Thai Text-to-Speech Synthesis: A Review

Chai Wutiwiwatchai[†a)], Chatchawarn Hansakunbuntheung[†], Anocha Rugchatjaroen[†], Sittipong Saychum[†], Sawit Kasuriya[†], Patcharika Chootrakool[†]

National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand

Abstract—**Thai text-to-speech synthesis (TTS) has been researched and developed for decades. Several systems have been launched commercially or publicly while the global TTS technology is still going on with novel algorithms such as deep neural network (DNN). As the Thai language has its special characteristics that make difficulties in computer processing, research work have been focused much on Thai text analysis for TTS pre-processing. Automatic word segmentation, phrase boundary detection, as well as F0 prediction in the Thai tonal language are some of such issues challenging the development of Thai TTS. This article conclusively provides a review of research related to Thai TTS focusing on the last decade (2007-2016). Although there have been consecutive research work on this area, there are still unsolved and challenging problems needed further research. Discussion on the existing issues requiring extensive future work is finally given.**

Index Terms— **Thai text-to-speech synthesis, Thai text processing**

I. Introduction

Text-to-speech synthesis (TTS) has played an important role in today digital communication. While TTS of many world major languages has been made available both publicly and commercially, the technology behind has still been researched and improved. Breakthroughs of synthesis algorithm have been since 1930s from the basic Voder [1] to the unit concatenation method. In the corpus-based technique, units to be concatenated have been flexibly selected from a large corpus, resulting in a clearly more natural synthesized speech. In 2000, the statistical hidden Markov model (HMM) has been introduced with its advantage in highly smoothed sound produced by the parameters re-generated from the HMM [2]. With the success of deep neural network (DNN) for acoustic modeling in automatic speech recognition, the DNN has also been explored to replace the HMM in TTS [3]. Besides the improvement of acoustic modeling, prosody modeling and prediction is also widely explored with a major purpose to make the synthesized speech more *communicative*, i.e. as natural as human conversation. Several basic prosodic parameters such as fundamental frequency (F0) and unit duration, as well as complex parameters such as intonation and stress, have been taken into account in this aspect [4]. Last but not least, the development of TTS for new languages especially those with resource sparsity and the development of multi-lingual TTS have been recently reported [5].

Thai TTS has been researched since the late 90s driven by the need of accessibility options for people with visual disability. Fig.1 illustrates the evolution of Thai TTS research and development. In brief, Thai TTS started from a unit-concatenation based system in around 1999. The use of a large speech corpus for unit-selection based approach was around 2003, and the statistical HMM based approach in 2010. The Thai language has its special characteristics that introduce the difficulties in developing TTS. There is no explicit word nor sentence boundary marker, and segmenting to such units is not trivial. Thai is a tonal language with 4 explicit tone marks but 5 tonal sounds. Tonal coarticulation happens in common while tonal mispronouncing is very sensitive to native listeners. By these examples of complexity, there have been a number of research work trying to solve such issues during the past decade.

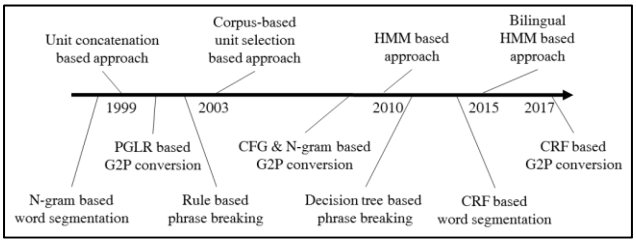

Unit concatenation based approach
Corpus-based unit selection based approach
HMM based approach
Bilingual HMM based approach
1999
2003
2010
2015
2017
PGLR based G2P conversion
CFG & N-gram based G2P conversion
CRF based G2P conversion
N-gram based word segmentation
Rule based phrase breaking
Decision tree based phrase breaking
CRF based word segmentation

Fig.1 Research and development history of Thai TTS

Research and development related to Thai TTS have been reviewed once in 2007 by Wutiwiwatchai and Furui [6]. This article, following the previous review article, will extend the review by grouping especially the past decade research work into 5 categories: 1) language resources; 2) text processing including word segmentation, part-of-speech (POS) tagging, and grapheme-to-phoneme conversion (G2P); 3) prosody prediction including phrase breaking, duration prediction, and F0 prediction; 4) speech syn-

[†]The authors are with National Electronics and Computer Technology Center, National Science and Technology Development Agency, 112 Phahon-Yothin Rd., Klong-Luang, Pathumthani, 12120, Thailand.
a) chai.wutiwiwatchai@nectec.or.th

thesis; and 5) other TTS related issues. The paper organization starts by a brief introduction to the Thai language, an overall diagram of Thai TTS, and an explanation of the development difficulties. Then, research work in the 5 categories will be deeply reviewed. Discussion on unsolved and challenging future topics are given at last.

