Obstruent Consonant Landmark Detection in Thai Continuous Speech

Siripong Potisuk
Department of Electrical and Computer Engineering, The Citadel, Charleston, SC 29409, USA
Email: siripong.potisuk@citadel.edu

***Abstract*—The presence of obstruent consonants constitutes key landmark events with cues that indicate abrupt acoustic discontinuities in the speech signal. Such discontinuities allow further analysis and recognition to be performed in knowledge-based speech recognition systems. This paper describes an acoustical investigation on Thai obstruent consonant detection using average level crossing rate (ALCR) information. Simple and easy to compute, ALCR information alone was successfully used in an automatic speech segmentation system for English. Comparable and, in some cases, slightly better performance than the spectral-domain methods using the Mel frequency cepstrum coefficients (MFCC) was reported. However, ALCR has never been applied to Thai. As a result, the objective of the study is to apply ALCR information to ascertain its usefulness in detecting significant temporal changes involving obstruent consonants in Thai continuous speech. Preliminary results suggest that ALCR and RMS energy can be combined to detect the phonetic boundary between initial obstruent consonant and preceding/following vowel or final consonant of the preceding syllable. An experiment was conducted on a small speech corpus containing 21 sentences designed to highlight the occurrences of all 21 possible leading consonants in various syllable structures. The overall detection rate is 83.5% for data from four speakers. The proposed method also reduces the insertion error due to amplitude variations within a phonetic segment.**

***Index Terms*—average level crossing rate, automatic speech segmentation, Thai obstruent detection**

I. Introduction

In automatic speech recognition systems, the goal is to uncover a possible sequence of words from a given speech signal. To date, the most successful system is based on HMM phone models, and the training of such acoustic models depends heavily on large annotated speech corpora. HMM-based algorithm is very computationally intensive and requires a large amount of training data. This is very prohibitive in extending the approach to under-resourced languages since no significantly large speech corpus exists to provide training data for the HMM acoustic models. Alternatively, Stevens [1] has proposed a knowledge-based speech recognition system which aims at uncovering a word sequence by utilizing the knowledge about acoustic landmarks and distinctive features determined from the input speech signal. The system consists of three modules: landmark detection, feature extraction, and sentence reconstruction. Three types of acoustic landmarks (e.g., consonant, vowel and glide) are first located in terms of time positions or boundaries. Then, a sequence of distinctive features bundles, from which a hypothesized word sequence is derived, is estimated around those boundaries. The final sentence reconstruction involves the Lexical Access from Features (LAFF) model based on human auditory processing of speech.

The proposed system described above is considered a segment-based approach as opposed to the typical frame-based approach employed by systems with HMM-based acoustic models and n-gram language models. The frame-based approach is more computationally intensive because it treats each consecutive frame of the speech signal independently, whereas the segment-based approach selectively focuses on those near landmarks where acoustic discontinuities occur. It is generally agreed that phonetic information, such as the place and manner of articulation of segments, is not distributed uniformly across an utterance. And, a large amount of such information is concentrated at or near acoustic landmarks.

Landmark detection process usually involves some types of automatic segmentation algorithm depending on the type of landmark being detected. Roughly speaking, automatic speech segmentation systems can be separated into two categories: implicit (blind) and explicit segmentation [2]. First, in the implicit case, the segmentation algorithm is designed without any prior linguistic knowledge about the phone sequence of the input speech signal. On the other hand, the design of the explicit type of segmentation algorithm relies on the linguistic knowledge associated with the input speech signal, such as its phonetic transcription or the knowledge of its phoneme sequence and hence the number of phonemes present. Thus, the system is only required to optimally locate the boundary locations that best coincide with the phoneme sequence given.

In this paper, the focus is on blind or implicit detection of consonant landmark, especially the detection of obstruent consonants (stops or plosives, fricatives and affricates). Obstruent consonants are those that get produced by partially or completely blocking the airflow from the lungs through the oral cavity. For example, a signature acoustic characteristic of stop consonants is the complete obstruction (i.e., closure) followed by sharp

Manuscript received April 30, 2015; revised July 14, 2015.

doi: 10.18178/ijsps.4.3.214-219

release of energy while fricatives are produced by partial occlusion of the airstream resulting in a noise-like turbulence. Affricates are composite speech sounds similar to a stop consonant gradually released with audible friction like a fricative. In terms of spectral behavior, obstruent consonants are very transient and dynamical.

