Methods and Tool for Constructing Phonetically-Balanced Materials for Speech Perception Testing: A Development of Thai Sentence-Length Materials

Adirek Munthuli
Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Thailand
5310450027@student.tu.ac.th

Charturong Tantibundhit
Department of Electrical and Computer Engineering, Faculty of Engineering and Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation (CILS)
Thammasat University, Thailand
tchartur@engr.tu.ac.th

Chutamanee Onsuwan
Department of Linguistics, Faculty of Liberal Arts and Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation (CILS)
Thammasat University, Thailand
consuwan@tu.ac.th

Krit Kosawat
National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Thailand
krit.kosawat@nectec.or.th

Abstract

Phonemic content is one of many important criteria in a development of any kind of speech testing materials. In this paper, we explain a procedure and tool we created in the process of constructing phonetically-balanced (PB) sentence-length materials for Thai, as an assessment for speech reception thresholds. Our procedure includes establishing criteria, preselecting sentences, creating pool of replacement words, determining phonemic distribution, and constructing sentences. Importantly, a tool is created to determine whether set of words or sentences are phonetically balanced. Once the phoneme distribution and the set of words with transcription are specified, the tool efficiently computes phoneme occurrences among words or sentences (within a set) and can be used to manipulate words to achieve goal in phonetically balanced (PB). To show how this is accomplished, two sentence sets are constructed and evaluated by native speakers. The procedure and tool have characteristics that make them potentially useful in other applications and can be applied to other languages.

Keywords: Thai, sentence-length material construction, phonetically balanced speech materials

1 Introduction

It is well-established that an assessment technique for evaluating an individual’s hearing sensitivity based on pure-tone audiometry alone does not truly reflect the individual's speech understanding (Bilger et al., 1984; Egan, 1948). Importantly, measuring of speech intelligibility could be obtained by counting number of correct responses from speech testing materials, e.g., phonetically-balanced (PB) monosyllabic words, polysyllabic words, and sentences (Egan, 1948).

29th Pacific Asia Conference on Language, Information and Computation: Posters, pages 293 - 301
Shanghai, China, October 30 - November 1, 2015

For the Thai language, there are a few existing speech materials for intelligibility test, some of these were developed using monosyllabic word lists, e.g., RAMA.SD1, RAMA.SD2 (Komalarajun, 1979), and TU PB'14 (Munthuli et al., 2014) while some using phrase or sentence materials, e.g., Ramathibodi Synthetic Sentence Identification (RAMA.SSI), which contains Thai artificial sentences (with no real meaning) (Wissawapaisal, 2002), and "PB and PD sentences", which are long stretches of phrases and sentences derived from Thai continuous speech corpus for an evaluation of automatic speech recognition system (Wutiwiwatchai et al., 2002). However, a majority of speech testing methods using sentence materials requires that the sentences are representative of the real communication system, which includes many factors such as meaning, context, rhythm, etc. (Egan, 1948). It is quite clear that the existing Thai sentence materials would not satisfactorily meet this requirement.

Sentence speech materials have been created in many languages, e.g., Dutch (Plomp and Mimpen, 1979), Mandarin Chinese (Fu et al., 2011), German (Kollmeier and Wesselkamp, 1997). For English (American), the most widely used are Speech Perception in Noise test (SPIN) (Kalikow et al., 1977) and Hearing in Noise Test (HINT) (Nilsson et al., 1994). SPIN is a test for measuring speech intelligibility at fixed S/N ratio, but it was found to have variability in terms of sentence difficulty (Kalikow et al., 1977). Therefore, HINT was designed and developed as a Hearing in Noise Test, composed of lists of sentences, which are shown to have no significant difference in terms of difficulty. Among those, different strategies were used (but no specific tool had been mentioned) to construct the phonetically balanced materials. It should be noted that those materials were created to obtain similar phoneme distributions among sentence sets (see Phonemic content in Table 1) rather than to reflect the true phonemic distributions of the language. Our approach tries to achieve both ends by using a semi-automatic tool. A fully automated tool of this type would be ideal, but would require other crucial components such as a language model. A list of important characteristics of SPIN and HINT are given in Table 1.

Due to the lack of Thai 'natural' sentence materials for speech perception testing, and especially those for assessing hearing-impaired individuals. Our goal is to construct Thai phonetically balanced sentence-length materials for assessing speech reception thresholds. In this paper, we describe the methods and tool for constructing a subset of these meaningful sentences.

