^{h}un /

   'The workers went on strike because they are afraid of not getting paid. There's a rumor that the payer has no money.'

   b) คนถูกหวยงวดนี้กันเพียบเลย เขาว่าคนจ่ายขาดทุน

   / k^{h}áw wâa k^{h}**oncàaj k^{h}àatt^{h}**un /

   'A lot of people strike it rich with this week's lottery. There's a rumor that the payer loses.'

8. a) งานก่อสร้างต้องหยุดชะงักลงกลางครัน เขาว่าคนจ้างขาดทุน

   / k^{h}áw wâa k^{h}**oncâaŋ k^{h}àat** t^{h}un /

   'The construction is suddenly halted. There's rumor that the contractor runs out of funds.'

   b) หลังจากเจรจาตกลงกันแล้ว ดูเหมือนพวกลูกจ้างจะได้กำไร เขาว่าคนจ้างขาดทุน

   / k^{h}áw wâa k^{h}**oncâaŋ k^{h}àatt^{h}**un /

   'After the negotiation, it seems that the workers profit from the deal. The employer is at a loss'

9. a) เศรษฐกิจยังไม่กระเตื้องขึ้นเลย เขาว่าคนซื้อขาดทุน

   / k^{h}áw wâa k^{h}**onsɯ́ɯ k^{h}àat** t^{h}un /

   'The economy is still not recovered. It's often said that consumers don't have the buying power.'

   b) การค้าแบบผูกขาดไม่ดีหรอก เขาว่าคนซื้อขาดทุน

   / k^{h}áw wâa k^{h}**onsɯ́ɯ k^{h}àatt^{h}**un /

   'Monopolized business is no good. It's often said that consumers are at a loss.'

10. a) ร้านนั้นดูท่าทางขายดีนะ แต่ยังขยายกิจการไม่ได้ เขาว่าคนขายขาดทุน

    / k^{h}áw wâa k^{h}**onk^{h}ăaj k^{h}àat** t^{h}un /

    'It seems that the store's business really picks up. But, it can't quite expand. People speculate that the store lacks investment capital.'

    b) ร้านอาหารข้างบ้านปิดไปแล้ว เขาว่าคนขายขาดทุน

    / k^{h}áw wâa k^{h}**onk^{h}ăaj k^{h}àatt^{h}**un /

    'The restaurant nextdoor was closed. People say that the owner didn't make a profit.'

11. a) ไม่เชื่อกันนี่ บอกแล้วว่าฟืนไหม้ดี
/ bɔ̀ɔk lɛ́ɛw wâa **fɯɯn mâj** dii /
'See? You don't believe me. I told you logs burn really well.'
b) ไม่เชื่อกันนี่ บอกแล้วว่าฟืนไม่ดี
/ bɔ̀ɔk lɛ́ɛw wâa **fɯɯn mâj** dii /
'See? You don't believe me. I told you the log is no good.'

12. a) ไม่เชื่อกันนี่ บอกแล้วว่าถ่านไหม้ดี
/ bɔ̀ɔk lɛ́ɛw wâa **tʰàan mâj** dii /
'See? You don't believe me. I told you charcoal burns really well.'
b) ไม่เชื่อกันนี่ บอกแล้วว่าถ่านไม่ดี
/ bɔ̀ɔk lɛ́ɛw wâa **tʰàan mâj** dii /
'See? You don't believe me. I told you the battery is no good.'

13. a) ไม่เชื่อกันนี่ บอกแล้วว่าข้าวไหม้ดี
/ bɔ̀ɔk lɛ́ɛw wâa **kʰâaw mâj** dii /
'See? You don't believe me. I told you rice burns really well.'
b) ไม่เชื่อกันนี่ บอกแล้วว่าข้าวไม่ดี
/ bɔ̀ɔk lɛ́ɛw wâa **kʰâaw mâj** dii /
'See? You don't believe me. I told you the rice is no good.'

14. a) ไม่เชื่อกันนี่ บอกแล้วว่าไม้ไหม้ดี
/ bɔ̀ɔk lɛ́ɛw wâa **máaj mâj** dii /
'See? You don't believe me. I told you wood burns really well.'
b) ไม่เชื่อกันนี่ บอกแล้วว่าไม้ไม่ดี
/ bɔ̀ɔk lɛ́ɛw wâa **máaj mâj** dii /
'See? You don't believe me. I told you it's a lousy kind of wood.'

