Thai Phonetically Balanced Word Recognition Test: Test-retest Reliability and Error Analysis

A. Munthuli[1], *C. Tantibundhit*[1, 2], *C. Onsuwan*[2, 3], *K. Kosawat*[4], *N. Klangpornkun*[1]

[1] Department of Electrical and Computer Engineering, Thammasat University, Thailand
[2] Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University, Thailand
[3] Department of Linguistics, Thammasat University, Thailand
[4] National Electronics and Computer Technology Center (NECTEC), Thailand

tchartur@engr.tu.ac.th, consuwan@tu.ac.th, krit.kosawat@nectec.or.th

Abstract

To develop a new Thai word recognition test with emphasis on phonetic balance, reliability, and inter-list equivalency, Thammasat University Phonetically Balanced Word Lists 2014 were created with five different lists. TU PB'14 reflects Thai phoneme distribution [1] based on large-scale written Thai corpora, InterBEST [2]. To further evaluate its test-retest reliability, the lists are given in test and retest sessions to 30 normal-hearing subjects. Percent correct discrimination scores between the two sessions are not significantly different. Detailed analysis of listeners' errors reveals that errors occurred predominantly in the case of initial only, final only, and initial along with final consonants.

Index Terms: Thai, phonetically balanced word lists, speech audiometry, TU PB'14, error analysis

1. Introduction

One of the tools, which is widely used in measuring word recognition score (WRS) is phonetically balanced word lists (PB lists). A form of PB list is usually employed during a hearing examination session. Therefore, to prevent learning effect and memorization, several test lists, which are interchangeable, are available [3].

Currently, there is a set of five word lists (henceforth OTL), each with 25 monosyllabic words, being used in hearing clinics across Thailand. However, they have several shortcomings. Firstly, there is a large degree of asymmetrical phoneme occurrences among the lists. It is unlikely that the lists were evolved from a set of reliable phoneme distribution data. Secondly, there are several cases of duplicate words across different lists.

Focusing on three major criteria, phonetic balance, reliability, and list equivalency, Thammasat University Phonetically Balanced Word Lists 2014 (TU PB'14) were created.

2. Development of TU PB'14

Due to the lack of large spoken corpus in the Thai language [1], we have obtained phoneme distributions from existing Thai large-scale written corpora, InterBEST [2] [4], composed of approximately nine million words divided into 12 genres [2]. Since Thai does not use spaces between words, a grapheme to phoneme (G2P) software [5] is needed to break a sentence into words and generate their pronunciations. Each word was transliterated. Finally, phoneme distributions of initial consonants, vowels, final consonants, and tones were obtained [1].

Details of a design and construction process of TU PB'14 word lists are given in [6] and are briefly summarized here. To create five lists, each with 25 monosyllabic words (CVC or CVV(C)), relative frequencies (%) of 81 Thai phonemes (consonants, vowels, and lexical tones) were multiplied by 125 and rounded to the nearest integer. Then, each phoneme was equally distributed as much as possible into each list as shown in Tables 1–2.

Words with desirable phoneme configuration were pooled and selected according to their commonness and familiarity, i.e., they are learned at an elementary school level (a minimum education required for all Thais). The complete TU PB'14 word lists are shown in Table 3. It is important to note that more than 70% of words in each list are ranged between average to high frequency based on number of occurrences in InterBest [2].

To assure test validity and inter-list equivalency, all 10 lists of TU PB'14 and OTL were given to 30 normal-hearing subjects to obtain discrimination scores in five intensity levels, i.e., 15, 25, 35, 45, and 55 dB HL [6]. Discrimination scores of the five TU PB'14 lists were highly comparable with those of the OTL lists, with those of the former being slightly lower (more difficult). Importantly, all TU PB'14 lists exhibited relatively equal range of difficulty. Therefore, good phonetic balance, reliability, relative symmetrical phoneme occurrence and inter-list equivalency were achieved in TU PB'14 lists, with no duplicate words among various lists [6].

In this paper, test-retest relationships of TU PB'14 are evaluated by using the five lists in test and retest sessions. Moreover, detailed analysis of listeners' errors is carried out to examine possible meaningful trends in error patterns. Section 3 presents details of the setup. Section 4 shows experimental results. Section 5 discusses the findings and future work.