	SPIN (1977)	HINT (1994)
Sentence length	6-8 syllables	6-9 syllables
Number of lists	8	25
Number of sentences per list	50	10
Measurement	Speech intelligibility (count only 'keyword' at the last monosyllabic noun of the sentence)	Speech intelligibility (count every word of the sentence) and sentence speech reception threshold (sSRT)
Phonemic content	Balanced within class of phonemes from Dewey's written corpus	Phonemically balanced of 43 phoneme sounds among lists
Others	Low predictability (LP) and high predictability (HP) sentences	Sentence difficulty: 1-grade reading level

Table 1: Important characteristics of SPIN and HINT tests.

2 Establishing Criteria

The first requirement in constructing PB sentence materials is phonetic/phonemic balance. Other common criteria include word familiarity, naturalness, sentence length, homogeneity, test-retest reliability, and inter-list difficulty (Bilger et al., 1984). Our approach is to incorporate most of the above criteria. However, in this paper, our focus is on the initial phase, which is designing lists of natural sentences with phonetic balance, equal length, and familiar words. The next phase, testing and evaluating, not discussed here, will be to ensure homogeneity, test-retest reliability, and inter-list difficulty.

Our PB sentence lists are based on phoneme distribution of Thai speech LOTUS-CELL2.0 (LT-CS) corpus (Section 3.2). To minimize effect of subject's different language background, we opt for familiar words. This is carried out by selecting words and sentences, which match desired phonemic content, from children's textbooks and stories, (Thai Children Stories, 1990; Ministry of Education, 1986; Sripaiwan, 1994; Sangworasin, 2003). In terms of sentence length, we follow SPIN (Kalikow et al., 1977), HINT (Nilsson et al. 1994), and RAMA.SSI (Wissawapaisal, 2002) and limit each sentence to six to eight syllables with no words greater than two syllables long. The PB sentences will compose of 10 lists, each with 10 sentences.

In addition, to address a question of whether different levels of predictability affect sentence intelligibility (Kalikow et al., 1977), in all five lists will be created to fit the 'low' predictability status and another five the 'high' predictability. However, degrees of predictability are beyond the scope of our developed tool, and are determined by semantics and overall sentence contexts. (see Sections 4 and 5).

3 Procedure and Concept Design for Tool

For SPIN (Kalikow et al., 1977) and HINT (Nilsson et al., 1994), pre-selection of sentences were carried out prior to phonemic distribution analysis and matching. HINT sentences were selected from Bamford-Kowal-Bench (BKB) corpus. SPIN sentences were constructed by generating sets of 'low' predictability and 'high' predictability sentences and manipulated key words (monosyllable nouns) in sentence final position by determining their semantics link to preceding words in the sentence. The key words were drawn from Thorndlike-Lorge corpus. Consequently, for HINT, there is 68% (of 252 sentences) where phonemes are off ± 1 from the target phonemes (Nilsson et al., 1994).

We found their approach quite difficult to achieve for Thai sentences as there are 4 phoneme types (initials, vowel, finals, and lexical tones) to account for. Therefore, we have taken a slightly different approach by starting with pre-selection of sentences in the same fashion, but the sentences will be further modified by replacing and reconstructing some words in sentences until it yields desired phonemic contents as described in Section 3.1.

Kalikow et al. (1977) asserted that recognition of keywords in sentence is based on familiarity of word. Therefore, for our lists, we have to make certain that the selected words (candidates) are familiar words in the language. We do so, by selecting words from children's textbooks and stories. In addition, to estimate frequency of word occurrences, we utilize the largest available Thai written corpus InterBEST (Kosawat et al., 2009).

Most importantly, our PB word candidates are considered to be as phonetic balanced as possible, i.e., less than 10% difference from targeted phoneme distribution.

3.1 Preselecting Sentences and Pool of Replacement Words

The first step to create PB sentences is based on preselection of sentences. All sentences from a collection of 89 children's stories (Thai Children Stories, 1990) are analyzed and only simple sentences, (i.e., subject-verb-(adverb), subject-verb-complement/object), are kept. These result in 313 sentences in total. Then, each sentence is transcribed and its phonemic distribution of initials, finals, vowels, and tones are tallied. Attempts are made to group a set of 10 sentences in to a list (10 lists in all) such that the phoneme distributions are as close to the ones shown in Tables 2-5 as possible.

However, from a limited number of simple sentences (313 sentences) that were preselected, the best outcome we could obtain was 10 lists of useable PB sentences with very low off-target from the desired phoneme distributions.

Therefore, we propose an additional step, which is to modify our preselected sentences by replacing and reconstructing some words using a pool of replacement words so that it finally yields ten mutually exclusive groups of 10 sentences that match the desired phoneme distributions.

Our pool of replacement words came from the collection of 89 children's stories (Thai Children Stories, 1990) and word corpora based on three children's textbooks (Ministry of Education, 1986; Sripaiwan, 1994; Sangworasin 2003).