II. Research Motivation

Obstruent detection problem has been studied over the past decade for various languages [3], [4] and [5]. In [6], the energy of six different frequency bands was calculated from the broadband spectrogram of the speech signal and abrupt change in the amplitude of each band energy is located with a two-pass peak-picking algorithm. Recently in 2014, Vachhani, Malde, Madhavi, and Patil [7] proposed a novel approach to the design of an automatic obstruent detection of English using spectral transition measure (STM) based on the Mel frequency cepstrum coefficients (MFCC) feature. The method does not take into account any prior information, such as phonetic sequence, speech transcription, and/or number of obstruents. The detection efficiency and estimated probability are around 77% and 0.77, respectively with a 30-ms duration agreement criterion and 0.4 STM threshold setting.

Our obstruent consonant landmark detection system for Thai continuous speech is based on the first type (implicit or blind case) of the segmentation procedures described in section I above. The system tries to detect the phonetic boundary between initial obstruent consonant and preceding/following vowel or final consonant of the preceding syllable. It is worth noting that HMM models are not used despite high degree of segmentation accuracy. Because Thai is an under-resourced language, no significantly large speech corpus exists to provide training data for the HMM acoustic models. Since the use of HMM approach is not chosen, it is imperative that a simpler alternative method that is capable of detecting acoustic boundaries from the input speech be found.

Segmentation algorithms based on time-domain features called average level crossing rate (ALCR) and Extrema-based signal track length (ESTL) were investigated in [8], [9]. Simple and easy to compute, ALCR information was successfully used in an automatic speech segmentation system for English. Comparable and, in some cases, slightly better performance than the spectral-domain methods using MFCC's was reported. Accuracy around 80% was reported with computational time reduction by approximately 75% compared with that of spectral domain methods. As for Thai speech segmentation, several types of energy calculations, such as absolute, root-mean-square, square, Teager, and modified Teager energy, were used for syllable segmentation of Thai connected speech based on the local extrema of those energy contours [10]. Recently in [11], 1-D stationary wavelet transform (SWT) whose detail coefficient component representing the high frequency part of input signal was used to analyze the boundaries of syllables. In [12], Theera-Umpon, Chansareewittaya, and Auephanwiriyakul proposed a phoneme recognition system with soft phoneme segmentation using discrete hidden Markov models based on the Mel frequency with perceptual linear prediction and the Mel frequency cepstrum coefficients. After a careful literature review, the use of temporal-based features was not found to have been utilized in Thai speech segmentation. Thus, following [8], our objective is to apply ALCR information to ascertain its usefulness in detecting significant temporal changes in Thai continuous speech in this paper. A novel method of combining ALCR and RMS Energy measures to improve segmentation performance is also proposed.

III. ALCR Description

In this section, the computation of ALCR and some observations on the characteristics of the ALCR contours will be described.

A. Front-End Preprocessing

The input speech signal was sampled at 22050 samples/s and quantized to 16 bits/sample. The resulting discrete-time sequence was then normalized with respect to its maximum to lie within [-1, 1] and then made zero mean to get rid of a DC offset. The sequence was then pre-emphasized with a first-order low-pass filter according to the following LCCDE:

$$y[n]=x[n]-0.95x[n-1] \quad (1)$$

B. ALCR Feature Extraction

Following [8], the level crossing rate (LCR) at any sample point and for a certain level is defined as the total of all the level crossings that have occurred for that level over a short interval around that sample point divided by the interval duration. Then, the average level crossing rate (ALCR) at each sample point was obtained by summing LCR's over all levels.

