

Constructing a short reading passage with phonemic and phonetic coverage for Thai: Traimit passage

Chutamane Onsuwan^{1,2}, Kanwara Bairerkdee¹, Juthatip Duangmal¹, Nawasri Chonmahatrakul¹, Atita Amornlaksananon³, Pimchanok Denmuenwong¹

¹ Department of Linguistics, Thammasat University, Thailand

² Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University, Thailand

³ Faculty of Liberal Arts, Rajamangala University of Technology Phra Nakhon, Thailand
consuwan@tu.ac.th

ABSTRACT

This work describes main characteristics of an alternative Thai reading passage, Traimit, as compared with existing passages (i.e., the translated North Wind and the Sun). Traimit was designed to meet a set of criteria which could be beneficial for Thai linguistics and speech research and possibly a potential tool for quick speech assessment. The 102-word passage is short, contemporary, and easily read. It contains a complete Thai phonemic inventory (including clusters and diphthongs) and takes into account five lexical tones in various rhyme units. There are mono- and multisyllabic words as well as utterances/sentences in various length. Traimit has neutral contents with more than 40% of words being “high frequency” and a small set of words repeated to allow for observation of production consistency. Measurements of a subset of vowels from two Thai adults reading the passage revealed its promise of extracting the vowel phonemes from the running speech.

Keywords: Thai, phoneme, vowel, reading passage, speech data

1. BACKGROUND

Standard reading passages/texts (in various languages) have proven to be useful materials for speech science researchers and clinicians as they provide quick and structured samples of the individual’s speech and characteristics [1]. Situated within natural reading contexts, they were believed by some to be “closet to spontaneous speech compared to isolated words or sentences” [1].

While some passages were constructed for a general research purpose such as the (English) North Wind and the Sun [2][3], others have been designed and tailored for specific tasks such as the (English) Caterpillar passage for assessment of motor speech disorders [4]. Although a great variety of passages and descriptions/comparisons of their features have

existed in many languages, it is only recently that general consensus guidelines have been proposed [1].

Passage features that have often been discussed are phonemic inventory, inclusion of clusters, presence of vowel-to-vowel coarticulation, phonetic balance, passage length and topic, mean length of utterance, word familiarity, reading level, prosodic variations, word length and (syllable) complexity, and word repetition [1][4]. Apart from phonemic coverage (i.e., covering all phonemes), allophonic coverage (i.e., covering many positional contexts—word positions, lexical stress, and proximity to other phonemes) has been analysed and quantified [5].

Following ratings and selections by experts in the international Delphi consensus study, a minimal set of criteria for standard reading passages for an overall speech and voice assessment in adolescents and adults was put forward [1]. The criteria (ranked from highest to lowest) include (1) complete phonemic inventory (2) phonetic balance (3) inclusion of various intonation patterns (4) control of the phonetic complexity of words (5) covering the positions of the phonemes in the words (as well as their vowel/consonant context) (6) inclusion of consonant clusters (7) repeated presence of some words (8) inclusion of various sentence lengths [1].

Various forms of short reading passages have been around for English and Indo-European languages [1][4][5]. To our knowledge, a few reading passages (see Section 2) have been available for Thai, one for a general research propose and the other two for a small clinical survey. Traimit passage was designed mainly for speech and linguistics research practice. Our goal was to create a relatively short, simple, contemporary and well-structured text, with neutral contents and natural sounding. It should contain all phonemes and capture different rhyme units to allow for study of tonal production/variation in checked and unchecked syllables—one of important aspects of Thai [6][7].

With further evaluations and modification, it may become an effective tool for both clinical and research purposes. Nevertheless, it should not be used

as a reading proficiency test and certainly not recommended for illiterate populations or young children.

2. THAI READING PASSAGES

Four short reading passages in Thai script along with broad phonemic transcription (according to the conventions given in [8]) are provided (hence no reflections of allophonic variations). Only English translation of the Traimit passage is given. As Thai writing system is usually written with no space between words and a space is often (but not always) used to indicate an end of a phrase, a sentence, or between separate items in a list, here a “|” indicates a space in typed written passage and often corresponds to a break or pause between words and utterances/ sentences in real production. Importantly, main characteristics and limitations of each passage are discussed.