II. Characteristics of the Thai Language

a. Writing System

Detailed information on the Thai writing system can be found in several books [7, 8]. Thai is a tonal language, like Chinese, and is represented in text form with the Thai alphabet. This native alphabet comprises 44 consonants, 15 basic vowels, and 4 additional tone markers. Text is written from left to right, with no intervening spaces, to form syllables, words, and sentences. Vowels are written above, below, before, or after the consonant they modify, however the consonant is always pronounced first when the syllable is spoken. The vowel characters (and a limited number of consonants) can be combined in various ways to produce numerous compound vowels such as diphthongs.

The grammar of the Thai language is considerably simpler than the grammar of most Western languages, and for many foreigners learning Thai, this compensates for the additional difficulty of learning tones. It is a "Subject + Verb + Object" language with no definite or indefinite article, no verb conjugation, no noun declension, and no object pronouns. Most significantly, words are not modified or conjugated for tense, number, gender, or subject-verb agreement. Articles such as English "a", "an", or "the" are not used. Tenses, levels of politeness, and verb-to-noun conversion are accomplished by the simple addition of various modifying words (called "particles") to the basic subject-verb-object format. One of the major problems for Thai language processing is a lack of word boundaries and explicit sentence markers. White space can be used as sentence, phrase, and word boundaries without strict rules. An analogous example in English is the word "GODISNOWHERE", which can be perceived as "GOD IS NO WHERE", "GOD IS NOWHERE", or "GOD IS NOW HERE" depending on the context.

b. Sound System

A general description of Thai sound system can be found in [9]. Luksaneeyanawin [10] has also published a comprehensive description of the Thai sound system, which is briefly reviewed in this subsection. Thai sound is often described in a syllable unit in the form of /C_i-V-C_f^T/ or /C_i-V^T/, where C_i, V, C_f, and T denote an initial consonant, a vowel, a final consonant, and a tonal level, respectively. The C_i can be either a single or a clustered consonant, whereas the V can be either a single vowel or a diphthong. Table 1 illustrates all Thai consonants, vowels, and tones. As seen in Table 1, some of the phonemes /p, p^h, t, t^h, c, c^h, k, k^h/ can combine with each of the phonemes /r, l, w/ to form a clustered consonant. Diphthongs are double-vowels beginning with one of the vowels /i, ɨ, u, iː, ɨː, uː/ followed by /a/. Five tones in Thai can be divided into 2 groups: the static group consists of 3 tones, the high /á/, the middle /ā/, and the low /à/; the dynamic group consists of 2 tones, the rising /ǎ/ and the falling /â/. Figure 1 shows a graph comparing the F0 contours for the 5 tones that appear in Thai. Recently, some loan-words which do not conform to the rules of native Thai phonology, such as the initial consonants /br, bl, fr, fl, dr/ and the final consonants /f, s, ch, l/ have begun to appear.

Table 1. Thai phonemes in IPA.

Initial consonant, C_i	Single consonant	p, p^h, t, t^h, c, c^h, k, k^h, b, d, m, n, ŋ, f, s, j, r, l, ʔ, h
	Consonant cluster	pr, pl, p^hr, p^hl, tr, t^hr, kr, k^hr, kl, k^hl, kw, k^hw
Vowel, V	Short vowel	i, ɨ, u, e, ɜ, o, æ, a, ɔ, ia, ɨa, ua
	Long vowel	iː, ɨː, uː, eː, ɜː, oː, æː, aː, ɔː, iːa, ɨːa, uːa
Final consonant, C_f	p, t, k, m, n, ŋ, w, j	
Tone, T	ā, à, â, á, ǎ	

III. Overview of Thai TTS and Its Challenge

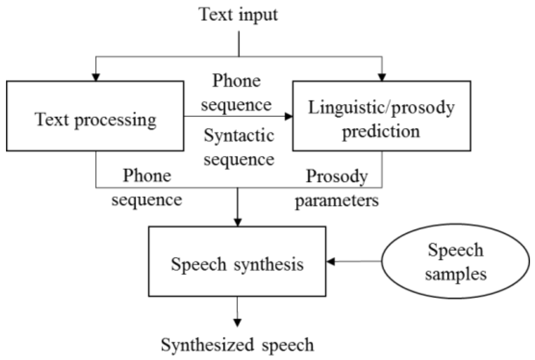


Fig.2 An architecture of Thai TTS

Fig. 2 illustrates a common architecture of Thai TTS [6]. In a general text processing module, text input is tokenized to several types of text chunk such as Thai text, English text, digits, and symbols. Each chunk will be processed with regarding its context in some cases such as the digits. Thai text will furthered segmented into words. Then each word will be tagged by its POS and, at the G2P step, will be converted to its pronunciation in the form of phone sequence. Word, POS, and phone sequences from the text processing module are used by the prosody prediction module for phrase breaking, phone duration prediction, and F0 and intonation prediction. At last, the phone sequence and predicted prosody parameters are the input of the speech synthesis module for constructing a desired sound. It is noted that not all the sub-modules described above are necessary for building Thai TTS, and there are certainly more sub-modules not described which will help enhancing the synthesis performance. Based on this common architecture, several Thai specific problems have been observed but not limited to as follows.