3. Experimental Setup

An experiment is carried out to investigate whether there is a significant difference in subject's performance between test and retest sessions using TU PB'14. The interval between the test and retest sessions ranged from 6–28 days, with a mean interval of 14.03 days. In addition, subjects' errors and confusion patterns are examined.

Table 1: Number of occurrences of initial phonemes in TU PB'14 lists.

List	pʰr	pr	kʰw	kʰr	kl	kr	tr	p	pʰ	b	t	tʰ	d	tɕ	tɕʰ	k	kʰ	ʔ	f	s	h	m	n	ŋ	l	r	w	j
1	1			1				1	1		1	2	1	1		2	2	1	1	2	1	1	2		1	1	1	1
2			1				1	1	2		1	2	1	1	1	2	2	1		1	1	2	2		1	1		1
3		1				1			1	1	1	2	1	1	1	1	1			2	1	2	2		2	2	1	1
4	1				1			1	1	1	1	3	1	1	1	1	1			2		2	2		1	2	1	1
5		1						1	1	1	1	2	1	1	1	2	2	1		2		1	2	1	1	2	1	

Table 2: Number of occurrences of vowels, final phonemes, and lexical tones in TU PB'14 lists. Note: In isolation, short-vowel syllables with no final consonant are phonetically ended with [ʔ]; ø denotes the lack of final consonant. /ɯa ɯa: ua ua:ia ia:/ are vocalic diphthongs.

List	a	a:	i	i:	u	u:	e	e:	ɛ	ɛ:	ɤ	ɤ:	o	o:	ɔ	ɔ:	ɯ	ɯ:	ia	ia:	ua	ua:	ɯa	ɯa:	ø	ʔ	p	t	k	m	n	ŋ	w	j	ˉ	ˋ	ˆ	ˊ	ˇ
1	7	5	1	2	1	1							2		1	2				1		1			6	3	1	2	1	2	4	3		3	9	5	4	4	3
2	7	5	1	2	1	1				1			1	1		1		1		1		1		1	5	4	1	2	1	2	4	2	1	3	8	6	5	4	2
3	8	5	1	1			1	1		1		1	1	1		2		1				1		1	5	4	1	2	1	2	4	3	1	2	8	6	5	4	2
4	8	6	1	1	1		1	1	1	1		1	1			1	1								5	4	1	2	2	1	4	3	1	2	9	5	5	4	2
5	8	5	1	2	1	1	1		1	1			1			2	1								5	4	1	2	1	1	4	3	1	3	9	5	5	4	2

Table 3: TU PB'14 word lists given in IPA symbols and English glosses.

List 1	List2	List3	List4	List5
kàt bite	kàʔ estimate	kràʔ freckle	klàn distill	kām karma
kā: raven	kā:w glue	kē:m game	kéʔ small drawer	kā:n task
kʰâ: me	kʰàp drive	kʰàj egg	kʰān itch	kʰì: ride
kʰrî:p fin	kʰwā:n grope	tɕàʔ will	tɕē: vegetarian	kʰɯ̆ŋ stretch
kʰá:ŋ remain	kʰɔ̄: neck	tɕʰɤ̄:ŋ manner	tɕʰɔ̂: bouquet	ŋɔ́ʔ rambutan
tɕām remember	tɕî: rob	dɛ̄:n territory	dì:t flick	tɕǐw tiny
dɔ̄:j hill	tɕʰāŋ hate	tèʔ kick	tā: eyes	tɕʰɛ̂: chill
tān trap	dɯ̀:m drink	tʰâm cave	tʰà:ŋ open	dàk trap
tʰān in time	trīa:m prepare	tʰá: challenge	tʰɯ̀k massive	tɔ̀: renew
tʰīa:m artificial	tóʔ table	nî: these	tʰɔ̀ʔ rabbit year	tʰà:n coal
ná: aunt	tʰɛ̄:n replace	bìʔ break off	sāj banyan tree	tʰû:p joss stick
nāj in	tʰāj Thai	pràʔ dot	nā:ŋ Madame	nā: field
pòk cover	náʔ final particle	pʰăn fluctuate	nâw rotten	nā:j mister
pʰà: chop	nâŋ sit	mót ant	bɛ̀:p model	bɔ̀: well
fǔ:ŋ group	pùa:j sick	má: horse	pà: forest	prùʔ perforate
pʰrɔ́ʔ because	pʰǐ: ghost	jɔ̄:n throw	pʰǎ: cliff	pɔ̀ʔ pile up
mɯ̄: hand	pʰùʔ decay	rán stubborn	pʰráʔ monk	pʰāŋ collapse
rɔ̂:t survive	mâ:n curtain	rɯ̄:a boat	mân confident	má:j wood
lóʔ discard	mít covered	lɔ̂:k imitate	júʔ incite	ráŋ hold back
sàʔ pool	jâ:k difficult	lɔ̂:t pass through	rā:j case	rî:t press
sì: four	rō:j sprinkle	wâ:p flash	rɤ̂:m begin	léʔ messy
jǐŋ female	láʔ leave	sàŋ command	wík wig	wān day
hûn share	sɯ̂a: shirt	sà:j swing	sân short	sàt animal
wǎj be able	hǎ:n divide	nùa:ŋ delay	mòt clear	sâj stuffing
ʔūa:n seine	ʔù:t camel	hǎ:w yawn	lǎŋ behind	ʔân restrain