2.1. North Wind and the Sun passage translated to Thai (NW-Th) [8]

The (original) North Wind and the Sun passage is a well-known English standard text (113 words) which has allowed a great number of researchers to compare and describe pronunciation of different English varieties despite absence of some English phonemes [2][3]. In JIPA Illustrations, the International Phonetic Association has encouraged researchers from different languages to capture and document language phoneme diversity using the translated versions of North Wind and the Sun passage [2][3]. For each language in the JIPA Illustrations, there are descriptions of sound inventory and sets of minimal pairs followed by the translated North Wind and the Sun passage, which could be considered as a supplement [8]. However, there are many shortcomings with the Thai version (NW-Th) as in many translated versions in other languages [2]. The NW-Th has 118 words and 172 syllables. Since there are high repetitions of words (26 words get repeated, some more than three times), it fails to capture 14 Thai phonemes (see Section 3).

Thai script: ขณะที่ลมเหนือและพระอาทิตย์กำลังเถียงกันว่าใครจะมีพลังมากกว่ากันก็มีนักเดินทางผู้หนึ่งเดินผ่านมา|ใส่เสื้อกันหนาว|ลมเหนือและพระอาทิตย์จึงตกลงกันว่าใครที่สามารถทำให้นักเดินทางผู้นี้ถอดเสื้อกันหนาวออกได้สำเร็จก่อน|จะถือว่าเป็นผู้ที่มีพลังมากกว่า|และแล้วลมเหนือก็กระพือพืดอย่างสุดแรง|แต่ยิ่งพัดแรงมากขึ้นเพียงใด|นักเดินทางก็ยังดึงเสื้อกันหนาวให้กระชับกับตัวมากขึ้นเพียงนั้น|และในที่สุดลมเหนือก็เลิกล้มความพยายาม|จากนั้นพระอาทิตย์จึงสาดแสงอันร้อนแรงออกมา|นักเดินทางก็ถอดเสื้อกันหนาวออกทันที|ในที่สุดลมเหนือจึงจำต้องยอมรับว่าพระอาทิตย์มีพลังมากกว่าตน

(A “|” corresponds to a space in the typed passage in Thai script (p. 27) of [8])

Transcription: [kʰàʔ.nàʔ.tʰiː.lom.nú.a.léʔ.pʰráʔ.ʔaː.tʰít.kam.laŋ.tʰiǎŋ.kan.wáː.kʰraj.tɕàʔ.miː.pʰáʔ.laŋ.mâːk.kwàː.kan|kʰiː.miː.nák.

k.dʰɔːm.tʰa.ŋ.pʰúː.nùŋ.dʰɔːn.pʰàːn.maː|sàʔ.sú.a.kan.náːw|lom.nú.a.léʔ.pʰráʔ.ʔaː.tʰít.tɕuŋ.tòk.loŋ.kan.wáː|kʰraj.tʰiː.sáː.mâːt.tʰa.m.háj.nák.dʰɔːn.tʰa.ŋ.pʰúː.níː.tʰòːt.sú.a.kan.náːw.ʔòːk.dâːj.sám.r.rét.kòːm|tɕàʔ.tʰiː.wáː.pen.pʰúː.tʰiː.miː.pʰáʔ.laŋ.mâːk.kwàː|léʔ.lé.éːw|lom.nú.a.kʰiː.kràʔ.pʰuː.pʰát.jàːŋ.sùt.rɛːŋ|tɛː.jiŋ.pʰát.rɛːŋ.m.mâːk.kʰúŋ.pʰiǎŋ.daj|nák.dʰɔːn.tʰa.ŋ.kʰiː.jiŋ.duŋ.sú.a.kan.náːw.háj.kràʔ.tɕʰáp.kàp.tua.mâːk.kʰúŋ.pʰiǎŋ.nán|léʔ.naj.tʰiː.sùt.lom.m.nú.a.kʰiː.lɛːk.lóm.kʰwam.pʰáʔ.jaːjaːm|tɕàːk.nán.pʰráʔ.ʔaː.tʰít.tɕuŋ.sàːt.séːŋ.ʔan.rɔːŋ.rɛːŋ.ʔòːk.maː|nák.dʰɔːn.tʰa.ŋ.kʰiː.tʰòːt.s.sú.a.kan.náːw.ʔòːk.tʰán.tʰiː|naj.tʰiː.sùt.lom.nú.a.tɕuŋ.tɕam.tʰiŋ.j.jɔːm.ráp.wáː.pʰráʔ.ʔaː.tʰít.miː.pʰáʔ.laŋ.mâːk.kwàː.ton]