- There is no explicit word boundary as the example given in the previous section, and the often use of compound word makes the word unit hard to clearly define. Combining with a lot of loan words from Pali, Sanskrit, and transliterated words from foreign languages, Thai word segmentation becomes highly ambiguous.
- There is even no explicit sentence boundary. Complicated writing comes from the use of conjunctions while subject is often omitted and serial verbs are often used. Defining a sentence becomes extremely difficult. Processing text on a basis of sentence could hence be error prone.
- Although Thai writing is a spelling system, the use of 44 consonantal alphabets, 15 vowel forms, and 4 tonal markers, especially for loan and transliterated foreign words, makes Thai G2P not a straightforward job. There are also words with pronunciations exceptionally not complied with common rules. Correct word segmentation is needed for accurate G2P.
- As similar to other languages, phrase and sentence intonation is of crucial for natural sound. The absolute phrase or sentence intonation in Thai is even more complicated by the combination of F0 driven by the phrase or sentence itself, and those restricted by syllabic tones. Furthermore, these syllabic tones are highly co-articulated regarding their syllable and word context. F0 modeling and prediction becomes one of important tasks for Thai.

IV. Language Resources

In today data-driven approaches, language resources are necessary as a basis for analysis and modeling. Table 2 summarizes the Thai language resources useful for TTS related work. In addition to the fundamental resources used to develop Thai TTS since a decade ago like the ORCHID text corpus [11] and the TSynC-1 speech corpus, the current advance has come from several modern Thai resources including a large word-boundary and POS tagged corpus called BEST [12], a large pronunciation lexicon called LEXiTRON-Pro [13], and a re-designed speech corpus for speech synthesis evaluation in particular [14]. The LEXiTRON-Pro has its special tag called *pseudo-morpheme* (PM), in which its detail will be given in the next section. In addition to speech corpora and lexicon, in 2015, Hansakunbuntheung et al. [15] has proposed a framework for creating a text corpus specifically for Thai G2P evaluation.

Table 2. Thai TTS language resources.

Resource	Detail
ORCHID	Word boundary and POS tagged text corpus of 568,316 words
TSynC-1	Prosody tagged speech corpus of 13-hour, 1-female speaker
TSynC-2	10.5-hour speech corpus reduced from TSynC-1
BEST	Word boundary tagged text corpus of 5 million words
LEXiTRON-Pro	Pronunciation and syllable tagged dictionary of 140,000 words

V. Text Processing

Published research and development regarding Thai text processing useful for TTS can be grouped to two general topics: word segmentation and POS tagging, and G2P.

a. *Text Segmentation and POS Tagging*

Word segmentation is a classical task in Thai text processing. Wutiwiwatchai and Furui [6] has summarized the work in this topic until around 2005. During 2009 to 2012, Thai word segmentation has been extensively improved under the BEST [12] series contest using the large annotated text corpus. In these contests, various statistical approaches have been conducted and trained by the 5 million words BEST corpus. A state-of-the-art performance was over 97% word segmentation accuracy by a hybrid word and character-cluster model and discriminative online learning, proposed by Kruengkrai et al. [16]. Conditional Random Fields (CRF) have also given high accuracies in this task [17].

Thai word segmentation could be partly disambiguated by using word POS. Therefore, some have investigated in word segmentation and POS tagging in a single learning machine. SWATH [18] is one in the past. Recently, Boonkwan et al. [19] has investigated on a unified model for word segmentation and POS tagging using Deep Neural Network (DNN).

Although there have been some work on sentence breaking in order to discover Thai sentences [20, 21]. As the sentence unit is hardly defined in the current complex Thai writing, there have been only small size data for training and evaluation. Errors of sentence breaking mostly come from the unclear definition of sentences itself. Hence there have been, if not none, only few extensions of sentence breaking in Thai language processing applications. The work by Tangsirirat et al. [22] was among ones that tried to tackle this problem especially for social media text. They introduced a combined rule-based and statistical approach and obtained a 93% sentence breaking correctness.

There have also been work on syllabification and syllable-like unit segmentation. Although Thai spelling is a /C_i-V-C_f^T/ system, vowel characters could be located around the initial consonant characters and even across syllables. Some consecutive textual syllables are then not possible to segment. Aroonmanakun [23] has proposed a nice syllable-like unit which combines these un-segmented syllables as a single unit. This unit has been called Pseudo-morpheme (PM) by Jongtaveesataporn et al. [24] and used for Thai speech recognition lexical unit modeling. Syllabification based on this unit achieved over 99% segmentation accuracy by using syllable-pattern parsing and a syllable 3-gram model on about 19,000 word test set in Aroonmanakun [23]. In 2015, Kongyoung and Rugchatjaroen [25] has produced about 98% PM segmentation accuracy by using Conditional Random Fields (CRF) on about 21,000 word test set without any syllable parsing needed. In 2011, there has been another attempt for Thai syllabification by Jucksriporn and Sornil [26]. They extended the idea of Thai character cluster (TCC) [27] to a Thai minimum cluster (TMC) and used it as a basis for segmentation. With this unit, they achieved around 97% segmentation accuracy.