3.2 Phonemic Content

In this section, phoneme frequency occurrence and its distribution (ranking) derived from written and spoken Thai corpora (Kosawat et al., 2009;

Chotimongkol et al., 2009; Chotimongkol et al., 2010) are discussed (Munthuli et al., 2015). More generally, InterBEST, which is one of the Thai largest written corpora, is composed of 12 text genres with approximately nine million words. LOTUS-CELL2.0 is a collection of telephone conversation recordings of 50 hours long, where data were transcribed according to different speaking styles: formal style (LT-FS) and causal speech style (LT-CS) (Chotimongkol et al., 2010). LOTUS-BN is a Thai television broadcast news recordings of 100 hours long. Munthuli et al. (2015) show phoneme distribution from InterBEST, LOTUS-CELL2.0 (LT-FS and LT-CS) and LOTUS-BN. Among the written and two spoken corpora, there are notable differences (largely due to lexical differences and phonetic variations in conversational speech) in terms of frequency occurrence and the distribution for initial consonants, vowels, final consonants (but not for lexical tones) (Munthuli et al., 2015). In addition, many existing speech testing materials (e.g., HINT) favored the use of spoken corpus (Nilsson et al., 1994). Therefore, our approach here is to employ the phoneme frequency occurrence and distribution derived from causal speech style (LT-CS). The next step is to modify our preselected sentences by replacing and reconstructing some words (using the pool of replacement words in Section 3.1) so that it finally yields ten mutually exclusive groups of 10 sentences that match the desired phoneme distributions as shown in Tables 2-5 as much as possible.

3.3 Selecting and Replacing Words

A tool is developed to facilitate the process at which the preselected sentences are modified by replacing and reconstructing some words using a pool of replacement words so that it finally yields ten mutually exclusive groups of 10 sentences that match the desired phoneme distributions. The steps involved are as follows:

1. Consider target number of phoneme occurrence of all 65 phonemes shown in Tables 2-5 (29 initials, 21 vowels, 10 finals, and 5 tones) that are required for construction of PB sentences.
2. Start with construction of PB sentences of List 1. Consider all combinations of the preselected sentences; choose 10 sentences ($^{313}C_{10}$). Then, the selected 10 sentences will be transcribed and the resulting phonemes are tallied.
3. For each case of the selected 10 sentences, calculate absolute difference between numbers of occurrences of Step 1 and Step 2 for each phoneme. Then, calculate percentage of summation of absolute differences for all phonemes.
4. Select the best combination of 10 simple sentences (6-8 syllables per sentence), where the sentences provide the lowest percentage of summation of absolute differences for all phonemes. After this selection, these 10 sentences will be removed from the list of preselected sentences.
5. Consider the phonemes in Step 4, where numbers of occurrences are higher than target numbers in Step 1. These phonemes will be among the first phonemes to be removed.
6. Consider all words from the selected sentences in Step 4, which compose of phonemes (initials, finals, vowels, or tones) in Step 5. These words will be removed in order based on which one has a higher number of exceeding phonemes per syllable. In case of tie, the one with lower word frequency of occurrences based on InterBEST corpus (Kosawat et al., 2009) has higher priority to be removed. It should be noted that a two-syllabic word has a higher priority than a monosyllabic word. Then, update number of occurrences of all phonemes.
7. Repeat Step 6 until no exceeding phoneme available.
8. Now, all phonemes have numbers of occurrences below target numbers. Then, insert a new word from a pool of replacement words. Words with higher frequency of occurrences will have higher priority. Then, update number of occurrences of all phonemes.
9. Repeat Step 8 until numbers of occurrences of all phonemes of preselected sentences have absolute error less than 10%, i.e., any phoneme in any group of initials, finals, vowels, and tones can be

out of target at most 2 times and 6 times in total.

10. Use words in the replacement pool to construct 10 sentences, where each sentence is a simple sentence composed of seven syllables.

Repeat Steps 1 to 10 to construct PB sentences for Lists 2 to 5.

4 Tool

Graphical user interface is developed to facilitate insertion or removal of words by considering each list of PB sentences one by one.

Figures 1-2 show asterisks on the chart. Each of which signifies a target number of occurrences of any phoneme. Bar refers to current number of occurrences of any phoneme, where positive/negative number signifies that number of occurrences is higher/lower than a target number of occurrences of that phoneme.

Figure 1 shows an example of the PB sentence constructing process with the tool starting with Step 1 described earlier in Section 3.3. Here, we show construction of sentences in list I (low predictability sentences). After Steps 2 to 3 are performed, select the best combination of 10 sentences stated in Step 4. Then, each sentence (one by one) is put in the tool as shown in Fig. 1.