2.2. Rama-I (RM-I) and Rama-II passages (RM-II) [9]

The RM-I and RM-II passages were constructed for a different purpose from the NW-Th. They were designed for an investigation of speech defects in Thai school children with more focus on consonants than vowels [9]. Many vowel phonemes were not taken into account even when the two passages were combined (see Section 3). The RM-I put an emphasis on initial and final consonants whereas the RM-II on the initial cluster set (see Section 3) [9]. The RM-I contains 64 words and 100 syllables.

Thai script: ผนพามีวีแรววว่าจะตจ|ยายฉิมชวนหนุแฉวจเรือไปหากันน่งที่บางปอ|หนุแฉวพอกหน้าเปบึงด้วยคินสอพองดูงามดี|พอถึงที่ผงพมาติดตามต่อนหน้าต่อนหลัง|ป้าเอื้อนอะอะเอ็ดคะโรแล้วอ้อมเบิกบาน|เจ้าแก้วหลานรักริมาทักทาย|ผู้ใหญ่พูดคุยหัวเราะขบขัน|สองคนซุกซนเล่นซ่อนหา|หนุแฉวตกต้นไม้ร้องไห้แง|ยายฉิมว่าอย่าจับกันขังไม่พ้ง

Transcription: [fõn.fáː.miː.wíː.wɛːw.wáː.tɕàʔ.tòk|jaːj.tɕʰim.tɕʰu.an.núː.tɕʰɛw.tɕɛw.ru.a.paj.háː.kam.nan.tɕʰɛŋ.tʰiː.ba.ŋ.bòː|núː.tɕʰɛw.pʰiːk.náː.pàʔ.pɛːŋ.dú.j.din.sɔː.pʰòːŋ.duː.ŋaːm.diː|pʰwɔː.tʰiŋ.tʰiː.fũ.ŋ.máː.tít.taːm.tòːm.náː.tòːm.láŋ|páː.jú.an.ʔéʔ.ʔáʔ.ʔét.tàʔ.roː.léːw.jím.bɔːk.baːŋ|tɕáw.kéːw.láːn.rák.ríː.maː.tʰák.tʰaːj|pʰúː.jàj.pʰúː.t.kʰuj.hú.a.róʔ.kʰòp.kʰán|sɔʔ.ŋ.kʰon.súk.son.lén.sòːm.háː|núː.tɕʰɛw.tòk.tõn.máːj.róʔ.ŋ.háːj.ŋuː.ŋɛː|jaːj.tɕʰim.wáː.jɛː.tɕàp.kõn.kòp.jaŋ.máj.pʰaŋ]

The RM-II has 61 words and 72 syllables.

Thai script: หนุแฉวกินกล้วยเพลิน|เปลลถูกเพื่อนผลักกลางกรุงคิ่นซุกซน|พระไม่เห็นใครก็ว่าแขนไว้มีบ|ครูปอบแฉว|แฉวกราบพระ|ครูดตรวจตราพบขาดผลกว้างกว่า|นี้|จริงใช้ผ้าโปร่งปรับเป็นผ้าพันแผล|แล้วส่งหมอนามัยกลางตรอก|แฉวไม่จับตัวเปล่า|คราวนี้จำเลยขริมโดนกวาดร้องฮักฮัก

Transcription: [núː.tɕʰɛw.kin.klú.japʰɔːn|pʰɔːn.tʰiː.k.pʰiŋ.an.pʰi.àk.loŋ.kʰiːwɔːŋ.kla.ŋ.kruŋ.dín.kʰiːk.kʰiːk.pʰráʔ.máj.hén.kʰraj.kʰiː.kʰwáː.kʰiːn.wáj.máp|kʰruː.plòːp.tɕʰɛw|tɕʰɛw.kràːp.pʰráʔ|kʰruː.t.rùat.tràː.pʰóp.bàːt.pʰiː.kwáːŋ.kwàː.níw|tɕuŋ.tɕʰáʔ.pʰáː.p.ròːŋ.p.ràp.pen.pʰáː.pʰan.pʰiː|léːw.sòŋ.mɔː.ʔáʔ.naː.maj.kla.ŋ.tròːk|tɕɛw.máj.tɕɛp.tua.plàːw|kʰraːw.níː.tɕam.ɔːj.kʰiːrúm.dom.kùat.róː.ŋ.hák.hák]

