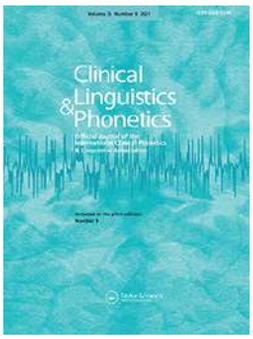

Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

The development of Thai monosyllabic word and picture lists applicable to interactive speech audiometry in preschoolers

Apit Hemakom, Sujinat Jitwiriyanont, Anocha Rugchatjaroen & Pasin Israsena

To cite this article: Apit Hemakom, Sujinat Jitwiriyanont, Anocha Rugchatjaroen & Pasin Israsena (2021) The development of Thai monosyllabic word and picture lists applicable to interactive speech audiometry in preschoolers, Clinical Linguistics & Phonetics, 35:9, 809-828, DOI: 10.1080/02699206.2020.1830301

To link to this article: https://doi.org/10.1080/02699206.2020.1830301

Published online: 04 Nov 2020.

CLINICAL LINGUISTICS & PHONETICS
2021, VOL. 35, NO. 9, 809–828
https://doi.org/10.1080/02699206.2020.1830301

Taylor & Francis
Taylor & Francis Group

The development of Thai monosyllabic word and picture lists applicable to interactive speech audiometry in preschoolers

Apit Hemakom[a], Sujinat Jitwiriyanont[b], Anocha Rugchatjaroen[a], and Pasin Israsena[a]

[a]National Electronics and Computer Technology Center, Pathumthani, Thailand; [b]Department of Linguistics and Southeast Asian Linguistics Research Unit, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

ABSTRACT

Interactive speech audiometry is the assessment of speech comprehension and phonological discrimination through automated means. In order for the performance of such assessments in preschoolers to be successful, the employed list of words and pictures must be easily recognized both linguistically and visually. That is, the children must be able to easily associate the sound they hear with the picture they see with a high degree of certainty. To this end, a Thai monosyllabic word and picture list called NCU-20 (NECTEC-CU-20) is proposed. The word lists for Thai vowel and consonant hearing tests are designed with an awareness of phonetic environments. Regarding Thai vowels, both monophthongs and diphthongs, with all qualities and quantities, are examined. Initial consonants are categorized based on places and manners of articulation. The effectiveness of the list is objectively and subjectively verified through Thai Textbook Corpus, Thai National Corpus, Zipf scores, a listening test of preschoolers with normal hearing, and our proposed ranking systems referred to as Tier-1st, Tier-3/3, and Overall Tier. The final suggested word and picture list comprises 45 items (words) covering 35 vowels and consonant groups in the Thai Language.

ARTICLE HISTORY
Received 12 May 2020
Revised 25 September 2020
Accepted 26 September 2020

KEYWORDS
interactive speech audiometry; preschoolers; monosyllabic words; word and picture lists; Thai language

Introduction

Speech audiometry is a clinical practice typically carried out as a complement to pure-tone audiometry (PTA) in order to evaluate not only the absolute hearing threshold but also speech comprehension and phonological discrimination of an individual (Chaix & Lewis, 2016). Speech audiometry comprises three types of tests (Tanaka et al., 2018). Speech recognition threshold (SRT) measures the lowest level of sounds in decibel hearing level (dB HL) at which a person can hear and repeat disyllabic words 50% of the time. However, if a person has poor hearing capability and is only aware that test words have been presented but cannot recognize and repeat those words, a test called speech detection (reception, awareness) thresholds (SDTs) may be used as an alternative. It determines the lowest dB HL where the presence of speech (disyllabic words) is correctly detected. The goals of the SRT and SDT tests are 1) to validate the results obtained from PTA and 2) to further determine the presentation level employed in the so-called speech discrimination test. The goal of the speech discrimination test is to determine an individual's optimum performance for word recognition under controlled and standardized conditions. For this test, a list of

CONTACT Apit Hemakom apit.hemakom@nectec.or.th National Electronics and Computer Technology Center, Pathumthani, Thailand

monosyllabic words, typically phonemic balancing (PB), is presented to the patient at SRT +40 dB HL and/or the most comfortable loudness (MCL). This test can be administered in either quiet conditions or in the presence of noise. Word recognition score (WRS) is then obtained by scoring the percentage of monosyllabic words which are correctly repeated. A WRS score of 90–100% is considered normal. It should be noted that SRT indicates the ability to distinguish phonemes, through disyllabic words, while WRS gives an indication of speech comprehension, through monosyllabic words.