Table 4: Distribution of intensity test levels (in dB HL) across a set of five subjects for the test and retest sessions (T stands for TU PB'14 and numerical information represents list number).

Subject	dB HL				
	15	25	35	45	55
I	T-5	T-4	T-1	T-2	T-3
II	T-4	T-3	T-5	T-1	T-2
III	T-3	T-2	T-4	T-5	T-1
IV	T-2	T-1	T-3	T-4	T-5
V	T-1	T-5	T-2	T-3	T-4

Table 5: Average percent correct discrimination scores with 95% CI obtained with TU PB'14 during the test and retest sessions.

dB HL	Average		95% CI	
	Test	Retest	Test	Retest
15	18.1%	16.1%	6.0% –30.3%	1.1% –31.1%
25	52.4%	45.1%	43.4% –61.4%	36.0% –54.2%
35	80.0%	80.3%	75.6% –84.4%	77.5% –83.1%
45	91.2%	92.8%	86.3% –96.1%	90.6% –95.0%
55	92.3%	94.3%	86.7% –97.8%	91.5% –97.1%

To do so, 125 monosyllabic words from five TU PB'14 lists (25×5 = 125) were read three times and recorded at a sampling rate of 44.1 kHz in a sound attenuated chamber by a 39 year-old Thai male speaker who was born and grew up in Bangkok. Then, one of the three tokens of each word was selected based on impressionistic hearing evaluation and spectrographic inspection.

Each of the five lists was presented at one of five intensity levels, i.e., 15, 25, 35, 45, and 55 dB HL. These intensity levels were chosen based on our preliminary experiments such that floor and ceiling would be achieved. It should be noted that rather than performing a straightforward test of 125 stimuli × 5 intensity levels that would have created a test of 625 stimuli (considerably long and could cause subject's fatigue and learning effect [7]), we decided to increase a number of subjects five times. Consequently, the total number of words/trials for each subject stays at 125 trials and it took about one hour per subject. In all, the test sessions were equally divided by five intensity levels across six sets of five subjects as shown in Table 4.

The psychoacoustic tests were performed individually on untrained 30 normal hearing subjects consisted of 15 males and 15 females ranging in age from 19 to 23 years, with a mean of 20.5 years. They were drawn from the student

Table 6: Confusion matrix of initial consonants across test and retest sessions at 35dB HL (each cell represents row-wise based normalization (%) with raw data in parenthesis). Last column contains 3-tuple (stimulus, response, % (raw data)): (/k/, /pl/,1.0(1)), (/r/, /pl/, 2.1(2)), (/r/, /dr/, 1.0(1)), (/t/,/pl/, 1.7(1)), (/tɕ/, /kʰl/, 1.7(1)), (/ʔ/, /pl/, 2.8(1)), (/h/, /pʰl/, 2.8(1)), (/pʰr/, /pʰl/,4.2(1)), (/pr/, /pl/, 4.2(1)), and (/kʰw/, /kw/, 8.3(1)).