2.3. Traimit passage (TM)

The TM was created primarily for speech and linguistics research practice, but has also met a set of criteria—complete phonemic inventory, control of the phonetic complexity of words, covering a number of positions of the phonemes in the words (especially the rhymes), inclusion of consonant clusters, repeated

presence of some words and inclusion of various sentence lengths which could potentially be useful as a material for quick speech assessment [1].

We believe that a text adapted from traditional stories and tales would yield high repetitions of words while a text deliberately created from minimal pairs would make it unnatural. The TM was a random text found in a school textbook and (subsequently) heavily modified to cover all Thai phonemes—at least one instance of each phoneme (see Section 3) with a reasonable amount of pauses across the passage.

2.3.1. Words and utterances

It contains 102 mono- and multi syllable words and 139 syllables. Of those, 69 were content words, 29 function words and 4 proper nouns. Moreover, 47.6% of words are “high frequency” (occurring more than 3,000 times) and 34% of words occurring between 100-2,900 times based on Thai corpus data [10]. The TM is composed of 25 utterances/sentences in various length (1 to 10 syllables per utterance; average of 5.6 syllables per utterance). The text contains 5 minimal sets, 14 minimal pairs, and 8 near-minimal pairs.

2.3.2. Repetitions of words

The TM has low repetitions of words to allow for observation of production consistency. Thirteen words get repeated once and one word five times.

2.3.3. Rhyme units

One important aspect of the TM is that it could be useful for a research investigation on tonal production and variation across various checked and unchecked syllables [6] in natural speech. Table 1 summarizes the TM rhyme-tone distribution, shaded cells show three only gaps in the distribution. A “*” indicates a combination that is deemed as ‘rare occurrence’ and “**” as ‘no occurrence’ in the language [7].

	mid	low	falling	high	rising
short V + stop	0**	18	0*	21	0**
short V + nasal	14	2	2	0	5
short V + glide	6	1	1	0	1
long V + stop	0**	5	5	1*	0**
long V + nasal	15	1	5	4	5
long V + glide	1	0	3	4	1
long V + no coda	7	4	4	1	2

Table 1: Rhyme-tone distribution in the TM.

Thai script: โรงเรียนไตรมิตรรามอินทราของฉันทและเธอ|เป็นโรงเรียนเล็กเล็กริมคลองบาง
ซื่อ|อยู่ห่างจากหมู่บ้านสองแคว|ประมาณหนึ่งกิโลเมตร|พวกเราสมัครใจสมัคร|เราฝึกปลูกพืชผัก
สวนครัว|และผลไม้หลายชนิด|เช่น|ต้นกะเพรา|มะเขือเปราะ|ผักกระเฉด|กล้วยน้ำว้า|เงาะ|ลูก
พลับ|และมะเฟืองหวาน|โดยฉันทกับเพื่อนเพื่อน|จะช่วยกันรดน้ำ|พรวนดินและใส่ปุ๋ย|จนผักผลไม้
งอกงามอยู่เสมอ|ในตอนเช้าและตอนเย็น|น้องน้องทุกคน|ช่วยกันเก็บขยะที่ละหอย|กวาดใบไม้
แห้งที่ร่วงหล่น|ลงบนถนนรอบอาคารเรียน|และหน้าเสาธงจนสะอาด