After Steps 5 to 7 are performed, insert words from pool of replacement words. Figures 1-2 show words ranked in ascending order based on frequency of occurrences (from InterBEST corpus).

Figure 2 shows phoneme distributions after performing Step 9. Then, Step 10 is performed and 10 PB sentences with low predictability are shown in Table 6. As another example, we use the tool to construct 10 PB sentences with high predictability as shown in Table 7.

It should be noted that degrees of predictability are beyond the scope of this tool, and are determined by semantics and overall sentence contexts. At this stage, the tool users are expected to make several attempts in word selecting and replacing to achieve desired level of predictability, which could be later evaluated (see Section 5).

	b	tɕ	tɕ^{h}	d	f	h	j	k	k^{h}	k^{h}r	k^{h}w	kl	kr	kw	l	m	n	ŋ	p	p^{h}	p^{h}l	p^{h}r	pr	r	s	t	t^{h}	w	ʔ
List 1	2	2	2	3	0	2	3	5	6	0	0	1	0	0	7	5	5	1	3	3	0	0	0	1	5	2	4	3	5
List 2	2	3	2	3	0	2	2	5	7	0	0	0	0	0	8	5	4	1	3	3	0	1	0	0	5	3	4	2	5
List 3	2	3	2	2	1	2	2	5	7	0	0	0	0	0	7	6	4	1	3	3	0	0	1	0	5	2	5	2	5
List 4	2	2	2	3	0	2	3	5	6	0	0	0	1	0	7	5	5	1	3	3	0	0	0	1	5	2	4	3	5
List 5	2	3	1	3	1	2	2	5	6	0	0	0	0	1	7	5	5	0	3	3	0	0	0	1	5	3	4	2	6
List 6	2	2	2	3	0	2	2	6	6	1	0	0	0	0	7	5	5	1	3	3	0	0	0	1	5	2	4	3	5
List 7	2	2	2	3	1	1	3	5	6	0	1	0	0	0	7	5	5	1	3	3	0	0	0	1	5	2	4	3	5
List 8	1	3	2	3	0	2	2	6	6	0	0	0	0	1	7	5	4	1	4	2	0	0	0	1	5	3	4	2	6
List 9	2	2	2	3	1	1	3	5	6	0	1	0	0	0	7	5	5	1	3	3	0	0	0	1	5	2	4	3	5
List 10	1	3	2	3	0	2	2	6	6	0	0	0	0	0	8	5	4	1	3	3	1	0	0	1	4	3	4	2	6

Table 2: Initial consonant occurrences across ten sentence lists.

	ɔ	ɔɔ	a	aː	e	eː	i	iː	iaː	o	oː	ɤ	ɤː	u	uː	uaː	ɯ	ɯː	ɯaː	ɛ	ɛː
List 1	1	7	22	11	2	2	2	5	1	3	1	0	1	2	2	1	1	2	1	0	3
List 2	1	7	21	12	2	1	3	4	2	2	1	0	2	2	1	1	1	2	1	1	3
List 3	1	7	21	12	2	2	2	5	1	2	1	0	2	2	2	1	1	2	1	0	3
List 4	1	6	22	12	2	1	3	4	2	2	1	0	2	1	2	1	1	2	1	1	3
List 5	1	7	21	12	2	1	3	4	2	2	1	0	2	2	1	2	1	1	1	1	3
List 6	0	7	22	12	2	1	2	5	1	3	1	0	2	1	2	1	1	2	1	0	4
List 7	1	7	21	12	2	1	3	4	2	2	1	1	2	1	2	1	1	2	1	0	3
List 8	0	7	22	12	1	2	2	5	1	3	1	0	2	2	1	2	1	1	1	1	3
List 9	1	7	21	12	2	1	3	4	2	2	1	1	1	2	2	1	1	2	1	0	3
List 10	0	7	22	12	1	2	2	5	1	3	1	0	2	2	1	1	2	1	1	1	3

Table 3: Vowel occurrences across ten sentence lists.

	j	k	m	n	ŋ	p	t	w	‘x’	ø
List 1	9	3	3	10	6	3	4	4	9	19
List 2	9	3	3	9	6	4	3	4	10	19
List 3	9	3	3	9	6	4	3	4	10	19
List 4	8	3	4	10	6	3	4	4	9	19
List 5	9	3	3	9	7	3	4	3	10	19
List 6	8	3	4	9	6	4	3	4	9	20
List 7	9	3	3	9	7	3	4	3	10	19
List 8	8	3	4	9	6	4	3	4	9	20
List 9	9	3	3	9	6	4	4	3	10	19
List 10	8	3	3	10	6	3	4	4	9	20

Table 4: Final consonant occurrences across ten sentence lists.

