Thai speech processing technology: A review

Chai Wutiwiwatchai [a,*], Sadaoki Furui [b]

[a] *Speech Technology Research Section, Information R&D Division, National Electronics and Computer Technology Center, 112 Pahonyothin Rd.,*
Klong-luang, Pathumthani 12120, Thailand
[b] *Department of Computer Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan*

Abstract

This paper presents a comprehensive review of Thai speech technology, from its impetus in the early 1960s to 2005. Thai is the official language of Thailand, and is spoken by over 60 million people worldwide. As with Chinese, it is a tonal language. It has a spelling system using a Thai alphabet, but has no explicit word boundaries, similar to several Asian languages, such as Japanese and Chinese. It does have explicit marks for tones, as in the languages of the neighboring countries, Laos and Vietnam. Therefore, with these unique characteristics, research and development of language and speech processing specifically for Thai is necessary and quite challenging. This paper reviews the progress of Thai speech technology in five areas of research: fundamental analyses and tools, text-to-speech synthesis (TTS), automatic speech recognition (ASR), speech applications, and language resources. At the end of the paper, the progress and focus of Thai speech research, as measured by the number of publications in each research area, is reviewed and possible directions for future research are suggested.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Speech analysis; Speech processing; Speech synthesis; Speech recognition; Thai speech

1. Introduction

The history of Thai speech research spans over 40 years. Similar to other languages, early research was limited by computational capability. Initial research mainly involved linguistic studies, where only small sample sets were analyzed by humans regarding some specific characteristics of Thai, such as lexical tones and stress. Representative studies from this area, made during the 1960s–1970s, include that by Abramson (1962), Gandour (1979) and Luangthongkum (1977).

In the late 1980s, when computers became much more powerful and affordable, computational linguistics became a major inter-disciplinary field. Research moved from the study of the basic characteristics of Thai speech to natural language and speech processing, including the development of Thai speech synthesis systems (Luksaneeyanawin, 1989; Saravari and Imai, 1983). Such complicated systems required the parallel development of several fundamental tools for morphological and phonological analysis, which take into account the unique characteristics of the Thai language.

Research on Thai speech recognition first appeared in 1987 in a study by Pathumthan (1987). In this study, he developed a system capable of recognizing a very limited vocabulary of isolated Thai words. During the same period, several similar works on isolated digits or words were explored (Kiat-arpakul, 1996; Pensiri and Jitapunkul, 1995). Another topic widely researched was the analysis and modeling of tones (Abramson, 1962; Anivan, 1988; Gandour, 1979) and other prosody features, such as intonation and stress in Thai (Abramson and Svastikula, 1982; Potisuk et al., 1996a).

The late 1990s to the early 2000s saw a huge expansion of Thai speech research, with major developments in Thai speech recognition and synthesis. Nevertheless, no large speech corpus was available and hence most of the speech

---

* Corresponding author. Tel.: +662 564 6900x2230; fax: +662 564 6873.
  E-mail address: chai@nectec.or.th (C. Wutiwiwatchai).

recognition research emphasized only isolated-word recognition. In 2000, the National Electronics and Computer Technology Center (NECTEC) in Thailand initiated a project to develop the first Thai large vocabulary continuous speech recognition (LVCSR) corpus (Kasuriya et al., 2003b). This corpus was not only used for Thai continuous speech recognition research at NECTEC (Kanokphara, 2003a; Tarsaku and Kanokphara, 2002), but also for many other research at different sites. In the field of speech synthesis, systems reached a commercial level of sophistication by the end of 1990s. Most of the systems currently available on the market are based on concatenation of subword units taken from an inventory of subword units. It has also been reported that the naturalness of synthesized voices can be improved by using variable-length units selected from a large corpus of speech utterances (Sagisaka et al., 1992). NECTEC has also succeeded in improving sound quality for their TTS system using the same technique (Hansakunbuntheung et al., 2005a).

At present, research on Thai speech recognition has moved on to various applications of continuous speech recognition (Pisarn et al., 2005; Suebvisai et al., 2005; Thienlikit et al., 2004). Research areas for speech synthesis include improving intelligibility of text processors and enhancing naturalness of synthetic speech by improving prosody prediction. In addition to the basic research already described, advanced applications of speech, such as a spoken dialogue system (Suchato et al., 2005; Wutiwiwatchai and Furui, 2003b) and a speech-to-speech translation system (Schultz et al., 2004) have also been attempted. Since basic tools and resources for recognizing Thai are still limited, developing these applications remains a formidable task.

Notable reviews and surveys of progress in Thai speech research published in the past include Luksaneeyanawin (1993), Jitapunkul et al. (1998), and Thongprasert et al. (2002). All these papers mainly presented an overview of the work carried out by their own organizations. In this paper, we attempt to review all published literature on speech research in Thai, not only that conducted by particular organizations. Section 2 introduces both the writing and sound systems of the Thai language. Section 3 describes research works on fundamental analyses and tools. Sections 4 and 5 review investigations of Thai speech synthesis and recognition, respectively. Section 6 presents research on high-level speech applications. Suggestions for, and discussion on, potential research directions are given at the end of the paper.

2. Thai writing/sound systems

Thai is the official language of Thailand, and is spoken by over 60 million people worldwide. Thai is but one of many languages and dialects belonging to the Tai sub-family (itself a member of the Sino-Tibetan family of languages). The following subsections review the text and sound characteristics of Thai.

2.1. Thai writing system

Detailed information on the Thai writing system can be found in several books (Haas, 1980; Higbie and Thinsan, 2002). In this subsection, we review some important characteristics of written Thai in order to lay the foundation for explanations of research presented in the rest of this paper. 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 horizontally, 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.

2.2. Thai sound system

A general description of Thai sound system can be found in (Tingsabadh and Abramson, 1999). Luksaneeyanawin (1992, 1993) 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 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, iː, ɨ, ɨː, u, uː/ followed by /a/. Five tones in Thai can be divided into two groups: the