Transcription: [ro.ŋ.rian.traj.mít.ra.m.ŋin.tʰra:kʰɔ̃:ŋ.tɕʰǎn.léʔ.t
ʰɔ̃:|pen.ro.ŋ.rian.lék.lék.rim.kʰɔ̃:ŋ.ba.ŋ.sú:|jù:.hà.ŋ.tɕà:k.mù:
bâ:n.sǎŋ.kʰwe:|pràʔ.ma.m.nùŋ.kíʔ.lo:mét|pʰúak.raw.mi:sǔa
n.tít.pà:.doŋ|raw.fùk.plù:k.pʰú:t.pʰàk.sǔan.kʰrua|léʔ.pʰǎn.láʔ.
máj.lǎj.tɕʰáʔ.nít|tɕʰǎn|tǎn.káʔ.pʰraw|máʔ.kʰúa.prɔʔ|pʰàk.krà
ʔ.tɕʰɛrt|klúaj.ná:m.wá:|ŋɔʔ|lú:k.pʰláp|léʔ.máʔ.fuaŋ.wá:n|doj.t
ɕʰǎn.kàp.pʰúan.pʰúan|tɕàʔ.tɕʰúaj.kan.rót.ná:m|pʰruan.din.lé
ʔ.sàj.pǔj|tɕon.pʰàk.pʰǎn.láʔ.máj.ŋí:k.ŋam.jù:sàʔ.mɔʔ|naj.tɔ:m.
tɕʰá:w.léʔ.tɔ:n.jen|nó:ŋ.nó:ŋ.tʰúk.kʰon|tɕʰúaj.kan.kèp.kʰáʔ.jàʔ.
tʰi:lǎʔ.tʰɔʔ|kwà:t.baj.máj.hé:ŋ.tʰi.rúaj.lòŋ|loŋ.bon.tʰáʔ.nón.r
ɔ:p.ʔa:kʰa.n.rian|léʔ.ná:sǎw.tʰoŋ.tɕon.sàʔ.ʔà:t]

English translation: Our Traimit Ram-Intra School is a small school located near Bang-Sue Canal and is about one kilometre from Song-Kwae Village. We have a garden near the forest. We practice growing plants, vegetables, and many kinds of fruit trees such as basil, eggplant, water spinach, banana, rambutan, persimmon, and sweet star apple. My friends and I always water, loosen the soil, and fertilize vegetables and fruit trees until they grow well. In the morning and evening, all students help picking up dirty trash and sweeping fallen dry leaves around the school building and the area in front of an in-ground flagpole.

3. PHONEME DISTRIBUTIONS

Table 2 summarizes all phoneme distributions (i.e., initial consonant, initial cluster, vowel, lexical tone, and final consonant) across the four passages, A “✓” indicates a complete distribution while “×” marks the incomplete. The TM gets all “✓” marks.

Phoneme distribution of 21 initial consonants and 12 initial clusters from the four passages is given in Table 3, of nine finals in Table 4, of 18 monophthongs and three diphthongs in Table 5 and that of five tones in Table 6. It is important to note that the TM has covered all of the Thai phonemes.

	initial C	initial cluster	monoph thong	diph thong	tone	final C
NW-Th	×	×	×	✓	✓	✓
RM-I	✓	×	×	×	✓	✓
RM-II	×	×	×	×	✓	✓
TM	✓	✓	✓	✓	✓	✓

Table 2: Summary of phoneme distributions across four passages.

initial C	NW -Th	RM -I	RM -II	TM	initial C	NW -Th	RM -I	RM -II	TM
p	1	4	1	3	r	6	6	1	12
pʰ	13	6	5	9	l	17	4	3	16
b	0	4	1	4	j	6	7	0	4
t	5	8	1	4	w	4	4	1	2
tʰ	20	5	1	7	pr	0	0	2	2
d	8	4	2	3	pʰr	4	0	2	2
k	15	4	3	6	pl	0	0	2	1
kʰ	3	4	1	5	pʰl	0	0	5	1
ʔ	8	3	1	3	tr	0	0	3	1
te	7	7	7	4	tʰr	0	0	0	1
teʰ	1	4	1	8	kr	0	0	2	1
f	0	3	0	2	kʰr	4	0	5	1
s	12	5	1	8	kl	0	0	3	1
h	2	4	3	2	kʰl	0	0	3	1
m	12	5	5	11	kw	3	0	2	1
n	20	6	4	9	kʰw	1	0	1	1
ŋ	0	3	0	3					

Table 3: Initial consonant distribution.

final C	NW -Th	RM I	RM -II	TM
p	3	3	6	4
t	14	2	3	9
k	15	7	7	12
ʔ	18	7	3	25
m	13	6	2	5

final C	NW -Th	RM -I	RM -II	TM
n	23	16	9	32
ŋ	25	11	10	16
j	9	10	8	13
w	5	8	8	5
no coda	47	30	16	18