15. a) ไม่เชื่อกันนี่ บอกแล้วว่าหนังไหม้ดี
/ bɔ̀ɔk lɛ́ɛw wâa **năŋ mâj** dii /
'See? You don't believe me. I told you cow hide burns really well.'
b) ไม่เชื่อกันนี่ บอกแล้วว่าหนังไม่ดี
/ bɔ̀ɔk lɛ́ɛw wâa **năŋ mâj** dii /
'See? You don't believe me. I told you it's a lousy movie.'

16. a) น้องสาวของแนนเป็นเด็กน่ารัก คนชอบมายอน้องแนน
/ kʰon cʰɔ̂ɔp maa **jɔɔ nɔ́ɔŋ** nan /
'Nan's sister is a cute kid. Everyone loves to praise Nan's sister.'

b) น้องแนนเป็นเด็กน่ารัก คนชอบมายอน้องแนน
/ kʰon cʰɔ̂ɔp maa **jɔɔ nɔ́ɔŋ**nan /
'Nan is a cute kid. Everyone loves to praise Nan.'

17. a) น้องสาวของแนนเป็นเด็กน่ารัก **คนชอบมาแหย่น้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **jὲɛ nɔ́ɔŋ** nan /
'Nan's sister is a cute kid. Everyone loves to tease Nan's sister.'

b) น้องแนนเป็นเด็กน่ารัก **คนชอบมาแหย่น้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **jὲɛ nɔ́ɔŋ**nan /
'Nan is a cute kid. Everyone loves to tease Nan.'

18. a) น้องสาวของแนนเป็นเด็กน่ารัก **คนชอบมาเยี่ยมน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **jîâm nɔ́ɔŋ** nan /
'Nan's sister is a cute kid. Everyone loves to visit Nan's sister.'

b) น้องแนนเป็นเด็กน่ารัก **คนชอบมาเยี่ยมน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **jîâm nɔ́ɔŋ**nan /
'Nan is a cute kid. Everyone loves to visit Nan.'

19. a) น้องสาวของแนนเป็นเด็กน่ารัก **คนชอบมาล้อน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **lɔ́ɔ nɔ́ɔŋ** nan /
'Nan's sister is a cute kid. Everyone loves to tease Nan's sister.'

b) น้องแนนเป็นเด็กน่ารัก **คนชอบมาล้อน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **lɔ́ɔ nɔ́ɔŋ**nan /
'Nan is a cute kid. Everyone loves to tease Nan.'

20. a) น้องสาวของแนนเป็นเด็กน่ารัก **คนชอบมาหลงน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **lɔ̌ŋ nɔ́ɔŋ** nan /
'Nan's sister is a cute kid. Everyone seems crazy about Nan's sister.'

b) น้องแนนเป็นเด็กน่ารัก **คนชอบมาหลงน้องแนน**

/ k^{h}on c^{h}ɔ̂ɔp maa **lɔ̌ŋ nɔ́ɔŋ**nan /
'Nan is a cute kid. Everyone seems crazy about Nan.'

21. a) หมอเพิ่งออกไปข้างนอก เดี๋ยวคงกลับ **น่าจะไปรอหมอดู**

/ nâa càa paj **rɔɔ mɔ̌ɔ** duu /
'The doctor has just stepped outside. He'll be back soon. We should wait for the doctor.'

b) หมอดูเพิ่งออกไปข้างนอก เดี๋ยวคงกลับ **น่าจะไปรอหมอดู**

/ nâa càa paj **rɔɔ mɔ̌ɔ**duu /
'The fortuneteller has just stepped outside. He'll be back soon. We should wait for the fortuneteller.'

22. a) หมอสั่งยาไม่ได้เรื่องเลย **น่าจะไปด่าหมอดู**

/ nâa càa paj **dàa mɔ̌ɔ** duu /
'The doctor didn't prescribe high-potent drugs for us. We should let the doctor know how we feel.'

b) หมอดูหลอกเรานี่หว่า น่าจะไปด่าหมอดู

/ nâa càˆ paj **dàa mɔ̌ɔduu** /
'The fortuneteller is quite deceitful. We should give the fortuneteller a piece of our minds.'

23. a) โรงพยาบาลเรายังขาดหมออีกมาก น่าจะไปจ้างหมอดู

/ nâa càˆ paj **câaŋ mɔ̌ɔ** duu /
'Our hospital is still in need of doctors. We should try to hire more doctors.'

b) คดีนี้มืดมนเหลือเกิน น่าจะไปจ้างหมอดู

/ nâa càˆ paj **câaŋ mɔ̌ɔduu** /
'This case is really going nowhere. We should hire a psychic.'