Many research work have utilized word part-of-speech (POS) for disambiguating word segmentation. These systems hence produce word sequences with word POS simultaneously. SWATH [18], since 1997, has employed POS n-gram modeling and achieved 92% word segmentation accuracy. Besides POS analysis, Thai has been considered better structuring by Categorial Grammar (CG). Supnithi et al. [28] conducted CG analysis for Thai and the CG model has been integrated in Thai phrase break analysis for TTS [29].

b. *Grapheme-to-phoneme Conversion*

According to Thai G2P difficulties described in Section III, there have been consecutive research work on this topic. In 2006, Thangthai et al. [30] investigated on a syllable n-gram model for scoring possible syllable segmentation paths and selecting the best path as a result. Syllabification was performed in prior by using Context Free Grammar (CFG). With their model, new syllable patterns could be automatically induced from large raw text by using the CFG rules and new patterns with high probabilities were stored. Although they have achieved a considerably high G2P performance, there have still been unsolved issues needed further improvement. Issues such as hidden syllables described in Section III, tone and vowel length distortion, have been deeply analyzed and alleviated in Rugchatjaroen et al. [31]. In this work, two-stage CRFs were used for syllabification and syllable type prediction, and for phoneme prediction. The proposed model reduced the G2P errors from a 24.8% word error rate by their baseline system [30] to a 13.7% word error rate. The problem such as hidden syllable caused mainly by Pali-Sanskrit loan words has been drastically solved, but the problem such as tone distortion, i.e. tone sounds not complied with their tone marks, still open for research.

Modern Thai writing often mixes with English words, which make G2P more complex in terms of both text analysis and speech synthesis. Aroonmanakun [32] has systematically stated the problem and proposed a rule-based approach derived from a training corpus. With ignoring tone, 36.4% of test words were evaluated as good conversion, 18.9% for fair, and the rest for poor. In 2007, Thangthai et al. investigated on a Classification and Regression Tree (CART) based model for English words to Thai phonemes conversion. A novelty came from the use of English phonemes obtained from English pronunciation dictionary to help G2P conversion. However, only 53.3% syllable accuracy was obtained. In 2013, Pitakpawatkul et al. [33] proposed using an online discriminative training for phone prediction in combination of CART-based tone prediction. They achieved up to a 76% syllable accuracy. It is clear that transcribing English words by using Thai phonemes is still unsolved as correct pronunciations are not clearly regulated. An officially pronunciation defined, large dictionary is needed for extensive improvement of this necessary task.

VI. Prosody Prediction

a. *F0 Prediction*

In the unit-concatenation based speech synthesis, predicted F0 has been used as one parameter for unit selection, whereas the predicted F0 has been directly used to constrain the speech generation in the HMM-based approach. As Thai is a tonal language, F0 modeling and prediction has particularly been interested by many researchers. Thangthai et al. [34, 35] tried to model F0 of Thai syllables by using the Tilt model. They found that the Tilt model could represent F0 shapes of Thai tonal syllables with some modifications. Two years later, Prom-on [36] proposed a pitch target representation for Thai tones. In 2012, Chunwijitra et al. [37] has proposed F0 modeling in Thai tones using quantized F0 and phone or sub-phone units. Inconsistency of F0 labeling over tonal units could then be overcome and then modeling became more precise.

In addition to the main research stream, F0 modeling in Thai has also been explored for different aspects. Chompan [38] has worked on F0 modeling for Thai dialects using the classical Fujisaki model. Boonpiam et al. [39] conducted a cross-lingual F0 modeling. With a relatively closed languages in term of syllabic tones, they proposed to use a Mandarin Chinese corpus as a resource for Thai F0 modeling constrained on syllable units.

b. *Duration Prediction*

Hansakunbuntheung et al. [40] are among research groups conducting continuous work on duration modeling for Thai TTS. The proposed algorithm was based on linear regression of syllable unit duration with respect to several factors such as phone context. As proposed partly in Rugchatjaroen et al. [41], a similar linear regression based duration modeling has been integrated in a complete Thai TTS system. In 2008, Saychum et al. [42] proposed a method to improve the duration prediction for unit-concatenation based TTS. In this work, only important phones defined heuristically were given more weights than those for the other phones in unit selection criteria. This could make synthesized speech more intelligible.

c. *Phrase Breaking*

Since Thai has no explicit sentence boundary marker, phrase breaking has become an important module for natural speech synthesis as well as for fast response text-to-speech production. Similar to word segmentation, phrase breaking has been investigated for Thai for long. The TTS system by Rugchatjaroen et al. [41] has included their phrase breaking module based on CART. Their work had actually followed the same algorithm proposed by Hansakunbuntheung et al. since 2005 [43]. As mentioned in

Section V(a), Categorial Grammar (CG) has been introduced to be one of suitable Thai grammar models. Therefore in 2011, Saychum et al. [29] investigated on phrase breaking using the CG grammar tags in place of the simple POS. They found that using automatic CG tagging could produce a better phrase breaking performance than that using automatic POS tagging. This might due to the fact that POS tags used in their task were not yet well defined and hence POS tagging has become ambiguous.

