ว.วิทย. มข. 40(4) 1114-1127 (2555) KKU Sci. J. 40(4) 1114-1127 (2012)

ทรัพยากรข้อความภาษาไทย: ชุดข้อความภาษาไทยที่ควรใช้สำหรับการวัดคุณภาพเสียงและการศึกษาเปรียบเทียบ

Thai Text Resource: A Recommended Thai Text Set for Voice Quality Measurements and Its Comparative Study

Therdpong Daengsi[1*], *Apiruck Preechayasomboon*[2], *Saowanit Sukparungsee*[3] *and Chai Wutiwiwatchai*[4]

บทคัดย่อ

บทความปริทัศน์นี้ เป็นการนำเสนอโดยสังเขปเกี่ยวกับ ภาพรวมของการวัดคุณภาพเสียง ชุดข้อความภาษาอังกฤษที่เกี่ยวข้องกับการวัดดังกล่าว ภาพรวมของภาษาไทยและเสียงภาษาไทย และการพัฒนาชุดข้อความภาษาไทยที่มีชื่อว่า ทีทีเอสที ซึ่งเป็นชื่อย่อของชุดข้อความภาษาไทยสำหรับการวัดคุณภาพเสียงในระบบสื่อสารโทรคมนาคม (Thai text set for telephonometry) ทีทีเอสทีถูกนำไปเปรียบเทียบกับโลตัส (LOTUS) ซึ่งเป็นฐานข้อมูลภาษาไทยที่ถูกนำไปใช้สำหรับในงานวิจัยด้านการประมวลผลภาษา และเสียงพูดอย่างกว้างขวางในประเทศไทย บทความนี้ได้นำเสนอว่า ทีทีเอสที เป็นชุดข้อความที่ง่าย และคล้ายกับประโยคซีไอดี (CID Sentences) และบางส่วนของประโยคคูนี่ (CUNY sentences) นอกจากนั้นยังเป็นไปตามมาตราฐาน ITU-T P.800 ดังนั้น ทีทีเอสทีจึงเหมาะสำหรับใช้ในการสร้างกลุ่มตัวอย่างเสียงสำหรับการวัดคุณภาพเสียงในสภาพแวดล้อมของไทย ซึ่งใช้ภาษาไทยในการติดต่อสื่อสาร

[1]Department of Information Technology, Faculty of Information technology, King Mongkut's University of Technology North Bangkok, Bangkok, 10800

[2]Metropolitan Sale and Services, TOT Public Company Limited, Bangkok, 10150

[3]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800

[4]Human Language Technology Laboratory, National Electronics and Computer Technology Center Pathumthani, 12120

*Corresponding Author, E-mail: therdpong1@yahoo.com

ABSTRACT

This review paper briefly presents voice quality measurement overview, related English text sets used in measurement, an overview of the Thai language and Thai sounds, and the development of a Thai text set, called the Thai Text Set for Telephonometry (TTST). TTST has been compared with LOTUS which is a Thai corpus that has been used widely for language and speech processing research in Thailand. This paper presents that TTST is simple and similar to CID Sentences and some CUNY Sentences. Moreover, it is in compliance with ITU-T P.800 standard. Therefore, the TTST could be recommended to create speech samples for voice quality measurements in Thai environments, which use Thai for communication.

คำสำคัญ: ชุดข้อความภาษาไทย ฐานข้อมูลโลตัส การวัดคุณภาพเสียง
Keywords: Thai text set, LOTUS corpus, Voice quality measurement

Introduction

Voice quality assessment requires an appropriate set of speech, following standards that issued by ITU-T (1996a; 1996b). Therefore, it is important to have a well designed and constructed appropriate text set. Telchemy (2008) mentioned the set of English text called “Harvard Sentences”, which is recommended for voice quality measurements (IEEE, 1969). Unfortunately, that list of sentences cannot be used for testing in all countries because some countries have their own languages, for examples, China, Japan, Korea, Thailand, Laos, and Vietnam. Therefore, in those countries, they should have at least one recommended set of sentences in their languages to be used for voice quality measurements in their own environments.