Let $x(t)$ be a continuous-time signal. For a set of pre-defined levels, η_j, $1 \le j \le J$, where J is a total number of levels based on the signal amplitude dynamic range, the signal $x(t)$ is said to have crossed the level η_j at a given time instant τ if $x(\tau)=\eta_j$ and $x(\tau-\Delta)<\eta_j<x(\tau+\Delta)$ or $x(\tau-\Delta)>\eta_j>x(\tau+\Delta)$ for an infinitesimally small duration of time Δ. If the above condition is true, a level crossing indicator function is defined as $l(j,t)=\delta_j(t-\tau)$ and the level crossing rate for level η_j over a time interval (t_1, t_2) can be computed by

$$L(j,t_1,t_2)=\frac{1}{(t_2-t_1)}\int_{t_1}^{t_2} l(j,t)\,dt \quad (2)$$

By summing over all levels, the ALCR is given by

$$ALCR(t_1,t_2)=\frac{1}{(t_2-t_1)}\sum_{j=1}^{J}\left[\int_{t_1}^{t_2} l(j,t)\,dt\right] \quad (3)$$

By the same token, consider a discrete-time signal, $x[n]$. Mathematically, for a given sample and level η_j, a level crossing indicator function $l(j, n)$ occurs between $x[n-1]$ and $x[n]$ if the following condition is true:

$$l(j,n) = \begin{cases} 1, & (x[n]-\eta_j)(x[n-1]-\eta_j) \le 0 \\ 0 & otherwise \end{cases} \quad (4)$$

The level crossing rate (LCR) for level η_j over the interval $[n_1, n_2]$ is then expressed as

$$L(j,n_1,n_2) = \frac{1}{n_2 - n_1} \sum_{i=n_1}^{n_2} l(j,i) \quad (5)$$

And, the expression for the average level crossing rate (ALCR) over all possible levels is given as

$$ALCR(n_1,n_2) = \sum_{j=1}^{J} L(j,n_1,n_2) \quad (6)$$

To compute the level crossing rate, the distribution of levels can be constructed in either a uniform or non-uniform manner. In this study, our speech samples were recorded in a quite environment. Thus, it is safe to assume that the signal-to-noise ratio (SNR) is high, and, consequently, a high number of levels (80 levels) was used with a uniform distribution of levels within the normalized dynamic range of the signal between [-1, 1]. Regarding the length of the averaging interval (n_2-n_1=2Δ), it is recommended that the interval should be chosen such that 2Δ ≈ one pitch period. Since this is just an approximation, no accurate pitch extraction is required. For a male voice, the pitch typically ranges from 60 to 150Hz. Thus, for a sampling rate of 22050Hz, one pitch period corresponds to 147 to 368 samples. For a female voice, the pitch typically ranges from 200 to 300Hz. Thus, for a sampling rate of 22050Hz, one pitch period corresponds to 74 to 110 samples. For convenience, the interval length is kept constant at 400 samples.

It is noteworthy that ALCR can be calculated for every sample location of the input speech sequence. That is, ALCR over a certain interval of samples can be computed using an analysis window of a given length and advancing across the input speech signal by one sample at a time. In term of frame-based processing, instead of the usual frame step of 10ms, the amount of the frame step of the analysis window is only one sample, which is equal to the length of the sampling interval (e.g. 0.045ms for a sampling rate of 22050Hz). This fact helps increase segmentation accuracy (0.045ms vs 10ms) because the resolution of the automatic method is now the same as that of the manual method, which is at the sampling step of 0.045ms.

The side effect of choosing a small frame step of one sample is that the resulting ALCR contour is very choppy containing several spurious minima and maxima. This can be remedied by moving average filtering the contour. To avoid too much smoothing, the window length is chosen to be 201 samples, 100 samples on either side of the current value of the ALCR contour.

From the plot of an ALCR contour, it is observed that the range of the magnitude of the contour depends on the normalized amplitude of the input speech signal, i.e., on the recording level. Fig. 1 shows three plots of ALCR contours of the same utterance whose magnitude was uniformly scaled by 0.5 and 0.25 times the that of the original signal. It is interesting to note that while the magnitude ranges are different, the overall shape of the contours does not basically change. Only subtle differences can be detected. This implies that any phoneme boundary demarcation process should be designed to be insensitive to these differences. This means that one should normalize the ALCR contour with respect to its maximum so that it lies within [0, 1].