VII. Speech Synthesis

There have been not many publications tackling the speech synthesis part for Thai in particular. It is known that one of the problems found in unit-concatenation based TTS is the smoothness of connected points. Nuratch et al. [44] has faced this problem and tried to solve by detecting synthesized signal discontinuity. The problematic signal portion is removed and replaced by new PARCOR interpolated signals. Some other publications aim to report the overall development of Thai TTS with improvement at some minor issues. Chula TTS [45] has introduced a framework for integrating several TTS submodules to work together with easily configurable.

Another interesting issue of TTS of modern Thai writing which often contains English words. Thai-accent English is required to make the whole sound natural. Rugchatjaroen et al. [46] tackled this problem by constructing two separated monolingual HMMs for English and Thai using two speech corpora. Then the English HMM was adapted by the Thai HMM to make English phones more Thai accent, and the two different voice talents became closer to each other. Wutiwiwatchai et al. [47] suggested that Thai-accent English could be varied by the fluent of English of the speaker. This work thus conducted an interpolation of Thai and English HMM-based acoustic models. The interpolation weight was then a variable controlling the degree of English accent. This makes the TTS system available for users to adjust the accent level of English words as they like.

VIII. Run-time Processing and Applications

In addition to the main research topics for building Thai TTS, there have been work related to TTS run-time processing and applications. Since 2008, there have been attempts to port Thai TTS onto mobile devices which are resource limited but real-time required. Chinathimatmongkhon [48] developed Thai TTS to run on Palm OS mobile devices. With this low-resource device, rule-based text processing and HMM-based synthesizer were minimized and deployed. Wongpatikaseree et al. [49, 50] proposed an implementation scheme for unit concatenation based system on Windows phone devices. Their contributions were the optimization of text processing and the design of speech corpus suitable for different kind of unit in concatenation. Saychum et al. [51] has first integrated bilingual Thai-English TTS into a single system successfully running on Android devices. In a year later, Saychum et al. [52] introduced a novel method to make TTS on mobile devices respond faster. They proposed a fast-track way that skipped high computation modules such as POS tagging. Nearly first text chunks were fed to this fast track and the following chunks were processed as usual. Synthesized sounds from the fast track and the full-operation track were connected smoothly within multithread processing, so that the whole speech could be produced seamlessly.

TTS is an important component in the area of information accessibility especially for people with visual disability. In this aspect, Suchato et al. [53] has investigated on integrating Thai TTS into a voice-interacted web browser. Another interesting application was reported by Khorinphan et al. [54]. They implemented a formant synthesis based system for home robots and tried to synthesize appropriate emotional tones. This seems to be nearly the first speech synthesis work beginning to generate Thai emotional speech.

IX. Future Challenge

In the global aspect, the TTS technology is still under researched and improved. Based on parametric speech modeling, synthesized speech still suffers from sound buzzy caused by vocoding. Researchers in the area of statistical HMM as well as DNN acoustic modeling are on finding better algorithms to generate more intelligible sounds [??]. For Thai TTS in particular, there are still not many available speech and text corpora. Existing resources contain some compulsory tags such as word boundaries in the text corpus, phone transcriptions and automatically generated prosodic features in the speech corpus. In order to enhance the performance of TTS, deeper and broader tags are required. These include, for example, POS, phrase and sentence boundaries in the text corpus, and tone realization and stress markers in the speech corpus. Moreover, Thai is lack of standard, well-designed speech and text corpora specifically for TTS evaluation.

Modern Thai writing is more complicated for computation. Social media text have special characteristics requiring complex text normalization. Moknarong et al. [55] was one that tackled the problem of romanized Thai words often occurred in social media text. With a combination of decision tree and n-gram methods, over 89% of romanized Thai words were detected. Further research is required after the detection of romanized Thai words in order to properly synthesize speech in this bilingual environment.

Last but not least, today TTS has been requested to produce much more communicative sounds i.e. sounds containing speakers' emotion, expressive, and non-verbal information. This issue has raised much more difficulty in text analysis as well as prosody prediction. For the text analysis, long context may have to be taken into account when we would like to identify the speaker intention, whereas for the prosody prediction, it is hard to produce suitable prosodic parameters from pure text. Research work related to emotional or communicative speech analysis such as Chompan [56] is the first step toward such predictive engines.