24. a) หมอมีบุญคุณกับพวกเรามาก น่าจะไปเลี้ยงหมอดู

/ nâa càˆ paj **liáŋ mɔ̌ɔ** duu /
'We owe a great deal to the doctor. We should take the doctor out to dinner.'

b) จริง ๆ อย่างที่หมอดูบอกเลยว่าคุณจะได้งาน น่าจะไปเลี้ยงหมอดู

/ nâa càˆ paj **liáŋ mɔ̌ɔduu** /
'The fortuneteller is right on the money about your job prospect. We should take the fortuneteller out to celebrate.'

25. a) ลูกไม่เป็นอะไรมากแล้ว พากลับบ้านเถอะ น่าจะไปขอหมอดู

/ nâa càˆ paj **kʰɔ̌ɔ mɔ̌ɔ** duu /
'Our child is a lot better now. Let's take him home. We should ask the doctor.'

b) หวยใกล้จะออกแล้ว ยังไม่รู้จะซื้อเลขอะไรเลย น่าจะไปขอหมอดู

/ nâa càˆ paj **kʰɔ̌ɔ mɔ̌ɔduu** /
'It's about time for the lottery. I still haven't a clue to what number to play. We should ask the fortuneteller.'

APPENDIX C

STIMULI FOR EVALUATING AUTOMATIC TONE CLASSIFICATION ALGORITHM

1. หญิงงามอย่างนี้แมวมองไม่มีเมินอย่างแน่นอน

/ jǐŋ ŋaam jàaŋ níi mɛɛwmɔɔŋ mâj mii məən jàaŋ nêɛnɔɔn /
'Such a beautiful girl like this is definitely not ignored by the scout.'

2. อย่าลืมว่าเวลามันล่วงเลยมาชาวนานแล้ว

/ jàa lɯɯm wâa weelaa man lûaŋ ləəj maa jaaw naan lɛ́ɛw /
'Don't forget that time has passed for so long.'

3. งูเหลือมใหญ่เลื้อยอยู่ในหย่อมหญ้าอย่างเหนื่อยหน่าย

/ ŋuulɯ̌am jàj lɯ́aj jùu naj jɔ̀m jâa jàaŋ nɯ̀aj nàaj /
'A big python is winding tiredly in the grass.'

4. ลุงหวานยอน้าหญิงว่างามยิ่งเมื่อยามยิ้มแย้ม

/ luŋwǎan jɔɔ náa jǐŋ wâa ŋaam jîŋ mɯ̂a jaam jím jɛ́ɛm /
'Uncle Wahn praises aunt Ying for her beauty when she smile.'

5. หล่อนนั่งเหม่อลอยเมื่อโหน่งเล่าว่าหนิงยังไม่ยินยอม

/ lɔ̀ɔn nâŋ mə̀əlɔ̀ɔj mɯ̂a nòoŋ lâw wâa nǐŋ jaŋ mâj jinjɔɔm /
'A big python is winding tiredly in the grass.'

6. อย่าหวั่นไหวเมื่องานไม่ลื่นไหลเหมือนเมื่อยังหนุ่ม ๆ อยู่

/ jàa wànwǎj mɯ̂a ŋaan mâj lɯ̂ɯn lǎj mɯ̌an mɯ̂a jaŋ nòm nòm jùu /
'Don't be discouraged if work doesn't go your way like when you were young.'

7. เหน่งลิ้มลองเนื้อน้อยหน่าหนังเมื่องานเลี้ยงวันวาน

/ nèŋ límlɔɔŋ mɯ́a nɔ́ɔjnàanǎŋ mɯ̂a ŋaan liáŋ wan waan /
'Neng tasted the meat of a sugar apple at yesterday's party.'

8. นายยิ่งยงยังยุ่ง ๆ อยู่เลยไม่ลางานมาเยี่ยมหลาน ๆ

/ naaj jîŋjoŋ jaŋ jûŋjûŋ jùu lə̀əj mâj laa ŋaan maa jîam lǎan lǎnn /
'Yingyong is quite busy to take a leave of absence to come visit grandchildren.'

9. หน่อยนุนงงเมื่อนายนนท์เมามายแล้วมาลุ่มล่ามอย่างนี้

/ nɔ̀ɔj ŋunŋoŋ mɯ̂a naaj non mawmaaj lɛ́ɛw maa lûmlaâm jàaŋ nií /
'Nawj is puzzled that the drunken Non is trying to take advantage of her.'