	¯	`	ˆ	´	ˇ
List 1	23	14	16	11	6
List 2	22	15	16	11	6
List 3	23	14	16	11	6
List 4	22	15	16	10	7
List 5	23	15	15	11	6
List 6	23	14	16	11	6
List 7	23	15	15	11	6
List 8	22	15	16	11	6
List 9	23	9	16	11	6
List 10	22	10	16	11	6

Table 5: Lexical tone occurrences across ten sentence lists.

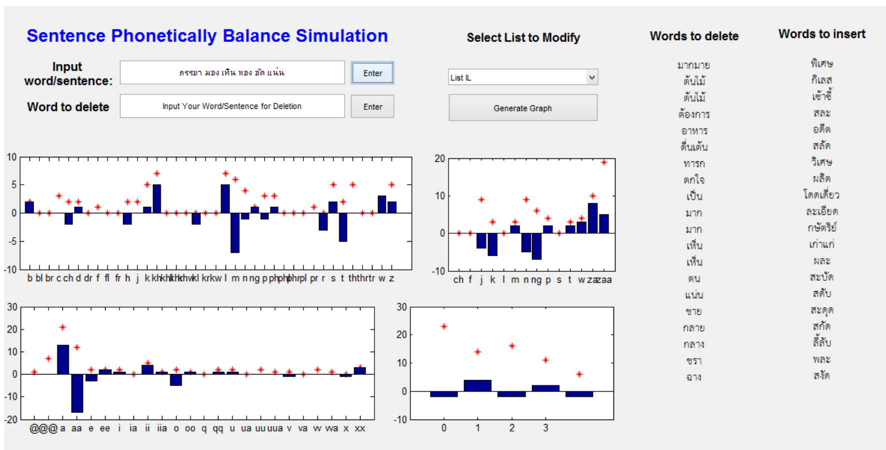

Figure 1: Simulation of tool in process of subtraction in a set of preselected sentences that has minimum absolute of summation errors. It should be noted that c is ɕ, ch is ɕ^{h}, kh is k^{h}, khl is k^{h}l, khr is k^{h}r, khw is k^{h}w, ng is ŋ, ph is p^{h}, phl is p^{h}l, phr is p^{h}r, th is t^{h}, thl is t^{h}l, za is ‘x’, zaa is ø, @ is ɔ, @@ is ɔː, aa is aː, ee is eː, ii is iː, iia is iaː, oo is oː, q is ɤ, qq is ɤː, uu is uː, uua is uaː, v is ɯ, vv is ɯː, vva is ɯaː, x is ɛ, xx is ɛː, 0 is mid tone, 1 is low tone, 2 is falling tone, 3 is high tone and 4 is rising tone.

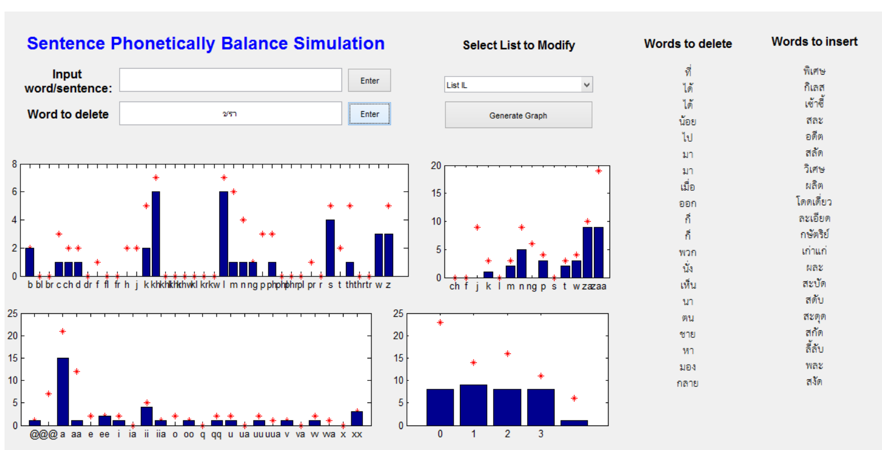

Figure 2: Simulation of tool in process of removing words in a set of sentences which has exceeding phonemes.

5 Preliminary Output and Evaluation

Tables 6 and 7 show two lists of Thai PB sentences that are successfully constructed using the tool. Importantly, differences from the target phoneme distributions are lower than 10% for each type of phoneme (initial consonant, vowel, final consonant, and lexical tone).