X. Conclusion

Text-to-speech synthesis (TTS) has played an important role in today digital age, where multimedia over the Internet and smart phone becomes compulsory. Even with the long history of TTS research from all around the world as well as in Thailand, the technology itself has still opened for new algorithms to improve the sound quality and naturalness. Thai TTS has been researched, developed and improved along the invention of new algorithms; from unit concatenation to corpus-based unit selection, and to statistical HMM. This paper has intensively reviewed major research contributions to Thai TTS since around 2007. The main contribution went to the text analysis part which is much more language specific. During the last decade, Thai word segmentation as well as G2P has been obviously improved, where there have been a number of work investigating on tone and intonation modeling. Toward communicative TTS where speech emotion and expressiveness have to be involved in the synthesized sound, Thai TTS still need some fundamental text analysis tool such as phrase and sentence breaking, and stress modeling. For Thai, these issues would be focused in the near future.

References

1. H. Dudley, "System for the artificial production of vocal or other sounds", US application 2121142, Bell Telephone Laboratories , June 1938.
2. K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura, "Speech parameter generation algorithms for HMM-based speech synthesis", Proc. of ICASSP, pp.1315-1318, 2000.
3. H. Zen, A. Senior, M. Schuster, "Statistical parametric speech synthesis using deep neural networks", Proc. of ICASSP, pp. 7962–7966, 2013.
4. K. Sreenivasa Rao, "Predicting prosody from text for text-to-speech synthesis", Springer-Verlag New York, 2012.
5. C. Traber, K. Huber, K. Nedir, B. Pfister, E. Keller, B. Zellner, "From multilingual to polyglot speech synthesis", Proc. of Eurospeech, pp. 835-838, 1999.
6. C. Wutiwiwatchai, S. Furui, "Thai speech processing technology: a review", Speech Communication 49 (1), 8-27, 2007.
7. M. R. Haas, "The Thai system of writing", Program in Oriental Languages. Spoken Language Services, Inc., New York, 1980.
8. J. Higbie, S. Thinsan, "Thai reference grammar: the structure of spoken Thai", Orchid Press, Bangkok, 2002.
9. K. Tingsabadh, A. S. Abramson, "Illustrations of the IPA: Thai", Handbook of the International Phonetic Association. Cambridge University Press, Cambridge, 1999.
10. S. Luksaneeyanawin, "Three-dimensional phonology: a historical implication", Proc. of International Symposium on Language and Linguistics, pp. 75–90, 1992.
11. T. Charoenporn, V. Sornlertlamvanich, H. Isahara, H., "Building a large Thai text corpus – part-of-speech tagged corpus: ORCHID", Proc. Of NLPRS, 1997.
12. K. Kosawat, M. Boriboon, P. Chootrakool, A. Chotimongkol, S. Klaithin, S. Kongyoung, K. Kriengket, S. Phaholphinyo, S. Purodakananda, T. Thanakulwarapas, C. Wutiwiwatchai, "BEST 2009: Thai word segmentation software contest", Proc. of SNLP, pp. 83–88, 2009.
13. P. Chootrakool, C. Wutiwiwatchai and K. Kosawat, "A large pronunciation dictionary for Thai speech processing", Proc. of ASIALEX, Bangkok, 2009.
14. C. Wutiwiwatchai, S. Saychum, A. Rugchatjaroen, "An intensive design of a Thai speech synthesis corpus", Proc. of SNLP, 2007.
15. C. Hansakunbuntheung, S. Thatphithakkul, "Context-dependent grapheme-to-phoneme evaluation corpus using flexible contexts and categorial matrix", Proc. of Oriental COCOSDA, 2015.
16. C. Kruengkrai, K. Uchimoto, J. Kazama, K. Torisawa, H. Isahara, and C. Jaruskulchai, "A word and character-cluster hybrid model for Thai word segmentation", Proc. of InterBEST 2009.
17. C. Haruechaiyasak and S. Kongyoung, "TLex: Thai lexeme analyser based on the Conditional Random Fields", Proc. of InterBEST 2009.
18. S. Meknavin, P. Charoenpornsawat, and B. Kijsirikul, "Feature-based Thai word segmentation", Proc. of NLPRS, Phuket, Thailand, 1997.
19. P. Boonkwan, V. Sutantayawalee, and T. Supnithi, "Language as Tensors: bidirectional deep learning of context representations for joint word segmentation and part-of-speech tagging", Proc. of DESGT, 2017.
20. G. Slayden, M. Y. Hwang, and L. Schwartz, "Thai sentence-breaking for large-scale SMT", Proc. of WSSANLP, pp. 8-16, 2010.
21. P. Charoenpornsawat, V. Sornlertlamvanich, "Automatic sentence break disambiguation for Thai", Proc. of ICCPOL, 2001.
22. N. Tangsirirat, A. Suchato, P. Punyabukkana, C. Wutiwiwatchai, "Contextual behavior features and grammar rules for Thai sentence-breaking", Proc. of ECTI-CON, 2013.
23. W. Aroonmanakun, "A unified model of Thai word segmentation and Romanization", Proc. of PACLIC, pp. 205–214, 2005.
24. M. Jongtaveesataporn, I. Thienlikit, C. Wutiwiwatchai, S.i Furui, "Lexical units for Thai LVCSR", Speech Communication 51 (4), 379-389, 2009.
25. S. Kongyoung, A. Rugchatjaroen, "Thai Pseudo Syllable Segmentation using Conditional Random Fields", Proc. of Oriental COCOSDA, Shanghai, China, 2015.
26. C. Jucksriporn, O. Sornil, "A minimum cluster-based trigram statistical model for Thai syllabification", CICLing, pp. 493-505, 2011.
27. T. Theeramunkong, V. Sornlertlamvanich, "Character cluster based Thai information retrieval", Proc. of IRAL, Hong Kong, pp. 75–80, 2000.
28. T. Supnithi, T. Ruangrajitpakorn, K. Trakultaweekool, and P. Porkaew, "AutoTagTCG: A framework for automatic Thai CG tagging", Proc. of LREC, pp. 971-974, 2010.
29. S. Saychum, C. Hansakunbuntheung, N. Thatphithakkul, T. Ruangrajitpakorn, C. Wutiwiwatchai, T. Supnithi, A. Chotimongkol, A. Thangthai, "Categorial-grammar-based phrase break prediction", Proc. of ECTI-CON, 2011.
30. A. Thangthai, C. Hansakunbuntheung, R. Siricharoenchai, C. Wutiwiwatchai, "Automatic syllable-pattern induction in statistical Thai text-to-phone transcription", Proc. of INTERSPEECH 2006.
31. S. Saychum, S. Kongyoung, A. Rugchatjaroen, P. Chootrakool, S. Kasuriya, C. Wutiwiwatchai, "Efficient Thai grapheme-to-phoneme conversion using CRF-based joint sequence modeling", Proc. of INTERSPEECH, 2016.
32. W. Aroonmanakun, N. Thapthong, P. Wattuya, B. Kasisopa, S. Luksaneeyanawin, "Generating Thai transcriptions for English words. In SEALS XIV, Vol. 1, Papers from the 14th annual meeting of the Southeast Asian Linguistics Society