Stimulus	Response																												
	/pʰr/	/pr/	/kʰw/	/kʰr/	/kl/	/kr/	/tr/	/p/	/pʰ/	/b/	/t/	/tʰ/	/d/	/tɕ/	/tɕʰ/	/k/	/kʰ/	/ʔ/	/f/	/s/	/h/	/m/	/n/	/ŋ/	/l/	/r/	/w/	/j/	Others
/pʰr/	79.2 (19)	0.0 (0)	0.0 (0)	8.3 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/pr/	0.0 (0)	79.2 (19)	0.0 (0)	4.2 (1)	0.0 (0)	12.5 (3)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/kʰw/	0.0 (0)	0.0 (0)	75.0 (9)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	16.7 (2)	0.0 (0)	0.0 (0)
/kʰr/	0.0 (0)	0.0 (0)	0.0 (0)	100.0 (12)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)
/kl/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	83.3 (10)	0.0 (0)	0.0 (0)	16.7 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/kr/	0.0 (0)	16.7 (2)	0.0 (0)	0.0 (0)	0.0 (0)	75.0 (9)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	8.3 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/tr/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	16.7 (2)	58.3 (7)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	25.0 (3)	0.0 (0)	0.0 (0)	3.1 (3)
/p/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	91.7 (44)	0.0 (0)	2.1 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.1 (1)	4.2 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/pʰ/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	12.5 (9)	73.6 (53)	1.4 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.4 (1)	1.4 (1)	8.3 (6)	0.0 (0)	0.0 (0)	0.0 (0)	1.4 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/b/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	5.6 (2)	0.0 (0)	88.9 (32)	2.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/t/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	5.0 (3)	0.0 (0)	0.0 (0)	93.3 (56)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.7 (1)
/tʰ/	0.0 (0)	0.0 (0)	0.0 (0)	0.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	9.1 (12)	0.0 (0)	0.0 (0)	68.2 (90)	0.8 (1)	0.8 (1)	0.8 (1)	0.0 (0)	14.4 (19)	0.0 (0)	0.0 (0)	3.0 (4)	1.5 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.8 (1)	1.7 (1)
/d/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	3.3 (2)	0.0 (0)	88.3 (53)	0.0 (0)	0.0 (0)	6.7 (4)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.7 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/tɕ/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.7 (1)	0.0 (0)	0.0 (0)	91.7 (55)	0.0 (0)	3.3 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.7 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/tɕʰ/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	6.3 (3)	0.0 (0)	0.0 (0)	93.8 (45)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/k/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	6.3 (6)	0.0 (0)	1.0 (1)	3.1 (3)	0.0 (0)	83.3 (80)	2.1 (2)	0.0 (0)	0.0 (0)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/kʰ/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (4)	0.0 (0)	0.0 (0)	5.2 (5)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	88.5 (85)	0.0 (0)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)
/ʔ/	2.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)	5.6 (2)	5.6 (2)	0.0 (0)	5.6 (2)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)	0.0 (0)	72.2 (26)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)
/f/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	91.7 (11)	8.3 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/s/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.9 (1)	0.0 (0)	0.0 (0)	1.9 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.9 (1)	96.3 (104)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (1)
/h/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)	13.9 (5)	0.0 (0)	0.0 (0)	2.8 (1)	0.0 (0)	13.9 (5)	0.0 (0)	2.8 (1)	0.0 (0)	61.1 (22)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (1)
/m/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	2.1 (2)	6.3 (6)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	89.6 (86)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	8.3 (1)
/n/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.8 (1)	95.8 (115)	1.7 (2)	0.8 (1)	0.8 (1)	0.0 (0)	0.0 (0)	0.0 (0)
/ŋ/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	8.3 (1)	91.7 (11)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/l/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	2.8 (2)	1.4 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	90.3 (65)	5.6 (4)	0.0 (0)	0.0 (0)	0.0 (0)
/r/	1.0 (1)	0.0 (0)	0.0 (0)	3.1 (3)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	3.1 (3)	0.0 (0)	2.1 (2)	0.0 (0)	4.2 (4)	0.0 (0)	0.0 (0)	2.1 (2)	1.0 (1)	0.0 (0)	0.0 (0)	5.2 (5)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (1)	0.0 (0)	72.9 (70)	0.0 (0)	0.0 (0)	0.0 (0)
/w/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.2 (2)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	95.8 (46)	0.0 (0)	0.0 (0)
/j/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	100.0 (48)	0.0 (0)
Others	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)

Table 7: Confusion matrix of final consonants across test and retest sessions at 35dB HL (each cell represents row-wise based normalization (%) with raw data in parenthesis).