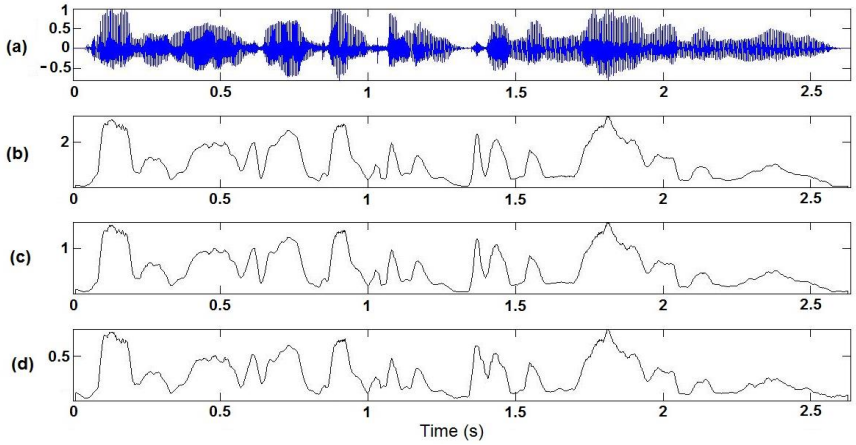

Figure 1. Different ALCR contours of an input speech signal whose normalized amplitude is scaled by 0.5 and 0.25 times the original amplitude. (a) A plot of speech signal; (b) ALCR contour without amplitude scaling; (c) ALCR contour with amplitude scaling by 0.5; and (d) ACLR contour with amplitude scaling by 0.25

C. Boundary Demarcation Process

In [9], it is reported that ALCR magnitude is directly proportional to the product between amplitude and frequency of a given sinusoidal signal. Since the speech signal can be thought of as a combination of sinusoid of different amplitudes and frequencies, temporal changes from one phoneme to the next occur with substantial changes in amplitude and frequency. By noting the points of change in ALCR curve expressed through its valleys,

the boundary between phonemes can be detected by searching for the locations of valleys or local minima of the ALCR curve. This task was proven difficult, and the authors recommended setting a threshold for picking valleys based on speech properties to avoid insertions and deletions.

In this study, it is proposed that the obstruent detection process be performed on the difference between the ALCR and the RMS energy contours by detecting the zero-crossings of the difference contour. The threshold-based method was not chosen because fixed thresholds limit the algorithm's ability to capture the range of phonetic variation. The next section describes segmentation results for obstruents detection and the rationale behind the boundary demarcation process.

IV. Experiment and Results

A. Speech Materials

The speech corpus contains 21 sentences (listed in appendix A) designed to highlight the occurrences of all 21 possible leading consonants in various syllable structures. Within those 21 sentences, there are a total of 144 obstruents distributed as follows: 19 fricatives, 26 affricates, 39 voiceless unaspirated stops, 45 voiceless aspirated stops, and 15 voiced stops. The algorithm was tested on 84 (21 sentences × 4 speakers) connected speech utterances read by 2 male and 2 female speakers in the 22-35 age range. All subjects were mono-dialectal speakers of Standard Thai. They were free of any speech or hearing disorders by self-report based on a screening interview and as later judged by the investigator during the recording session.

Recordings were made in a quiet office using recording feature of the Microsoft freeware 'Speech Analyzer' version 3.1.0 installed on a Dell Latitude laptop computer. The digitization is at a sampling rate of 22050Hz by means of a 16-bit mono A/D converter. Speakers were seated and wore a regular Logitech computer headset with microphone maintained at a distance of 5cm from the lips. Each speaker was asked to read those 21 sentences once at their conversational speaking rate. Before the recording session began, the speakers were allowed to familiarize themselves with the sentences. Each session lasted about 20 minutes.

B. Analysis Results

For obstruents detection, it is observed that the magnitude of the ALCR contour during an obstruent portion of the speech signal tends to be higher than the magnitude of the RMS energy contour of the same segment. It makes sense because for obstruents, its frequency is high while the amplitude is low. As mentioned before, ALCR magnitude is directly proportional to both amplitude and frequency of the input speech signal. On the other hand, RMS energy calculation depends on the amplitude only, i.e.

$$E_{RMS}(j, n_1, n_2) = \sqrt{\frac{1}{n_2 - n_1} \sum_{i=j-n_1}^{j+n_2} x_j^2[i]} \qquad (7)$$

where $x_j[.]$ is the jth sample of the speech sequence x.