2004, Wilaiwan Khanittanan and Paul Sidwell (eds). May 19-21, 2004, Bangkok, 13-22.

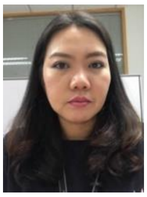
33. K. Pitakpawatkul, A. Suchato, P. Punyabukkana, C. Wutiwiwatchai, "Thai phonetization of English words using English syllables", Proc. of ECTI-CON, 2013.
34. A. Thangthai, N. Thatphithakkul, C. Wutiwiwatchai, A. Rugchatjaroen, and S. Saychum, "T-Tilt: a modified Tilt model for F0 analysis and synthesis in tonal languages", Proc. of INTERSPEECH, pp. 2270-2273, 2008.
35. A. Thangthai, A. Rugchatjaroen, N. Thatphithakkul, A. Chotimongkol, C. Wutiwiwatchai, "Optimization of T-Tilt F0 modeling", Proc. of INTERSPEECH, 2009.
36. S. Prom-on, Y. Xu, "Pitch target representation of Thai tones", Proc. of TAL, 2012.
37. V. Chunwijitra, T. Nose, T. Kobayashi, "A tone-modeling technique using a quantized F0 context to improve tone correctness in average-voice-based speech synthesis," Speech Communication, vol.54(2), pp.245-255, 2012.
38. S. Chomphan, "Fujisaki's model of fundamental frequency contours for Thai dialects", Journal of Computer Science 6 (11): 1246-1254, 2010.
39. V. Boonpiam, A. Rugchatjaroen, C. Wutiwiwatchai, "Cross-language F0 modeling for under-resourced tonal languages: a case study on Thai-Mandarin", Proc. of INTERSPEECH, 2009.
40. C. Hansakunbuntheung, H. Kato, Y. Sagisaka, "Syllable-based Thai duration model using multi-level linear regression and syllable accommodation", Proc. of ISCA Workshop on Speech Synthesis (SSW6), pp. 356-361, 2007.
41. A. Rugchatjaroen, A. Thangthai, S. Saychum, N. Thatphithakkul, C. Wutiwiwatchai, "Prosody-based naturalness improvement in Thai unit-selection speech synthesis", Proc. of ECTI-CON, Thailand, 2007.
42. S. Saychum, A. Rugchatjaroen, N. Thatphithakkul, C. Wutiwiwatchai A. Thangthai, "Automatic duration weighting in Thai unit-selection speech synthesis", Proc. of ECTI-CON, Krabi, pp. 549-552, 2008.
43. C. Hansakunbutheung, A. Thangthai, C. Wutiwiwatchai and R. Siricharoenchai. "Learning Methods and Features for Corpus-Based Phrase Break Prediction on Thai", Proc. of INTERSPEECH, pp. 325-328, 2005.
44. S. Nuratch, P. Boonpramuk, C. Wutiwiwatchai, "Shape and frequency continuity improvement in concatenation-based Thai text-to-speech synthesis", Proc. of SNLP, 2007.
45. N. Kertkeidkachorn, S. Chanjaradwichai, P. Punyabukkana, A. Suchato, "CHULA TTS: A modularized text-to-speech framework", Proc. of PACLIC, pp. 414–421, 2014.
46. A. Rugchatjaroen, N. Thatphithakkul, A. Chotimongkol, A. Thangthai, C. Wutiwiwatchai, "Speaker adaptation using a parallel phone set pronunciation dictionary for Thai-English bilingual TTS", Proc. of INTERSPEECH, 2009.
47. C. Wutiwiwatchai, A. Thangthai, A. Chotimongkol, C. Hansakunbuntheung, N. Thatphithakkul, "Accent level adjustment in bilingual Thai-English text-to-speech synthesis", Proc. of ASRU, 2011.
48. N. Chinathimatmongkhon, A. Suchato, P. Punyabukkana, "Implementing Thai text-to-speech synthesis for hand-held devices", Proc. of ECTI-CON, Krabi, Thailand, 2008.
49. K. Wongpatikaseree, A. Ratikan, A. Thangthai, A. Chotimongkol, C. Nattee, "A real-time Thai speech synthesizer on a mobile device", Proc. of SNLP, 2009.
50. K. Wongpatikaseree, A. Ratikan, A. Chotimongkol, P. Chootrakool, C. Nattee, T. Theeramunkong, T. Kobayashi, "A hybrid diphone speech unit and a speech corpus construction technique for a Thai text-to-speech system on mobile devices", Proc. of ECTI-CON, 2010.