10. นางนิ่มยืนยันว่าน้องนนนี่นำยาน้ำมันเหลืองมาเล่น

/ naaŋ nîm jɯɯnjan wâa nɔ́ɔŋ nonniî nam jaanáammanlɯ̌aŋ maa lêen /
'Nim insists that Nonni is the one playing with the yellow oil.'

11. หนูยุ้ยล้อเลียนหมอหยองว่างมงายไม่น้อยเลย

/ nǔujúuj lɔ́ɔlian mɔ̌ɔjɔ̌ɔŋ wâa ŋomŋaaj mâj nɔ́ɔj lɔ̀ɔj /
'Yuj is making fun of the fortune-teller, Yong, for being quite superstitious.'

12. ยามนี้ลุงเนื่องย่ำแย่เลยไม่น่ามีเงินเหลือ

/ jaam níi luŋnɯ̂aŋ jâmjɛ̂ɛ lɔ̀ɔj mâj nâa mii ŋɔ̀n lɯ̌a /
'For the moment, uncle Nuang is in trouble and has no money left.'

13. น้าแหววิงวอนว่าอย่าลงไม้ลงมือเลย

/ náawɛ̌ɛw wiŋwɔɔn wâa jàa loŋmájloŋmɯɯ lɔ̀ɔj /
'Aunt Waew is pleading for non-violent means.'

14. หม่อมหลวงเนื่องย้ำว่าน้องหมิวยังเยาว์วัยอยู่

/ mɔ̀ɔmlɯ̌aŋ nɯ̂aŋ jám wâa nɔ́ɔŋmǐw jaŋ jaw waj jùu /
'M.L. Nuang emphasizes that Mew is still quite young.'

15. น้องแมวไม่ยอมเล่าว่าลืมแหวนวงนั้นไว้ไหน

/ nɔ́ɔŋmɛɛw mâj joom lâw wâa lɯɯm wɛ̌ɛn woŋ nán waj nǎj /
'Maew refuses to tell where she misplaced that ring.'

16. วันนี้นั้นย่อมไม่เหมือนวันวานอย่างแน่นอนเลย

/ wanníi nán jɔ̂ɔm mâj mɯ̌an wanwaan jàaŋ nɛ̂ɛnɔɔn lɔ̀ɔj /
'Today is certainly not the same as yesterday.'

17. น้าหม่องแหย่น้องแหม่มว่าหน้าไม่เหมือนแม่นุ่นเลย

/ náamɔ̀ɔŋ jɛ̀ɛ nɔ́ɔŋmɛ̀m wâa nâa mâj mɯ̌an mɛ̂ɛnûn lɔ̀ɔj /
'Uncle Mong teases Mam for her unresemblance to her mother, Nuun .'

18. น้องแมนงอนแม่เลี้ยงเลยหนีมานั่งนิ่ง ๆ

/ nɔ́ɔŋman jɔɔn mɛ̂ɛliáŋ lɔ̀ɔj nǐi maa nâŋ nîŋnîŋ /
'Man was upset at his stepmother and ran from her to sit quietly alone.'

19. หมอเล่าว่าหมู่มั่นเหนื่อยง่ายแม้เวลาวิ่งเหยาะ ๆ

/ mɔ̌ɔ lâw wâa mùu mân nɯ̀aj jâaj mɛ́ɛ weelaa wîŋ jɔ̀ʔjɔ̀ʔ /
'The doctor says that corporal Mun gets tired easily even when lightly jogging.'

20. ยายเมี้ยนนำเนื้อวัวมาย่างไว้ยำเย็นวันนี้

/ jaajmián nam nɯ́awua maa jâaŋ wáj jam jen wannií /
'Old Mian grilled beef for making tonight's salad.'

21. หลวงลุงนำหมูหยองในย่ามมาโยนไว้ในหลุม

/ lŭaŋluŋ nam mŭujɔ̆ŋ naj jâam maa joon wáj naj lŭm /
'The old monk threw shredded fried pork into the hole.'

22. มาลินลืมล้อมหมูไว้ในเล้าเมื่อเย็นวานนี้

/ maalin lɯɯm lɔ̂ɔm mŭu wáj naj láw mɯa jen waanníi /
'Malin forgot to herd her pigs into the pen yesterday evening.'