Based on the observation above, obstruent detection was performed on the difference between the ALCR and the RMS energy contours by detecting the zero-crossings of the difference contour. Those zero-crossing points correspond to the points where changes in amplitude of the speech signal is more pronounced indicating a transition from one segment to the other adjacent segment. In other words, these points occur at the crossing of the ALCR and the RMS energy contours.

Fig. 2 shows three plots representing various calculations and boundary detection results for sentence no. 6 of the appendix A. From the middle plot, one can see that the magnitude of the ALCR contour is higher than that of the RMS energy contour during the '$\mathbf{c^h}$' segments of the speech signal, which is the initial consonant of every syllable. On the other hand, the magnitude of the ALCR contour is lower than that of the RMS energy contour during the rhyme portion of the syllable, i.e., the portion containing the nucleus (vowel) and the final consonant. The difference between these two contours shown superimposed on the speech signal clearly indicates distinct boundaries between the affricate and the preceding/following sound segment. The bottom plot shows the difference contour resulting from subtracting the RMS energy contour from the ALCR contour shown in the middle plot. The bottom plot also shows the segment boundaries demarcating the affricate, '$\mathbf{c^h}$', sounds, which are obtained by locating the zero-crossings of the difference contour. For each '$\mathbf{c^h}$' segment, the boundary on its right side demarcates the boundary between the initial consonant and the vowel of that syllable whereas the one on its left side demarcates the boundary between itself and the final consonant of the previous syllable.

It is important to note again that the horizontal axis does not represent the frame index because the calculations are done with a frame step of one sample as previously mentioned in Section III. In other words, the frame index coincides with the sample index. This is done to increase segmentation accuracy and to facilitate the comparison with manual segmentation.

To measure performance of the proposed method, percent detection efficiency is computed. Detection efficiency is defined as the ratio of the total number of obstruents detected to the total number of 576 (144×4 speakers) obstruents contained in the 84 test utterances. In addition, a detected boundary is considered to match with that from manual segmentation if they are within 20-ms duration agreement of each other. The insertion error due to amplitude variation within a phonetic segment was eliminated. The insertion error is primarily caused by a shallow dip in the ALCR contour (solid contour) of a segment (see the middle plot of Fig. 2 at 0.9-1.0 sec, 1.8-2.0 sec, 2.7-2.8 sec and 3.0-3.1 sec). A 90.3% detection rate was obtained in detecting fricatives, affricates, and voiceless aspirated stops. However, the detection rate for voiced and voiceless unaspirated stops reduces significantly to 72.2%. The overall detection rate is 83.5% for data from the four speakers.

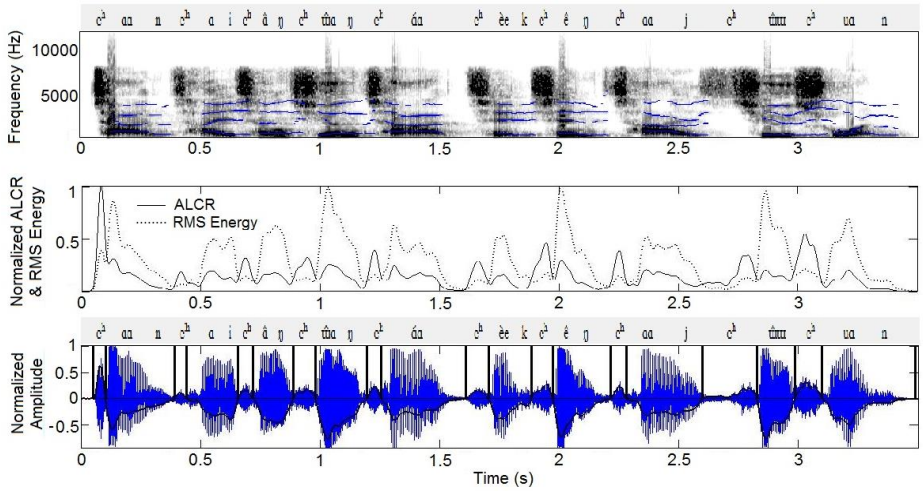

Figure 2. (Top) Spectrogram of sentence #6 along with segment-by-segment phonemic transcription; (middle) comparison of superimposed solid ALCR and dotted RMS Energy contours; and (bottom) the difference contour between ALCR and RMS energy contours superimposed on the speech signal, along with the segment boundaries demarcating the c^h segment obtained from locating zero-crossings of the difference contour.