51. S. Saychum, A. Thangthai, P. Janjoi, N. Thatphithakkul, C. Wutiwiwatchai, P. Lamsrichan, T. Kobayashi, "A bi-lingual Thai-English TTS system on Android mobile devices", Proc. of ECTI-CON, 2012.
52. S. Saychum, N. Thatphithakkul, C. Wutiwiwatchai, P. Lamsrichan, T. Kobayashi, "Fast-track text processing for real-time text-to-speech on mobile devices", Proc. of ECTI-CON, 2013.
53. A. Suchato, J. Chirathivat, P. Punyabukkana, "Enhancing a voice-enabled web browser for the visually impaired", Proc. of ICAS, Vientiane , Laos , 2006.
54. C. Khorinphan, S. Phansamdaeng, S. Saiyod, "Thai speech synthesis with emotional tone based on formant synthesis for home robot", Proc. of ICT-ISPC, 2014.
55. N. Moknarong, A. Suchato, P. Punyabukkana, "Detecting romanized Thai tokens in social media texts", Proc. of ICSEC, pp. 36-41, 2013.
56. S. Chomphan, "Modeling of fundamental frequency contour of Thai expressive speech using Fujisaki's model and structural model", Journal of Computer Science 7 (8) (2011) 1310-1317.

Chai Wutiwiwatchai received Ph.D. in Computer Science from Tokyo Institute of Technology in 2004. He is now the Director of Intelligent Informatics Research Unit, National Electronics and Computer Technology Center (NECTEC), Thailand. His research interests include speech processing, natural language processing, and human-machine interaction. His research work includes several international collaborative projects in a wide area of speech and language processing as well as nation-wide e-Learning. He is now a member of the International Speech Communication Association (ISCA) and the Institute of Electronics, Information and Communication Engineers (IEICE).

Chatchawarn Hansakunbuntheung received B.Eng. and M.Eng. degree in Electrical Engineering from Chulalongkorn University, Thailand, in 1998 and 2000, respectively, and, D.Sc. in Global Information and Telecommunication Studies from Waseda University, Tokyo, in 2010. He is a researcher at National Electronics and Computer Technology Center (NECTEC), Thailand. His research interests include speech and language processing, multilingual processing, language resources, and, human-computer interaction. His research works includes Asian text-to-speech synthesis, language proficiency evaluation, and Asian speech and language resources. He is a member of the International Speech Communication Association (ISCA).

Anocha Rugchatjaroen received her PhD in Electronics Engineering from the University of York in 2014. She is now a researcher at the Speech and audio technology laboratory (SPT), National Electronics and Computer Technology Center (NECTEC), Thailand. Her research interest is in all aspects of speech synthesis especially in corpus-based, statistical-based and articulatory-based speech synthesis.

Sittipong Saychum is an assistant researcher in the field of Speech and Audio Technology Laboratory, Human Computer Communication Research Unit, National Electronics and Computer Technology Center (NECTEC), Thailand.



Patcharika Chootrakool received MS.C. in Information Technology from King Mongkut's University of Technology Thonburi in 2004 and BA. (Linguistics) from Thammasart University in 1997. She is now a researcher in Speech Technology Team at Human Computer Communications Research Unit, HCCRU, National Electronics and Computer Technology Center (NECTEC). Her expertise is Thai Linguistics.