Background

Voice Quality Measurement

Voice quality measurement is mainly classified into subjective and objective methods (Daengsi., 2009; Guéguin et al., 2008; Rango et al., 2006; Goudarzi, 2008), as shown in Figure 1. Subjective listening tests are the most reliable method but cost, time and effort are important issues, because it is complex to design and conduct the test (Goudarzi, 2008; Guéguin et al., 2008). Therefore, objective methods have become more popular, especially a non-intrusive E-Model method, and an intrusive PESQ method (Rango et al., 2006). The output of voice quality measurement is presented in the range of 1-5, which means bad, poor, fair, good and excellent respectively. The average of scores is then called Mean Opinion Score (MOS) which is an official scale issued by ITU-T

(Daengsi, 2009). In Table 1 (Karapantazis and Pavlidou, 2009), the last column shows the MOS of some popular codecs, which are used in VoIP applications, that relates to the levels of call quality as in Table 2 (Telchemy, 2008).

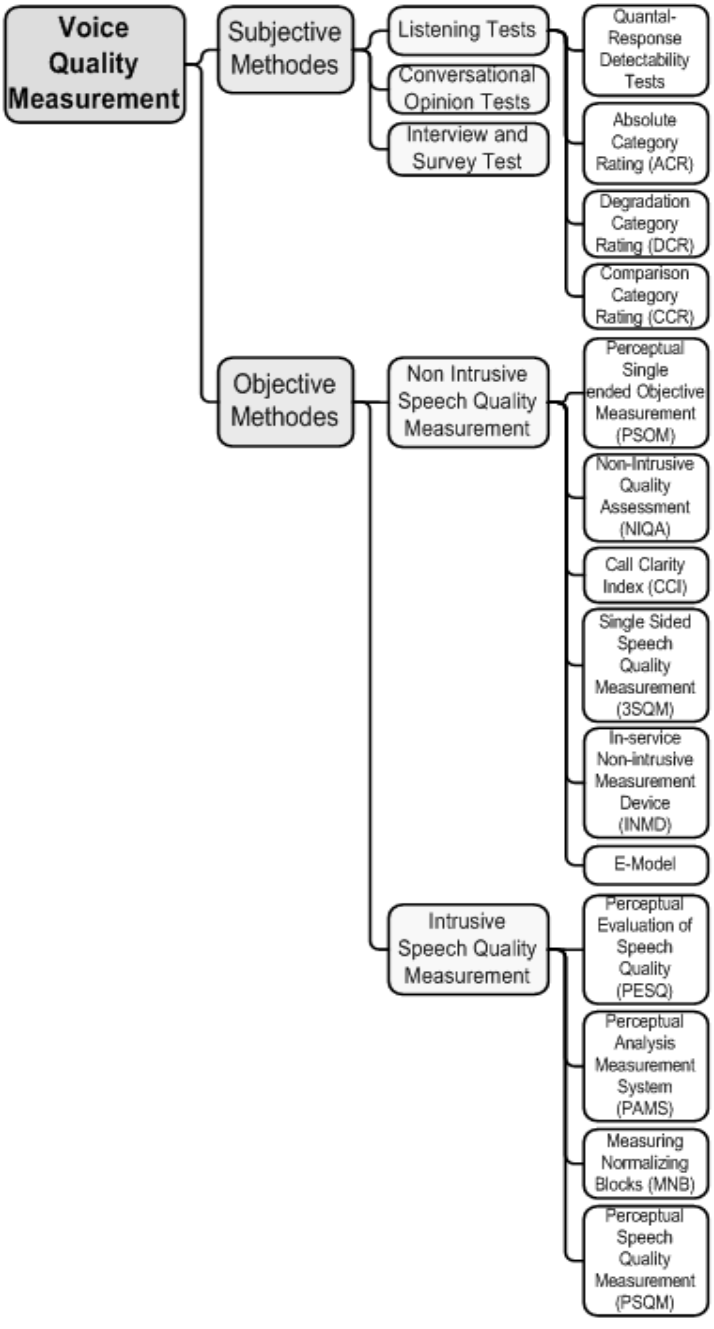


Figure 1. Overview of voice quality measurement methods

Table 1. Some popular codecs with MOS

Codec	Algorithm	Bit rate (Kbps)	Sample size (ms)	MOS
G.711	PCM	64	0.125	4.1
G.723.1	MP-MLQ	6.3	30	3.8
G.723.1	ACELP	5.3	30	3.6
G.729	CS-ACELP	8	10	3.92
G.729a	CS-ACELP	8	10	3.7

Table 2. Call quality levels

User opinion	MOS-ITU scaled	MOS-ACR scaled
very satisfied	4.3-5.0	4.1-5.0
satisfied	4.0-4.3	3.7-4.1
some users satisfied	3.6-4.0	3.4-3.7
many users dissatisfied	3.1-3.6	2.9-3.4
nearly all users dissatisfied	2.6-3.1	2.4-2.9
note recommended	1.0-2.6	1.0-2.4

English Text Sets

For speech or voice quality measurements, IEEE recommended to use the "Harvard Sentences" set (IEEE, 1969), which is a well-known text set and has been used widely (Delhorne et al., 1998; Kabal, 2002; Kain and Macon, 1998; Mansour, 1997; Park et al., 2007; Yankelovich et al., 2006). It is a set of phonetically balanced sentences consisting of 72 lists, each list consists of 