Stimulus	Response										
	/ø/	/ʔ/	/p/	/t/	/k/	/m/	/n/	/ŋ/	/w/	/j/	Others
/ø/	94.4 (272)	0.0 (0)	1.7 (5)	0.3 (1)	1.7 (5)	0.0 (0)	0.3 (1)	1.4 (4)	0.0 (0)	0.0 (0)	0.0 (0)
/ʔ/	0.0 (0)	86.9 (219)	0.8 (2)	4.8 (12)	7.5 (19)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/p/	18.3 (11)	0.0 (0)	68.3 (41)	3.3 (2)	1.7 (1)	6.7 (4)	0.0 (0)	1.7 (1)	0.0 (0)	0.0 (0)	0.0 (0)
/t/	0.0 (0)	0.0 (0)	3.3 (4)	92.5 (111)	1.7 (2)	0.0 (0)	2.5 (3)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/k/	1.4 (1)	0.0 (0)	2.8 (2)	6.9 (5)	86.1 (62)	1.4 (1)	0.0 (0)	1.4 (1)	0.0 (0)	0.0 (0)	0.0 (0)
/m/	2.1 (2)	0.0 (0)	5.2 (5)	0.0 (0)	0.0 (0)	88.5 (85)	4.2 (4)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
/n/	1.3 (3)	0.0 (0)	0.0 (0)	4.2 (10)	0.0 (0)	0.0 (0)	88.8 (213)	4.2 (10)	0.0 (0)	1.7 (4)	0.0 (0)
/ŋ/	1.8 (3)	0.0 (0)	0.0 (0)	0.6 (1)	2.4 (4)	0.0 (0)	2.4 (4)	92.9 (156)	0.0 (0)	0.0 (0)	0.0 (0)
/w/	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	100.0 (48)	0.0 (0)	0.0 (0)
/j/	0.6 (1)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	99.4 (155)	0.0 (0)
Others	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)

population at Thammasat University, Thailand. All the subjects passed a screening test to pure tones from 125 through 8,000 Hz at 20 dB HL in both ears and the right ear was served as the test ear.

Each subject sat in a sound attenuated chamber in an ENT clinic, Thammasat University Hospital and listened to playback speech stimuli (explained earlier) via a pair of headphones. He/she had to repeat the word they heard. If they did not recognize the word, they had to make a guess before moving to the next one. The test is divided into one training (10 stimuli) session and five test sessions (25 stimuli each). A short five-minute break was given to the subject at the end of each session.

During the test and retest sessions, the same procedure was carried out and identical distribution of intensity test levels across all subjects were maintained.

4. Experimental Results

Table 5 shows average percent correct discrimination scores with 95% CI between the test and retest sessions at all five intensity levels. Paired t-test confirms no significant difference of the scores between test and retest at every intensity level, ([$t(5) = 0.9947$, $p = 0.3655$] at 15 dB HL, [$t(5) = 0.8924$, $p = 0.4131$] at 25 dB HL, [$t(5) = -0.2840$, $p = 0.7878$] at 35 dB HL, [$t(5) = -1.0260$, $p = 0.3520$] at 45 dB HL, and [$t(5) = -1.6194$, $p = 0.1663$]) at 55 dB HL.

The subjects performed poorly at 15 dB HL (floor), about chance (about 50%) at 25 dB HL, and very well at 45 and 55 dB HL (ceiling). It should be noted that the most interpretable intensity level is at 35 dB HL, where the subjects had average percent correct discrimination scores of about 80.0% and 80.3% for the test and retest, respectively. Therefore, at this intensity level, subjects' errors are analyzed and confusion matrices (for initial and final consonants, vowel, and tones) are constructed. For this paper, only confusion matrices for initial (Table 6) and final consonants (Table 7) are shown as most errors were associated with those sounds. It is noteworthy that since no significant difference is found between the scores of test and retest, the data presented in Tables 6 and 7 is derived from the sum of the two test sessions at 35 dB HL.

Among 1,500 stimuli (25×5×6×2) presented at 35 dB HL, 298 words were responded incorrectly, which can be divided into three categories of error: one-sound, two-sound, and three-sound errors. One-sound error constitutes 67.81%, of which 45.64% was initial only, 0.36% was vowel only, 19.13% was final only, and 2.68% was tone only. Two-sound error constitutes 27.51%, of which 0% was initial + vowel, 18.12% was initial + final, 5.03% was initial+ tone, 0% was vowel+ final, 0% was vowel + tone, and 4.36% was final + tone. Finally, three-sound position error constitutes 4.72%, of which 1.34% was initial + vowel + final, 0% was initial + vowel + tone, 0.36% was vowel + final + tone, and 3.02% was initial + final + tone. No four-sound (all-sound) error ever occurred.