Another important step, which we incorporate into our procedure, is to analyze and evaluate our attempts in word selecting to achieve desired level of predictability. In so doing, we statistically compare evaluation responses from Thai raters and determine whether they rate 'low' and 'high' predictability sentences differently

We combine 20 sentences (constructed 'high' (Type 4) and 'low' (Type 2) predictability sentences) listed in Tables 6 and 7 with 20 sentences drawn from the list of our preselected sentences (Section 3.1). Ten of the twenty preselected sentences could potentially be considered as highly predictable (Type 3) and the other 10 with low predictability (Type 1). Twenty Thai adult participants are asked to rate each sentence in five-point scales (5 = very high predictability, 1 = very low predictability). Mean and average rating score of four types of sentences are given in Table 8.

หญิงสาว แบเบาะ ก็ สร้าง ชาติ 'young lady' 'baby' 'also' 'build' 'nation' A baby lady also builds a nation. [jǐŋ sǎ:w bɛ̄:øbɔ̀x kɔ̂:ø sâ:ŋ tɕʰâ:t]
เป็ด ดื้อ กลับ หา กำไร ได้ 'duck' 'stubborn' 'become to' 'find' 'profit' 'get' A stubborn duck is making a profit. [pèt d ɯ̂:ø klàp hǎ:ø kāmrāj dâ:j]
พวกตน เหวี่ยง ลิ้น ออก ไป 'we' 'fling' 'tongue' 'out' 'go' We fling the tongue out. [pʰûa:k tōn wìa:ŋ lín ʔɔ̀:k pāj]
ภรรยา มองเห็น คอ ขยับ 'wife' 'see' 'neck' 'move' A wife sees the neck moving. [pʰānráx jā:ø mō:ŋ hěn kʰō:ø kʰà x jàp]
ชาย ใจดำ เล้าโลม งู 'man' 'black-hearted' 'fondle' 'snake' Black-hearted man fondles a snake. [tɕʰā:j tɕājdām láwlō:m ŋū:ø]
แพะ สามารถ ทะเลาะ กับ เวลา 'goat' 'can' 'quarrel' 'with' 'time' A goat can quarrel with time. [pʰɛ́x sǎ:ømâ:t tʰáx lɔ́x kàp wē:ølā:ø]
ท่าน ลอย ไป แก้แค้น มา 'you' 'float' 'go' 'revenge' 'come' You float to get revenge. [tʰân lō:j pāj kɛ̂:kʰɛ́:n mā:ø]
เจ้าของ ก็ นั่ง เซ้าซี้ อีก 'owner' 'also' 'sit' 'importune' 'again' An owner sits and importunes again. [tɕâwkʰɔ̌:ŋ kɔ̂:ø nâŋ sáwsí:ø ʔì:k]
หนู มี เวทมนตร์ ใน ขณะ นี้ 'mouse' 'has' 'magic' 'in' 'while' 'this' A mouse is currently having magic power. [nǔ:ø mī:ø wê:tmōn nāj kʰàx nàx ní:ø]
ต้นไม้ ทอง ขึ้น อยู่ ที่อื่น ละ 'tree' 'gold' 'grow' 'at' 'elsewhere' 'already' A golden tree already grew up elsewhere. [tônmá:j tʰō:ŋ kʰɯ̂n jù:ø tʰî:ø ʔɯ̀:n lá x]

Table 6: Example of a set of 'low' predictability sentences (constructed sentences) ('x' signifies an ending of any short-vowel syllables with no final consonant whereas 'ø' a syllable with long vowel with no final consonant).

ข้าน้อย พับ เสื้อผ้า รอ เป็น วัน 'I' 'fold' 'cloth' 'wait' 'is' 'day' I folded clothes for a day while I am waiting. [kʰâ:ø nɔ́:j pʰáp sɯ̂a:øpʰâ:ø rō:ø pēn wān]
ยาย เล่า วิธี แกะสลัก 'grandmother' 'describe' 'method' 'carving' Grandmother describes how to carve. [jā:j lâw wíx tʰī:ø kɛ̀x sàx làk]
ท่าน หิว เพิ่ม ขึ้น ไป อีก 'you' 'hungry' 'increase' 'up' 'go' 'more' You get hungrier. [tʰân hǐw pʰɤ̂:m kʰɯ̂n pāj ʔì:k]
เขา ต้อง ขอบคุณ อาจารย์ มาก 'He' 'must' 'thankful' 'professor' 'many' He must be very thankful to professor. [kʰǎw tɔ̂ŋ kʰɔ̀:pkʰūn ʔā:øtɕā:n mâ:k]
เจ้า โอ้อวด แม้ ยัง สงสัย 'you' 'show off' 'even' 'still' 'doubt' You are showing off even if you still have a doubt. [tɕâw ʔô:øʔùa:t mɛ́:ø jāŋ sǒŋsǎj]
ขณะ นี้ น้อง ไม่ เฮฮา 'while' 'this' 'brother/sister' 'not' 'joyful' Brother/Sister is not joyful at this moment. [kʰàx nàx ní:ø nɔ́:ŋ mâj hē:øhā:ø]
เธอ น้อยใจ ก็ ทะเลาะ อีก