10 sentences (each sentence consists of about 5 key words). However, there are some other English text sets for measurements of language, speech and audio processing such as CID Everyday Sentences, CUNY Sentences, Haskins Sentences and TIMIT (Boothroyd, 2006; Delhorne et al., 1998; King, 2008; Lemmetty, 1999; Mansour, 1997, Richie et al., 2009).

Thai Language and Sound System Overview

The general Thai language that all Thai people understand is standard Thai which is the official spoken language used in official places and media broadcasts. Basically, Thai consists of 44 consonants (but 2 consonants are obsolete), 15 basic vowels and 4 tone markers. Similar to English, Thai text is written from left to right in a horizontal direction. However, there is no space between words in the same sentence and no explicit sentence

markers. Vowels can be found before, after, below or above the consonant. Combination of a few consonants and some vowel characters can make diphthongs. There is no article, verb conjugation, declension, object-pronouns, and tenses for the Thai grammar (Wutiwiwatchai and Furui, 2007).

Considering Thai tones, actually there are 5 tones, due to the middle tone unmarked, as shown in Figure. 2 (Wutiwiwatchai and Furui, 2007). The tonal feature is very important in the Thai language because its result changes the meaning of Thai words, as examples shown in Table 3.

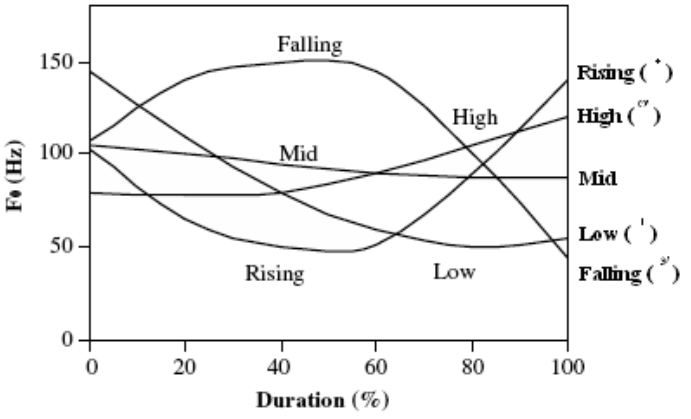


Figure 2. An example of fundamental frequency (F0) contours of five Thai tones by a man

Overview on Thai Text Sets

Before the development of TTST, there were several Thai language sets or resources available to use, summarized by Wutiwiwatchai and Furui (2007) and Daengsi et al. (2010a; 2010b). Those are classified into sets of text and speech. Most of Thai speech sets were associated with their transcription text sets. However, each has a different purpose of development, for examples, NECTEC-ATR and LOTUS. Table 4 shows some Thai Text Set samples (Wutiwiwatchai and Furui, 2007).