Table 4: Final consonant distribution.

vowel	NW -Th	RM -I	RM -II	TM
i	6	5	3	7
i:	11	6	1	3
e	2	3	3	6
e:	0	0	0	2
ɛ	4	4	4	6
ɛ:	6	7	4	2
a	50	17	17	31
a:	36	20	13	23
u	7	1	2	2
u:	2	0	0	2
ɤ	0	0	0	2

vowel	NW -Th	RM -I	RM -II	TM
ɤ:	6	1	3	2
u	3	2	2	2
u:	3	7	4	5
o	9	9	3	13
o:	0	1	2	4
ɔ	1	1	0	2
ɔ:	13	11	6	9
ia	3	0	0	3
ua	9	2	1	4
ua	1	3	4	9

Table 5: Monophthong and diphthong distribution.

tone	NW -Th	RM -I	RM -II	TM
mid	64	30	21	43
low	28	16	18	31
falling	36	25	9	20
high	30	9	12	31
rising	14	20	12	14

Table 6: Lexical tone distribution.

4. VOWEL MEASUREMENTS FROM THE TRAIMIT PASSAGE

Recordings of the TM passage have been made by 134 Thai adults (43 healthy controls, 45 adults with clinical depression and 46 adults with other psychiatric conditions) as part of spoken data profile for each individual in an ongoing research project “Artificial intelligence for screening of depression and other psychiatric conditions with speech features analysis from a Thai psychiatric screening test” [11][12]. The recordings were made with a designed mobile application onto a smartphone in a quiet room at Siriraj Hospital (Bangkok) at a sampling rate of 16.0 kHz.

To demonstrate its utility in extracting the vowel phonemes, the first two formants of a subset of vowels from two healthy adults (a 22-year old female (Figure 1) and a 27-year-old male (Figure 2)) reading the passage were measured at the vowel midpoint using PRAAT Version 6.0.05. One token per vowel was measured for the 18 monophthongs of each speaker. The words were [tit mī: kəp tɛh:t léʔ hɛ:ŋ phək tɛa:k fùk phú:t thʔ thɤ: thúk lú:k lɔŋ lɔ: ŋɔʔ and ŋɔ:k]. Measurements from more than one token form each vowel would be ideal, but as can be seen in Figures 1 and 2, the measurements from just one token for each vowel could produce a quick phonetic depiction of the speaker’s vowel space.

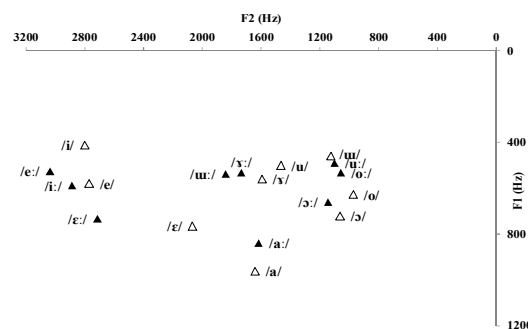


Figure 1: Plot of the first two formants for the Thai female reading the TM passage.

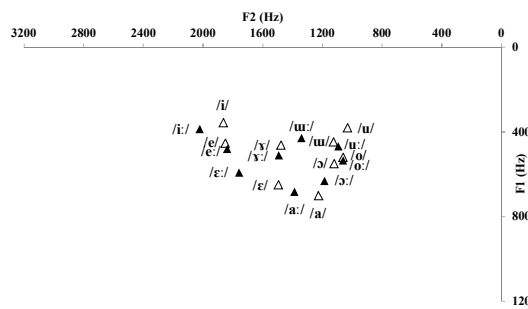


Figure 2: Plot of the first two formants for the Thai male reading the TM passage.

5. CONCLUSION AND FUTURE DIRECTION

As shortcomings of the Thai reading passages were discussed, the TM has been shown to address some of the limitations and could be useful for Thai linguists and speech researchers. With further evaluations and adjustments, it could be a good candidate for a speech assessing task. Being relatively short and simple, the TM may be used in conjunction with designed word list(s) allowing for supplementary phonetic descriptions in controlled phonetics study.