V. Conclusions and Future Directions

Preliminary results on the application of ALCR information and RMS energy to ascertain their combined usefulness in detecting significant temporal changes in Thai continuous speech have been presented. The results suggest that their difference can be used to detect the phonetic boundary between consonant and vowel as well as between some consonants. The proposed algorithm is based on the characteristic property of ALCR that its magnitude is directly related to the product between amplitude and frequency of the speech samples. In particular, ALCR can be reliably used to detect speech landmark based on the presence of Thai obstruents when combined with the RMSE energy feature. However, their difference fails as an effective acoustic feature for detecting the boundary between voiced/voiceless unaspirated stops and vowel as can be seen from a low detection rate. Results from previous acoustical experiments suggest that formant transition between them can be used to identify the manner and place of articulation of these stops. Thus, it is believed that a successful detection scheme must incorporate both temporal and spectral domain features in order to significantly improve performance.

As a final note, the next phase of this research is to continue assessing performance of this detection algorithm using a larger group of speakers (more than 20 spekers) in order to be certain of the effectiveness of the method. Furthermore, although the focus of this paper is on the application of the algorithm to Thai continuous speech, our goal is to attempt to extend the method to other languages. In particular, ongoing experiment is being conducted with American English using utterances from the standard TIMIT speech database. So far, preliminary results are very promising.

Acknowledgement

This research is supported in part by a faculty research grant from the Citadel Foundation. The author wishes to acknowledge the assistance and support from Mrs. Suratana Trinratana, Vice President and Chief Operation Officer, and her staff of the Toyo-Thai Corporation Public Company Limited, Bangkok Thailand, during the speech data collection process. The author would also like to thank the Citadel Foundation for its financial support in the form of a research presentation grant.

Appendix Speech Stimuli

The following is a list of utterances comprising the speech materials used in the experiment. Phonemic transcription and English translation accompany each of the Thai sentences. They are designed to highlight the occurrences of all 21 possible leading consonants in various syllable structures. Although Thai has five contrastive tones, no attempt was made to account for tone distribution.

1. สมศรีอยากสักลายเสือที่แขนข้างขวา
   sŏmsǐɪ jàak sàk laaɪ sŭa $t^hîɪ$ $k^hɛ̌$ɛn $k^hâa$ŋ $k^hwǎa$
   'SomSri wants to tattoo a tiger pattern on her right arm.'
2. อรอนงค์ออกอาการอึดอัดเมื่ออ๊อดออดอ้อน
   ʔɔɔnʔanoŋ ʔɔɔk ʔaakaan ʔùtʔàt mûa ʔɔ́t ʔɔɔtʔɔ̂ɔn
   Orn-anong felt uncomfortable when Aut pleaded with her.
3. แว่วคำหวานว่าวันวานยังหวานอยู่
   wɛ̂ɛʊ k^ham wǎan wâa wan waan jaŋ wǎan jʊ̀ʊ
   '(I) vaguely hear the sweet talk that things are still as sweet as yesterday.'
4. กู้เกียรติเป็นคนกร้าวแกร่งกว่าที่คิด
   kôʊkìat pen k^hon kraâʊ krèŋ kwàa $t^hîɪ$ $k^hít$
   'Kookiat is tougher than I thought.'
5. ทหารถอดพระที่ห้อยคอไว้บนหิ้ง
   t^hahaǎan $t^hɔ̀ɔt$ $p^hrá$ $t^hîɪ$ hɔ̂ɪ $k^hɔɔ$ wáɪ bon hîŋ
   The soldier took off a neck chain of amulets and put it on the shelf.'
6. ชาญชัยช่างเชื่องช้าเฉกเช่นชายชื่อชวน
   $c^haanc^haɪ$ $c^hâŋ$ $c^hûaŋc^ʰaa$ $c^hèek$ $c^hên$ $c^haaɪ$ $c^hûɯ$ $c^hʊʊan$
   'ChanChai was as slow as the man named Chuan.'