The Thai Text Set for Telephono metry (TTST)

TTST is actually the part of Thai Speech Set for Telephonometry (TSST) which was developed with the intention of voice quality measurements and telecommunication research in Thai environments, funded by NECTEC, Thailand (Daengsi et al. 2010a; Daengsi et al. 2010b). It is the part that must be completed before recording of TSST. It was designed in compliance with ITU-T P.800 standard which recommended the appropriate sentences for example:

- simple, meaningful, short sentences
- easy to understand
- very short and very long sentences should be avoided
- each spoken sentence should be short (2–3 seconds)

Table 3. examples of Thai words with similar sounds except tones

Item	Thai tone				
	Mid	**Low**	**Falling**	**High**	**Rising**
1	คา () to be stuck	ข่า () galangal	ข้า/ค่า/฿่า () I or me/value/to kill	ค้า () to trade	ขา () a leg
2	คาว () a bad odor	ข่าว () news	ข้าว () rice	ค้าว () a kind of freshwater fish	ขาว () white
3	ฟา () 4th note	ฝ่า () to violate	ฝ้า () scum	ฟ้า () sky	ฝา () a lid
4	ซอง () an envelope	ส่อง () to shine	ซ่อง () a bawdy house	ซ้อง () to acclaim with one voice	สอง () two
5	ไซ () a fish trap	ใส่ () to put in/on	ไส้ () intestine	ไซ้ () to preen	ใส/ไส () limpid/to plane

Table 4. examples of Thai language resources

Thai resources	Purpose	Details
NECTEC-ATR	Various Thai speech utterances for ASR research	- A set of 5000 frequently-used words - A set of phonetically-balanced sentences - A set of hotel reservation dialogues
LOTUS	Well-designed speech utterances for 5000-word dictation systems	- A set of phonetically-balanced sentences and phonetically-distributed sentences - Three 5000-word covered sets for training, development testing, and evaluation testing

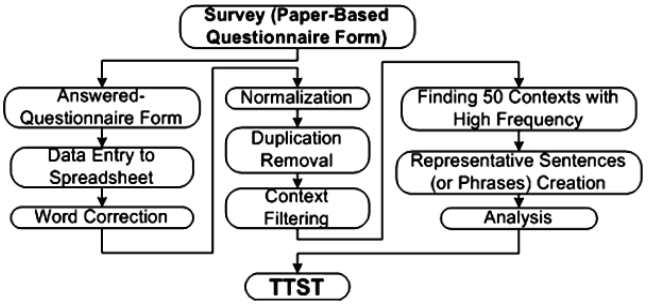


Figure 3. TTST development process

Table 5. 10 of 25 pairs of sentences of TTST

Pair No.	Sentences or Phrases	Meaning
5	ต่อไปเป็นข่าวในพระราชสำนัก	Next, it is the royal news.
	ไปไหนมาเหรอ	Where have you been?
6	กรุณาถือสายรอสักครู่ครับ[a]/ค่ะ[b]	Please hold on one moment.
	ยินดีต้อนรับครับ/ค่ะ	Welcome.
7	วันนี้จะไปเที่ยวที่ไหนดี	Where are we going today?
	ไม่ได้เจอกันตั้งนาน	Long time no see.
8	กำลังจะไปไหน	Where are you going?
	ขอรบกวนเวลาสักครู่นะครับ[a]/คะ[b]	Please give me sometime.
9	วันนี้ไปกินข้าวที่ไหนกันดี	Where will we go to eat today?
	จะกลับเมื่อไหร่	When will you come back?
11	ดูแลรักษาสุขภาพด้วยนะ	Take care of your health.
	ตกลงนะครับ/คะ	Are you ok?
12	กรุณาติดต่อกลับมาใหม่	Please contact again.
	กลับถึงบ้านหรือยัง	Have you reached home?
13	วันนี้เรียนวิชาอะไร	What subject are you going to study today?
	จะกลับถึงบ้านกี่โมง	When will you reach home?
14	กำลังทำอะไรอยู่เหรอ	What are you doing?
	วันนี้รถติดมากเลย	Traffic is/was bad today.
15	ขณะนี้เวลาแปดนาฬิกา	Now, the time is eight a.m.
	ขอโทษนะครับ/คะ	I'm sorry.