There are still areas of improvement for the TM: modification to allow for various prosodic variations [1][4], incorporation of more words of increasing complexity [1][4], and inclusion of words to fill in the three rhyme-tone gaps in Table 1. Future work should include further evaluations in extracting other Thai phonemes (especially the tones) and evaluations of its level of phonetic balance [13]. Preliminary examination of the recordings in Section 4 has suggested that the word แม็กว [k^hwɛ:]₁, which has contributed to many instances of pronunciation error, should be changed to [k^hwɛ:n]₂ แม็กวัน and that caution should be taken for analysis of segmental duration as the TM was not controlled for pattern of stressed and unstressed words across an utterance/sentence.

6. ACKNOWLEDGEMENTS

We thank Thanaporn Anansiripinyo for her role involving an earlier version of the TM passage and Kankamol Jaisin, M.D., Keerati Pattanaseri, M.D., Juthawadee Lortrakul, M.D., for their role in data collection process at Siriraj Hospital (Bangkok).

7. REFERENCES

- [1] Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. 2022. Criteria for creating new standard reading passages for the assessment of speech and voice: A Delphi consensus study. *Clinical Linguistics & Phonetics*, 1-20.
- [2] Baird, L., Evans, N., & Greenhill, S. J. 2022. Blowing in the wind: Using 'North Wind and the Sun' texts to sample phoneme inventories. *Journal of the International Phonetic Association*, Vol.52, no.3, 453-494.
- [3] Deterding, D. 2006. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, Vol.36, no.2, 187-196.
- [4] Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. 2013. "The caterpillar": a novel reading passage for assessment of motor speech disorders. *American Journal of Speech-language Pathology*, Vol.22, no.1, 1-9. [https://doi.org/10.1044/1058-0360\(2012/11-0134\)](https://doi.org/10.1044/1058-0360(2012/11-0134))
- [5] Gurevich, N., & Kim, H. 2022. Examination of Consonantal Phonetic Coverage in Standard Reading Passages. *Perspectives of the ASHA Special Interest Groups*, Vol.7, no.5, 1573-1582.
- [6] Gedney, W.J. 1989. A checklist for determining tones in Tai dialects. In R.J. Bickner, J. Hartmann, T.J. Hudak and P. Peyasantiwong (Ed.), *Selected Papers on Comparative Tai Studies*, 191-206. Center for South and Southeast Asian studies, the University of Michigan.
- [7] Gandour, J. 1974. Consonant types and tone in Siamese. *Journal of Phonetics*, Vol2, no.4, 337-350.
- [8] Tingsabadh, M. K., & Abramson, A. S.1993. Thai. *Journal of the International Phonetic Association*, Vol.23, no.1, 24-28.
- [9] Sindermsuk, D. 1986. *The survey of speech defects among prathom 4 students in Mitsampan school group*. MA. dissertation, Mahidol University.
- [10] Kosawat, K., Boriboon, M., Chotrakool, P., Chotimongkol, A., Klaithin, S., Kongyoung, S., Kriengkhet, K., Phaholphinyo, S., Purodakananda, S., Thanakulwarapas, T. & Wutiwiwatchai, C. 2009, October. BEST 2009: Thai word segmentation software contest. In 2009 *Eighth International Symposium on Natural Language Processing* (pp. 83-88). IEEE.
- [11] Klangpornkun, N., Ruangritchai, M., Munthuli, A., Onsuwan, C., Jaisin, K., Pattanaseri, K., Lortrakul, J., Thanakulakkarachai, P., Anansiripinyo, T., Amornlaksananon, A., Laohawee, S., & Tantibundhit, C. 2021, November. Classification of Depression and Other Psychiatric Conditions Using Speech Features Extracted from a Thai Psychiatric and Verbal Screening Test. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 651-656). IEEE.
- [12] Munthuli, A., Pooprasert, P., Klangpornkun, N., Phienphanich, P., Onsuwan, C., Jaisin, K., Pattanaseri, K., Lortrakul, J., & Tantibundhit, C. (to appear). Classification and analysis of text transcription from Thai Depression Assessment Tasks among patients with depression. *PLoS ONE*.
- [13] Munthuli, A., Sirimujalin, P., Tantibundhit, C., Kosawat, K., & Onsuwan, C. 2013. A corpus-based study of phoneme distribution in Thai. In *10th International Symposium on Natural Language Processing* (pp. 114-121).