7. ยายเมี้ยนย่างเนื้อวัวไว้ยำเย็นนี้
jaaɪ mían jâaŋ nɯ́a wáɪ jaam jen níɪ
'Grandma Mian is roasting some beef for making salad tonight.'
8. ท่องเที่ยวทั่วไทยไปกับบริษัททัวร์สยาม
tʰɔ̂ŋtʰîaʊ tʰôa tʰaɪ paɪ kàp tʰʊa sajǎam
'Travel all over Thailand with Siam Tour Company.'
9. สถานบริบาลเด็กอยู่แถวบ้านที่บางบัว
sàtʰǎanbarɪbaan dèklék jùʊ tʰɛ̌ʊ bâan tʰîɪ baaŋbʊa
'The daycare center is located in the vicinity of our BangBua home.'
10. รายการ คุยคุ้ยข่าว คนไม่ค่อยชอบดู
raaɪkaan kʰʊɪkʰɔ́ɪkʰàaɪ kʰon mâɪ kʰɔ̂ɪ cʰɔ̂ɔp dʊʊ
'People don't quite like to watch the *KhuiKhuiKhaw* show.'
11. เฟื่องฟ้าชอบแดดจัดแต่ไม่ชอบฝน
fûaŋfáa cʰɔ̂ɔp dɛ̀ɛt càt tɛ̀ɛ mâɪ cʰɔ̂ɔp fŏn
Bougainvillea likes full sun, but not rain.'
12. จันจิราเป็นคนขี้จุ้จิกเอาแต่ใจ
cancìraa pen kʰon côʊcîɪ còkcìk ʔaʊ tɛ̀ɛ caɪ
'Chanchira is a self-centered and nitpicking person.'
13. เด็กๆ ตึงดันจะดูแลเด่นคนัยด้วยกัน
dèkdèk dɯŋdan cà dʊʊlɛɛ dèndànaɪ dʊâɪ kan
'Children insisted on taking care of Dendanai together.'
14. ตฤณตรอมใจเพราะมองว่าตัวเองต่ำต้อย
trɪn trɔ̀ɔm caɪ pʰrɔ́ mɔɔŋ wâa tʊa ʔeeŋ tàmtɔ̂ɪ
Trin becomes depressed because of his inferiority complex.'
15. ป้าแป้นถูกส่งไปประจำที่ด่านปอยเปต
pâa pên tʰɔ̀ʊk sòŋ paɪ pràcam tʰîɪ dàan pɔɔɪpèet
'Aunt Pan is assigned a post at the Poipet border control.'
16. ลุงหาญชอบล้อลลิตาเรื่องลูกลี้ลูกลน
lʊŋ hǎan cʰɔ̂ɔp lɔ́ɔ lálítaa rûaŋ lóklíɪlóklon
'Uncle Hahn likes to tease Lolita's clumsiness.'
17. เรารู้สึกรันทดหดหู่เมื่อได้ยินเรื่องราวร้ายๆ
raʊ rʊ́ʊsʊ̀ʊk rantʰót hòthɔ̀ʊ mûa dâɪ jɪn rûaŋraaʊ ráɪ rái
'We feel depressed and dejected when hearing terrible news.'
18. แพรวพรรณขึ้นทอดผ้าที่งานศพพวงพยอม
pʰrɛɛʊpʰan kʰûn tʰɔ̂ɔt pʰâa tʰîɪ ŋaan sòp pʰʊʊaŋ pʰajɔɔm
'Phraewphan laid a yellow monk robe at Phuangphayom's funeral.
19. นลินีชอบทั้งขนมจีนน้ำยาและน้ำเงี้ยว
nálínɪɪ cʰɔ̂ɔp tʰáŋ kʰànɔ̌mcɪɪn náamjaa lɛ́ náamŋɪáʊ
'Nalinee likes both Namya and Nam-ngiaw rice noodle dish.'
20. เธอองอแงเพราะหงุดหงิดที่ไม่มีใครงอนง้อ
tʰəə ŋɔɔŋɛɛ pʰrɔ́ ŋʊ̀tŋ̀ɪt tʰîɪ mâɪ mɪɪ kʰraɪ ŋɔɔnŋɔ́ɔ
She made a big fuss because everyone ignores her.'
21. เขาซื้อมะม่วงมันมามากมายเมื่อตอนปีใหม่
kʰǎʊ sɯ́ɯ mámôaŋman maa mâakmaaɪ mûa tɔɔn pɪɪ màɪ
'He bought a lot of nutty-flavored mangoes during the New Year's celebration.