a. ครับ (is politeness expression for male.

b. คะ (and ค่ะ (are politeness expression for female.

TTST was from a survey with descriptive statistics, using open questionnaire forms to ask for the most frequently used sentences (or phases) in general conversation, telephone conversation and news broadcasts. The raw data was from answered sentences (or phrases) of over 3,460 from the returned questionnaire forms by over 700 participants. Using a survey can avoid some biases when selecting sentences to be included in TTST. Finally, the TTST consists of 25 pairs of sentences (or phrases). Each pair covers all Thai tones. The methodology to create TTST can be shown in Figure 3, whereas some examples of them are presented in Table 5.

Assessment of TTST

Methodology

From reviewing the Thai resources as in the previous section, all of them are meaningful, unlike English text set of Haskins which is anomalous and not in real life situations (Lemmetty, 1999). However, most of them are large size corpora (Wutiwiwatchai and Furui, 2007; Daengsi et al., 2010a; Daengsi et al. 2010b). Only the text part of LOTUS and NECTEC-ATR are medium in size compared with TTST, which is small in size. Nevertheless, NECTTEC-ATR is actually a part of LOTUS with additional sets of phonetically-balanced words and hotel reservation dialogues. Therefore, only TTST and LOTUS have been compared in this study. However, from the ITU-T recommendation, the issue about level of complication of the sentences is an important issue, therefore, only this issue has been investigated in this study to see whether TTST is more appropriate than LOTUS or not.

Subjective assessment has been selected as the methodology of this study. In general, subjective assessment requires 24–32 subjects with an assumption that the standard deviation of each subject is 0.5. In this study, 30 subjects have been asked to read some sentences of LOTUS (over 220 sentences) and all of TTST. They have to vote on LOTUS and TTST, with five scales as in Table 6, in the issues as follows:

1) complication of structure of sentences (or phrases)

2) complication of comprehension of sentences (or phrases)

However, to vote on the level of complication of Thai text sets, each subject or participant can vote on more than one value of a text set if he/she has an opinion that covers more than one level of complication.

Table 6. Meaning of level of complication for subjects

Subjective opinion	Level of complication
Very hard	5
Hard	4
Fair	3
Simple	2
Very simple	1

Result

The subjective assessment has been conducted using a paper-based questionnaire for assessment of TTST and LOTUS. 90% of 30 participants, 12 female and 18 male, were students in KMUTNB, both undergraduate and graduate. The majority of them are aged between 20-30 years old. After gathering the data from the assessment, the results of the assessment can be presented as in Figure 4 as follows:

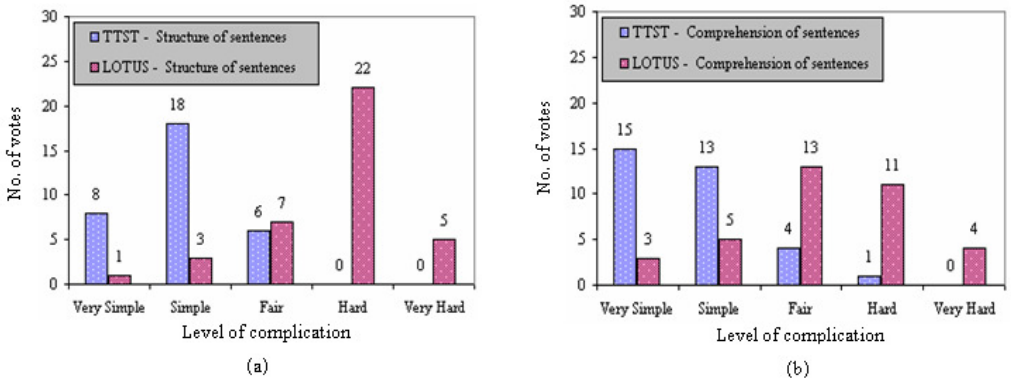


Figure 4. The results from subjective assessment of TTST and LOTUS, about complication of (a) structure of sentences and (b) comprehension of sentences

Comparisons

From reviewing previous English corpora (Boothroyd, 2006; Delhorne et al., 1998; King, 2008; Lemmetty, 1999; Mansour, 1997, Richie et al., 2009), it can be summarized similar to those of English text sets as shown in Table 7, Whereas, for Thai text sets, Table 8 comparatively summarize the key characteristics of TTST and LOTUS, about complication of structure and comprehension, by using the majority results from Figure 4.

Table 7. Comparisons among English text sets

English Text Set	Level of Complication		Example
	Structure	Comprehension	
CID	simple	simple	- Walking is my favorite exercise. - Good morning. - Where are you going?
CUNY	simple - fair	simple - fair	- Have you eaten yet? - How many cats do you have? - The store delivered the new washer and dryer.
Harvard	fair - slightly hard	fair - slightly hard	- Glue the sheet to the dark blue background. - The juice of lemons makes fine punch. - A gold ring will please almost any girl.
Haskins	simple - fair	very hard	- The great car met the milk. - The old corn cost the blood. - The short arm sent the cow
TIMIT	slightly hard - hard	hard	- Don't ask me to carry an oily rag like that. - Why put such a high value on being top dog? - Are planning a strategy of development emphasized sufficiently in your company?

Discussion

From Figure 4 (a) in section IV which is about complication of structure of sentences, it shows clearly that the majority of participants voted 'Hard' for LOTUS and 'Simple' for TTST. Whereas, Figure 4 (b) which is about complication of comprehension of sentences, shows that the majority of participants voted 'fair' and 'Hard' for LOTUS respectively, and 'Very Simple' and 'Simple' for TTST respectively.

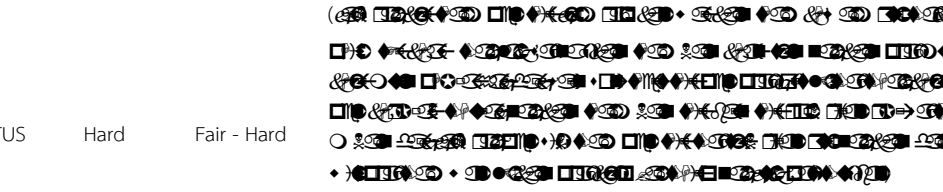
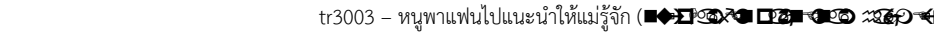
From the Table 7 and 8, one can see, Thai text sets particularly, that TTST consists of short everyday sentences mostly, whereas LOTUS consists of various kinds of sentences, among short, middle, long and very long sentences. Also one can see that TTST can be considered to be used instead of CID and some CUNY sentences, whereas LOTUS can be considered to be used instead of CUNY, Harvard, Haskins and TIMIT in Thai environments. However, TTST is in compliance with ITU-T Recommendation P.800, therefore, it should be used for speech recording instead of Harvard sentences for speech or voice quality measurements in Thai environments although it is simple and it is not phonetically

balanced (note: phonetically balance is not necessary for hearing by human but tone is important particularly for hearing by people that use tonal languages). Whereas, LOTUS is appropriate for applications that require a set of Thai phonetically balanced sentences, such as acoustic modeling in a Thai TTS system. Moreover, TTST could represent a set of everyday sentences; therefore, it could be also applied to test hearing-impaired people or test in hearing aid development, like CID Sentences and CUNY Sentences.