23. เชิญท่องเที่ยวทั่วถิ่นแคว้นแดนไทยไปกับทัวร์เอื้องหลวง

/ cʰɔɔn tʰɔ̂ŋtʰîaw tʰûa tʰìn kʰwɛ́ɛn dɛɛn tʰaj paj kàp tʰua ʔɯ̂aŋlŭaŋ /
'Come visit every inch of Thailand with the Royal Orchid Tour.'

24. ไอทีวีมีเรื่องราวหลากหลายให้ได้ชมกันทุกวัน

/ ʔajtʰiiwii mii rɯ̂aŋraaw làaklăaj hâj dâaj cʰom kan tʰúk wan /
'ITV offers a wide variety of programs for our viewing pleasure everyday.'

25. ทุกคนชื่นชมคนซื่อสัตย์เฉกเช่นชายชื่อชวน

/ tʰúkkʰon cʰɯ̂ɯncʰom kʰon sɯ̂ɯsàt cʰèk cʰên cʰaaj cʰɯ̂ɯ cʰuan /
'Everyone admires an honest person like Mr. Chuan.'

26. แชมพูสมุนไพรช่วยบำรุงผมให้กลับนุ่มเงางามได้

/ cʰɛɛmpuu samŭnpraj cʰûaj bamruŋ pʰŏm hâj klàp ŋawŋaam dâaj /
'Herbal shampoo helps revitalize your hair for soft and silky feel.'

27. รายการเหลียวหลังแลหน้ากล้าเจาะลึกประเด็นข่าวสำคัญ

/ raajkaan lǐawlăŋlɛɛnâa klâa cɔ̀ʔ lúk pràden kʰàaw săŋkʰan /
'The program "Glance Back and Look Ahead" dares to probe important issues.'

28. เชิญแวะมาชิมอาหารหลากรสได้ที่ร้านซุ้มสามสาว

/ cʰɔɔn wɛ́ʔ maa cʰim ʔaahăan làak rój dâaj tʰî̂i ráan súmsăamsăaw /
'You're welcome to sample many delicious dishes at Sam Sao restaurant.'

29. แม่ เตือนน้องข่ามว่าอย่าแข่นเขี้ยวเคี้ยวฟันเมื่อยามโกรธ

/ mɛ̂ɛ tɯan nɔ́ɔŋkʰàam wâa jàa kʰènkʰîawkʰiáwfan mɯ̂a jaam kròt /
'Mother warns Kham not to grind her teeth when angry.'

30. กับแกล้มบ้านลุงโกร่งพอจะกล้อมแกล้มไปได้บ้าง

/ kàpklɛ̂ɛm bâan luŋkrôŋ pʰɔɔ càʔ klɔ̂mkɬɛ̂ɛm paj dâaj bâaŋ /
'Appetizer at Uncle Krong's house will do.'

31. ไม่ควรด่วนตัดสินใจไขว่คว้าหาคู่ครองเมื่อยังเด็กอยู่

/ mâj kʰwuan dùan tàtsǐncaj kʰwàjkʰwáa hăa kʰûukʰrɔɔŋ mɯ̂a jaŋ dèk jùu /
'Don't decide to get marry at a young age.'

32. อรอนงค์ออกอาการอึดอัดเมื่ออ๊อดเอ่ยปากชวน

/ ʔɔɔnʔanooŋ ʔɔ̀ɔk ʔaakaan ʔɯ̀tʔàt mɯ̂a ʔɔ́ɔt ʔə̀əj paak cʰuan /
'Awn-anong felt uncomfortable when Aut invited her.'

33. น้องต้องตวาดหนูตุ๊ดตู่ว่าเดินต้วมเตี้ยมเป็นเต่าเลย

/ nɔ́ɔŋtɔ̂ŋ tawàat nŭu túttùu wâa dəən tûamtîam pen tàw ləəj /
'Tong snaps at Toot-too for walking so slow like a turtle.'

34. เธอควรเอาพระบนหิ้งมาห้อยเผื่อเวลาเคราะห์หามยามร้าย

/ tʰəə kʰuan ʔaw pʰráʔ bon hîŋ maa hɔ̂ɔj pʰɯ̀a weelaa kʰrɔ́ʔ hăam jaam ráaj /
'You should take Buddha images from the shelf and wear in case of bad luck.'

35. ความวัวยังไม่ทันหาย ความควายดันเข้ามาแทรก

/ kʰwaamwua jaŋ mâj tʰan hăj kʰwaamkʰwaaj dan kʰâw maa sêɛk /
'No sooner had one bad thing subsides than the occurrence of the other.'