‘she’ ‘feel slight’ ‘then’ ‘quarrel’ ‘again’ If she feels slighted, quarrel will begin again. [tʰɤ̄:ø nɔ́:jtɕāj kɔ̂:ø tʰá x lɔ́ x ʔì:k]
บัณฑิต แต่ละ คน มี ชื่อเสียง ‘graduate’ ‘each’ ‘person’ ‘has’ ‘famous’ Each Graduate is famous. [bāndìt tὲ:ø lá x kʰōn mī:ø tɕʰɯ̂:ø sǐa:ŋ]
บุตรหลาน ดูแล ไม่ ง่าย ‘children’ ‘take care’ ‘not’ ‘easy’ [Taking care of children is not easy. bùt lǎ:n dū:ølɛ̄:ø mâj ŋâ:j]
ฉัน กำลัง ตาม เก็บ กุหลาบ ‘I’ ‘being’ ‘follow’ ‘pick’ ‘rose’ I am picking roses. [tɕʰǎn kāmlāŋ tā:m kèp kù x là:p]

Table 7: Example of a set of ‘high’ predictability sentence (constructed sentences) (‘x’ signifies an ending of any short-vowel syllables with no final consonant whereas ‘ø’ a syllable with long vowel with no final consonant).

	Mean	Standard deviation
Type 1: ‘Low’ predictability (Preselected)	1.72	0.48
Type 2: ‘Low’ predictability (Constructed)	1.86	0.80
Type 3: ‘High’ predictability (Preselected)	4.48	0.30
Type 4: ‘High’ predictability (Constructed)	3.20	0.80

Table 8: Mean and average rating score of four types of sentences.

We perform ANOVA to test differences between high predictability and low predictability sentence types and use multiple comparisons to check whether each pair is statistically significant as shown in Figure 3. As expected, results show that significant differences are found in pairs of Types 1 and 3; and Types 2 and 4. An important point to be taken here is that levels of predictability could be estimated and later evaluated by native speakers (but this is beyond the scope of the developed tool).

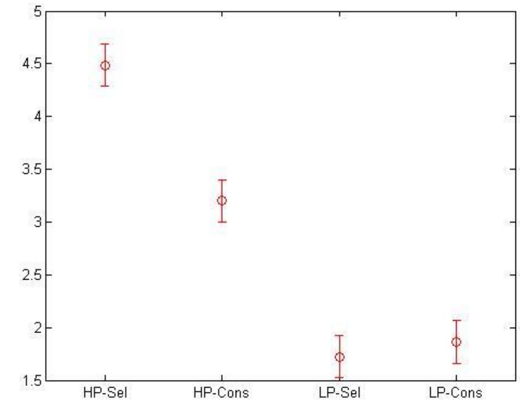

Figure 3: Multiple comparisons between 4 types of sentences. It should be noted that LP-Sel, LP-Cons, HP-Sel, and HP-Cons are referred to Type 1, Type 2, Type 3, and Type 4, respectively.

6 Discussion and Future Direction

We believe that we have successfully proposed and outlined procedure as well as constructed an efficient tool for constructing PB sentence sets. Importantly, a main advantage of our proposed procedure and tool is that it is easy to administer and create sets of words that are close to the desired distribution. As previously mentioned, a fully automated tool of this type would be ideal, but would require other crucial components such as a language model and other information associated with each word (e.g., part of speech).

The procedure and tool outlined here have characteristics that make them potentially useful in other applications and can be applied to other languages, but will certainly require a language specific set of data (i.e., phoneme distribution and language-specific grapheme-to-phoneme software).

Acknowledgments

This work was supported by Thailand Graduate Institute of Science and Technology (TGIST), NSTDA (TGIST 01-57-020) to the first author. Special thanks to Thanaporn Anansiripinyo for her help with Thai-to-English gloss and translation.

References

R. C. Bilger, J. M. Nuetzel, W. M. Rabinowitz, and C. Rzeczkowski. 1984. Standardization of a Test of Speech Perception in Noise. *Journal of Speech, Language, and Hearing Research*, 27(1): 32–48.

A. Chotimongkol, K. Saykhum, P. Chootrakool, N. Thatphithakkul, C. Wutiwiwatchai. 2009. LOTUS-BN: A Thai Broadcast News Corpus and Its Research Applications. *Proceedings of Oriental-COCOSDA*, Xinjiang, China.