Table 8. TTST vs LOTUS

Thai Text Set	Level of Complication		Example
	Structure	**Comprehension**	
TTST	Simple	Very Simple - Simple	See Table V.
LOTUS	Hard	Fair - Hard	dt0213 – โลกนี้เป็นบ่อเกิดของความเจ็บปวด ([illegible] This world is the cause of pain. et0443 – การมีอารมณ์ที่ไม่ดีย่อมเป็นผลร้ายต่อสุขภาพจิต ([illegible] Bad emotion affects mental health badly.
			pd601 – การทดลองค้นคว้าและการปฏิบัติการฝนหลวงช่วยเหลือราษฎรได้ก้าวหน้าและประสบความสำเร็จเป็นที่น่าพอใจ ([illegible] The research - experiment and performance of the Royal Rain (artificial rain) helps people progressively and successes satisfyingly.
			pd0699 - สำหรับประเทศไทยนั้นเรื่องการแพทย์แผนไทยและสมุนไพรได้ถูกละเลยกันไปประยะหนึ่ง ([illegible] For Thailand, Thai traditional medicine and herb had been ignored for a long period.
			tr0705 – ของสิ่งหนึ่งหรือสถานการณ์อย่างหนึ่ง อาจเป็นที่ปรารถนาของบางคน ([illegible] Something or some situation may be the desire of someone.

Table 8. TTST vs LOTUS (Continued)

Thai Text Set	Level of Complication		Example
	Structure	Comprehension	
LOTUS	Hard	Fair - Hard	tr2962 – อย่างไรก็ตาม เป็นที่ยอมรับกันแล้วว่า การทำความรู้จักกับคอมพิวเตอร์ เข้าใจลักษณะการทำงาน ขั้นตอนในการประมวลผลข้อมูล เพื่อให้ได้สารสนเทศที่เป็นประโยชน์ และจะใช้คอมพิวเตอร์เป็นเครื่องช่วยในการทำงานที่บ้าน ที่โรงเรียน หรือสถานที่ทำงานได้อย่างไร เป็นสิ่งจำเป็นที่จะต้องเรียนรู้ในการดำรงชีวิตประจำวัน และการประกอบอาชีพในยุคปัจจุบัน () However, it has been accepted that knowing computers, understanding its operation, procedures for data processing, to have useful information and how to use computers as tools to work at home, at school or at work is necessary to learn for everyday living and working at present.
			tr3003 – หนูพาแฟนไปแนะนำให้แม่รู้จัก () I invited my boyfriend and introduced him to my mother.

Conclusion

This paper has presented background information of voice quality measurement for speech communication/telecommunication, related English text sets, related information about Thai language and sound system, and Thai test sets, in order to reveal the importance of providing an appropriate set of Thai text, before analyzing and recommending to apply TTST, an ITU-T Recommendation P.800 complied text set, for voice quality measurements in Thai environments, including Thai sounds and Thai users/people. Moreover, it also includes frequently used sentences or everyday sentences and covers all Thai tones, which are important as they change the meaning of Thai words. This research can be applied to other tasks that require Thai everyday sentences for language and speech processing.

Acknowledgment

This paper is dedicated to Dr. Gareth Clayton, Lecturer, Department of Applied Statistics, Faculty of Applied Science, KMUTNB, the advisor of the first author, who passed away. Thank you very much to Mr. Gary Sherriff, International Coordinator, Faculty of Information Technology, KMUTNB, for his support and editing, also all participants for subjective assessment. Gratitude to NECTEC for permission to use TTST and TSST. Lastly, thank you very much to the reviewer for useful comments.