Table 6 shows confusion patterns of initial phonemes. Out of 21 single consonants, the four least confusable phonemes were /j/, /s/, /n/, and /w/ and the most confusable were /h/, /tʰ/, and /r/. As for consonant clusters, /kʰr/ was the least confusable and /tr/ the most confusable.

Table 7 shows confusion patterns of final phonemes. Out of nine consonants, the least confusable phonemes were /w/ and /j/ and the most confusable were /p/ and /k/.

Preliminary analysis shows that when a sound was misperceived, it is more likely that the misperception was associated with place of articulation rather than voicing contrast. For example, /tʰ/ was mostly misperceived as /kʰ/ and /pʰ/. The phonemes share voicing feature (voiceless aspirated), but differ in terms of place of articulation.

5. Discussions and Future Work

Our goal is to create a new Thai word recognition test with good phonetic balance, symmetrical phoneme occurrence and relative inter-list equivalency, and test-retest reliability. TU PB'14 lists are now successfully developed and evaluated in an experimental setting. When any of the 5 lists is used a second time after a period of more than 14 days, the discrimination score is highly reliable. To further examine variability of discrimination scores, we plan to carry out similar tests on sensorineural and conductive hearing impaired subjects, which are less homogeneous groups.

Detailed analysis of listeners' errors reveals that errors occurred predominantly in the case of initial only, final only, and initial along with final consonants. It suggests that consonants are more susceptible than vowels and tones as intensity level changes.

Largely, confusion patterns (initials and finals) reported here are in line with those found in noise condition [8]. In fact, the 'shared' least confusable phonemes in initial position are /j/ and /w/ and the 'shared' most confusable /tʰ/ and /r/. Likewise, in final position, the 'shared' least confusable phonemes are /w/ and /j/ and the 'shared' most confusable /k/. Interestingly, voicing was also found to be the most robust contrast while place-of-articulation was the least [8].

6. References

[1] A. Munthuli, P. Sirimujalin, C. Tantibundhit, K. Kosawat, and C. Onsuwan, "A corpus-based study of phoneme distribution in Thai.," in *10th International Symposium on Natural Language Processing*, Phuket, TH, 2013, pp. 114–121.

[2] K. Kosawat, M. Boriboon, P. Chootrakool, A. Chotimongkol, S. Klaithin, S. Kongyoung, K. Kriengket, S. Phaholphinyo, S. Purodakananda, T. Thanakulwarapas, and C. Wutiwiwatchai, "BEST 2009: Thai word segmentation software contest," in *8th International Symposium on Natural Language Processing*, Bangkok, TH, 2009, pp. 83–88.

[3] G. Lidén and G. Fant, "Swedish word material for speech audiometry and articulation tests," *Acta Oto-Laryngol*, vol. 116, pp. 189–210, 1954.

[4] A. Hammer, B. Vaerengerg, W. Kowalczyk, L. T. Bosch, and M. Coene, "Balancing word lists in speech audiometry through large spoken language," in *14th Annual Conference of the International Speech Communication Association (Interspeech)*, Lyon, 2013, pp. 3613–3616.

[5] A. Thangthai, C. Hansakunbuntheung, R. Siricharoenchai, and C. Wutiwiwatchai, "Automatic syllable-pattern induction in statistical Thai text-to-phone transcription," in *9th International Conference on Spoken Language Processing*, Pittsburgh, 2006.

[6] A. Munthuli, P. Sirimujalin, C. Tantibundhit, C. Onsuwan, K. Kosawat, and N. Klangpornkun, "Constructing Thai phonetically balanced word recognition test in speech audiometry through large written corpora," in *17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (Oriental COCOSDA)*, Phuket, TH, 2014.

[7] P. C. Loizou, *Speech Ehancement: Theory and Practice*. New York: CRC Press, Taylor & Francis Group, 2007.

[8] C. Tantibundhit, C. Onsuwan, S. Thatphithakkul, P. Chootrakool, K. Kosawat, N. Thatphithakkul, T. Saimai, and Saimai, "Subjective intelligibility testing and perceptual study of Thai initial and final consonants," in *17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, China, 2011, pp. 1970–1973.