REFERENCES

[1] K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," *Journal of Acoustical Society of America*, vol. 111, no. 4, pp. 1872-1891, Apr. 2002.
[2] A. Acero, "The role of phoneticians in speech technology," in *European Studies in Phonetics and Speech Communication*, G. Bloothooft, V. Hazan, D. Huber, and J. Llisterri, Eds., OTS Publications, 1995.
[3] K. T. Sung and H. C. Wang, "A study of knowledge-based features for obstruent detection and classification in continuous Mandarin Speech," in *Proc. 5th International Symposium on Chinese Language Processing (ISCSLP*), Heidelberg, 2006, pp. 95-105.
[4] J. Hoelterhoff and H. Reetz, "Acoustic cues discriminating German obstruents in place and manner of articulation," *Journal of Acoustical Society of America*, vol. 121, no. 2, pp. 1142-1156, 2007.
[5] J. I. Lee and J. Y. Choi, "Detection of obstruent consonant landmark for knowledge-based speech recognition system," *Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2417-2421, 2008.
[6] C. Park, "Consonant landmark detection for speech recognition," Ph.D. dissertation, Dept. Elect. Eng. & Comp. Sci., MIT, Cambridge, MA, 2008.
[7] B. B. Vachhani, K. D. Malde, M. C. Madhavi, and H. A. Patil, "A spectral transition measure based MELCEPS-TRAL features for obstruent detection," in *Proc. International Conference on Asian Language Processing (IALP)*, Sarawak, Malaysia, 2014, pp. 50-53.
[8] A. A. Sarkar and T. V. Sreenivas, "Automatic speech segmentation using average level crossing information," in *Proc. ICASSP*, Mar. 2005, pp. I-397-I-400.
[9] P. K. Ghosh, A. A. Sarkar, and T. V. Sreenivas. (Oct. 2014). Novel temporal features and their application to speech segmentation. [Online]. Available: http://www.scf.usc.edu /~prasantg/manuscript_ALCR_ESTL.pdf
[10] N. Jittiwarangkul, S. Jitapunkul, S. Luksaneeyanawin, V. Ahkuputra, and C. Wuttiwiwatchai, "Thai syllable segmentation for connected speech based on energy," in *Proc. IEEE Asia-Pacific Conference on Circuits and Systems*, Chiangmai, Thailand, 1998, pp. 169-172.
[11] J. Jitsup, U. Sritheeravirojana, and S. Udomhunsakul, "Syllable segmentation of Thai human speech using stationary wavelet transform," in *Proc. Asia-Pacific Conference on Communications*, Bangkok, Thailand, 2007, pp. 29-32.
[12] N. Theera-Umpon, S. Chansareewittaya, and S. Auephan-wiriyakul, "Phoneme and tonal accent recognition for Thai speech," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13254-13259, 2011.

Siripong Potisuk was born in Bangkok, Thailand in 1961. He received Bachelor of Science in Electrical Engineering from The Citadel, The Military College of South Carolina in 1984, Master of Science in Electrical and Computer Engineering and Doctor of Philosophy, both from Purdue University in 1986 and 1995, respectively. He was a lecturer in the Department of Electrical and Computer Engineering, Academic Division of Chulachomklao Royal Military Academy (CRMA) in Thailand from 1996 to 2004. He joined the Department of Electrical and Computer Engineering at the Citadel as an assistant professor in 2005 and was later promoted to associate professor in 2011. His research interests include speech and language processing, control systems, and artificial intelligence.