References

Boothroyd, A. (2006). Computer-Assisted Speech PERception testing and training at the SENTence-level (with the F0-only option). Online Available: http://www-rohan.sdsu.edu/~aboothro/files/CASPERSENT/CASPERSENT_MANUAL.pdf

Daengsi, T. (2009). Voice Quality Measurement for VoIP: Simple Method Using a Survey. EECON32.

Daengsi, T. Preechayasomboon, A. Clayton, G. and Wutiwiwatchai, C. (2010a). Development of Thai Text Set for Telephonometry. NCIT2010.

Daengsi, T. Prechayasomboon, A. Sukparungsee, S. Chootrakul, P. and Wutiwiwatchai C. (2010b). The Development of a Thai Speech set for Telephonometry. Oriental-COCOSDA 2010.

Delhorne, L. A. Besing, J. M. Durlanch, N. I. and Reed, C. M. (1998). Tactual Cued Speech as a Supplement to Speechreading. Cue Speech Journal 6: 71-87

Guéguin, M. Bouquin-Jeannès, R. L. Gautier-Turbin, V. Faucon, G. and Barriac, V. (2008). On the Evaluation of the Conversational Speech Quality in Telecommunications. EURASIP journal on Advances in Signal Processing. Online Available: http://downloads.hindawi.com/journals/asp/2008/185248.pdf

Havelock, D.I. and Green, D. (2005). Measuring the Dynamic Performance of Video Conferencing Systems. Proceeding of The Canadian Acoustical Association Conference.

IEEE (1969) Recommended practice for speech quality measurements. IEEE Trans. on Audio Electroacoust 17(3): 225–246

ITU-T Recommendation P.800. (1996a). Methods for subjective determination of transmission quality.

ITU-T Recommendation P.830. (1996b). Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs.

Kabal, P. (2002). TSP speech database. Tech. Rep. Depart. of Elect. & Comput. Eng. McGill University.

Kain, A. and Macon, M. (1998). Personalizing a speech synthesizer by voice adaptation. Proc. 3rd ESCA/COCOSDA International Speech Synthesis Workshop 225-230

Karapantazis, S. and Pavlidou, F.-N. (2009). Voip: A comprehensive survey on a promising technology. Comput. Networks 53(12): 2050-2090

King, S. E. (2008). Evaluation of sentence list equivalency for the TIMIT sentences by cochlear implant recipients. Independent Studies and capstones. Washington Universit School of Medicine.

Lemmetty, S. (1999). Review of speech synthesis technology. M.S. Thesis. Dept. Elect. Commun. Eng. Helsinki Univ. of Technology.

Mansour, S. A. (1997). Effects of Amplitude Envelope Cues as an Aid to Enhanced Speechreading. M.Eng. Thesis. Massachusetts Inst. of Technology.

Park, H. K. Bradley, J. S. and Gover, B. N. (2007). Rating sound insulation in terms of speech intelligibility. Int. Congr. on Acoust.

Rango, F. D. Tropea, M. Fazio, P. and Marano, S. (2006). Overview on VoIP: Subjective and Objective Measurement Methods. International Journal of Computer Science and Network Security 6(1B): 140-153

Richie, C. Warburton, S. and Carter, M. (2009). The Butler University Audio-Video Spoken Language Corpus. Online Available: http://www.ldc.upenn.edu/Catalog/docs/LDC2009V01/AV_Database_Spoken_American_English.pdf

Telchemy (2008). Voice Quality Measurement. Application Note.

Goudarzi M. (2008). Evaluation of Voice Quality in 3G Mobile Networks. M.S. Thesis. School of Computing. Communications and Electronics. Univ. of Plymouth.

Wutiwiwatchai, C. and Furui, S. (2007). Thai speech processing technology: A review. Speech Communication 49: 8-27.

Yankelovich, N. Kaplan, J. Provino, J. Wessler, M. and DiMicco, J. M. (2006). Improving audio conferencing: are two ears better than one? Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work 333–342.

❑❑❑❑❑