A. Chotimongkol, N. Thatphithakkul, S. Purodakananda, C. Wutiwiwatchai, P. Chootrakool, C. Hansakunbuntheung, A. Suchato, and P. Boonpramuk. 2010. The Development of the Large Thai Telephone Speech Corpus: LOTUS-Cell 2.0. *Proceedings of Oriental-COCOSDA*, Kathmandu, Nepal.

J. P. Egan. 1948. Articulation Testing Methods. *Laryngoscope*, 58(9): 955–991.

Q. Fu, M. Zhu, and X. Wang. 2011. Development and Validation of the Mandarin Speech Perception Test. *Journal of the Acoustical Society of America*, 129(6): EL267–273.

D. N. Kalikow, K. N. Stevens, and L. L. Elliot. 1977. Development of a Test of Speech Intelligibility in Noise using Sentence Materials with Controlled Word Predictability. *Journal of the Acoustical Society of America*, 61(5): 1337–51.

B. Kollmeier and M. Wesselkamp. 1997. Development and Evaluation of a German Sentence Test for Objective and Subjective Speech Intelligibility Assessment. *Journal of the Acoustical Society of America*, 102(4): 2412–21.

S. Komalarajun. 1979. *Development of Thai Speech Discrimination Materials*. Master's thesis, Department of Communication Disorders, Faculty of Graduate Studies, Mahidol University, Bangkok, Thailand.

K. Kosawat, M. Boriboon, P. Chootrakool, A. Chotimongkol, S. Klaithin, S. Kongyoung, K. Kriengket, S. Phaholphinyo, S. Purodakananda, T. Thanakulwarapas, and C. Wutiwiwatchai. 2009. BEST 2009: Thai Word Segmentation Software Contest. *Proceedings of 8th International Symposium on Natural Language Processing*, Bangkok, Thailand: 83–88.

Ministry of Education. 1986. *Basic words for teaching Thai language primary education*. Department of Curriculum and Instruction Development, Bangkok, Thailand.

A. Munthuli, P. Sirimujalin, C. Tantibundhit, K. Kosawat, and C. Onsuwan. 2014. Constructing Thai Phonetically Balanced Word Recognition Test in Speech Audiometry through Large Written Corpora. *Proceedings of 17th Oriental Chapter of COCOSDA*, Phuket, Thailand.

A. Munthuli, C. Tantibundhit, C. Onsuwan, K. Kosawat, and C. Wutiwiwatchai. 2015. Frequency of Occurrence of Phonemes and Syllables in Thai: Analysis of Spoken and Written Corpora. *Proceedings of 18th International Congress of Phonetic Sciences*, Glasgow, Scotland.

M. Nilsson, S. D. Soli, and J. A. Sullivan. 1994. Development of the Hearing in Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise. *Journal of the Acoustical Society of America*, 95(2): 1085–99.

R. Plomp and A. M. Mimpen. 1979. Improving the Reliability of Testing the Speech Reception Threshold for Sentences. *Audiology*, 18(1): 43–52.

R. Sangworasin. 2003. *Education of Kindergarten's Vocabulary 4 -5 years old in Bangkok.* Bachelor's Thesis, Department of Education, Primary Education, Faculty of Education, Chulalongkorn University, Bangkok. Thailand.

R. Sripaiwan. 1994. *Thai textbooks set "Mana Manee Piti Chujai*". Department of Curriculum and Instruction Development, Ministry of Education, Bangkok, Thailand.

Thai Children Stories. 1990. Retrieved from http://www.nithan.in.th/.

A. Thangthai, C. Hansakunbuntheung, R. Siricharoenchai, and C. Wutiwiwatchai. 2006. Automatic Syllable Pattern Induction in Statistical Thai Text-to-phone Transcription. *Proceedings of Interspeech*, Pittsburgh, PA.

N. J. Versfeld, L. Daalder, J. M. Festen, and T. Houtgast. 2000. Method for the Selection of Sentence Materials for Efficient Measurement of the Speech Reception Threshold. *Journal of the Acoustical Society of America*, 107(3): 1671–84.

W. Wissawapaisal. 2002. *Development of the Thai Synthetic Sentence Identification Test*. Master's thesis, Department of Communication Disorders, Faculty of Graduate Studies, Mahidol University, Bangkok, Thailand.

C. Wutiwiwatchai, P. Cotsomrong, S. Suebvisai, and S. Kanokphara. 2002. Phonetically Distributed Continuous Speech Corpus for Thai Language. *Proceedings of 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, Spain: 869–872.