Neonatal hearing loss can be detected using several techniques, such as automatic brain response (ABR), otoacoustic emission (OAE), and tympanometry. However, hearing loss can occur 2–3 years after birth and is not identified in advance by these techniques. Such hearing loss can be due to several causes, such as otitis media, recreational noise-induced hearing loss, and ototoxicity. Performing audiometry during early childhood development for early detection and intervention is, therefore, crucial as hearing loss affects learning abilities and hinders language, speech, and emotional development (Olusanya et al., 2014). The challenges of conducting speech audiometry with children are limitations in their lexicon, testing procedures, and the expert personnel required to perform such a complicated task. These experts, however, are considerably lacking in a low-resource, developing country such as Thailand, where a significant proportion of population with hearing impairment, accounting for 18.41% of the country's disabled people and approximately 0.6% of the country's population, have been reported (National Association of the Deaf in Thailand, 2018). Since young children could be inherently more responsive to a social interaction, particularly with their caregivers, 'scripted interactions' (Eisenberg et al., 2007) have recently been proposed as an audiometric technique whereby auditory testing is integrated into a type of scripted interaction between children and caregivers with the aim of enhancing the success rate in hearing evaluation in young children. Yimtae et al. (2018), adopting this approach, proposed an interactive tablet-based hearing screening system for Thai preschoolers. The system employed a list of 24 disyllabic words chosen from an elementary school book and varied in terms of contrastive tones. The user was required to choose the picture corresponding to the sound they heard. Average test time was reduced from 11.79 minutes (standard deviation 3.66 minutes, and 95% confidence interval 10.85–12.71 minutes), using conventional audiometry, to 150.52 seconds (standard deviation 19.07 seconds, and 95% confidence interval 145.71–155.32 seconds), using the proposed system. Compared to the speech reception threshold (SRT) at 20 dB, the system exhibited sensitivity and specificity of 77.42% and 82.11%, respectively. It must be noted that this system intrinsically indicated only children's phoneme distinguishability, via the use of disyllabic words. Interactive hearing assessment for phoneme distinguishability and/or speech comprehensibility in children should employ a set of words and corresponding pictures that can easily be recognized and are self-explanatory, requiring only a small amount of intervention from the parents, caregivers, or medical staff so as to ensure that the test results truly indicate the toddlers' real performances and, ultimately, are reliable.

Despite the availability of Thai word lists for speech audiometry testing, they (except for the word list employed in Yimtae et al., 2018) were designed for a conventional, non-interactive adult audiometric test and cannot be readily adopted for pediatric hearing tests, since they include words which are difficult to be used in the speech comprehension test. More precisely, they contain some words that are unknown to young children particularly at preschooler age and, more importantly, cannot be picturized, such as /bun/ 'merit', /mâak/

'numerous', /ŋâaj/ 'easy' (RAMA SD-1); /tɕʰəəŋ/ 'manner', /júʔ/ 'incite', /nùaŋ/ 'delay' (TU PB'14); /mɛ́ɛw/ 'a name of ethnic group in Thailand', /ráŋ/ 'hesitate or hold back', /nîi/ 'debt' (TU PB'15). To the best of our knowledge, no picturizable and monosyllabic Thai word lists specifically designed for pediatric, interactive audiometric evaluation as an assessment for both phoneme distinguishability and speech comprehension have been proposed.

For the application of interactive speech audiometry as an assessment for *both* phoneme distinguishability and speech comprehension in preschoolers, this study, therefore, proposed a monosyllabic Thai word list specifically designed with the awareness of children's word familiarity (through objective measures, namely TTC, TNC, and Zipf, see the Theory and background section for more details), picturizability, and linguistic factors. We subjectively evaluated the effectiveness of the proposed word list in terms of children's ability to recognize, comprehend, and associate sounds with corresponding pictures based on the experiment with normal-hearing preschoolers, thus verifying our objective word selection. Since the verified word list covers all Thai vowels and consonant groups, it will also be suitable for neural-based interactive speech audiometry proposed by Martin et al. (2008) and Morikawa et al. (2012), where the goal is to determine, through brain responses, if a participant can hear as well as comprehend and associate sounds with corresponding pictures. In these studies, a high magnitude of electrical activity of the brain was observed approximately 300 milliseconds (the so-called P300 response) after the participants were exposed to an auditory (or visual) stimulus which matched the preceding visual (or auditory) stimulus. Our proposed word list can be used in this form of audiometry by first presenting a picture from the list to a participant and then playing back 2 sounds from the list. To challenge the participant, those 2 sounds should be (near) minimal pairs, where one of those 2 sounds must match the picture, and the other must not. We can then determine if the participant can hear, comprehend, and associate the picture with the matched sound by observing whether there exists a P300 response after a matched sound is played. From our list of words familiar to preschoolers with phonetic balance, near or analogous pairs can be selected for diagnostic audiometry. The two sub-lists of vowels and consonants are designed with distributed important phonetic qualities. Therefore, the failure of sound-picture matching task can entail phonetic patterns such as back vowels or obstruent consonants. By doing so, neural-based interactive speech audiometry using our proposed word list will allow for simultaneous evaluation of speech comprehension and phoneme distinguishability through our meaningful, picturized monosyllabic words covering all vowels and consonant groups in the Thai language.

Theory and background

There are some well-constructed Thai word lists available. Some of them are in monosyllabic form, which could be appropriate for phoneme perception and comprehension tests. In 1968, Amatayakul presented RAMA.SD1 and RAMA.SD2 word lists containing 5 and 4 sub-lists, respectively (Amatayakul, 1968). The lists cover all Thai monophones and some cluster appearances, that is, cluster-phones or cluster-consonant phones (There are 7 of these in Thai: /pʰr/, /pr/, /kʰw/, /kʰr/, /kl/, /kr/, and /tr/). However, the lists were designed to be played with decibel intelligibility to depict word familiarity and difficulty equivalency, hence phoneme occurrence symmetry was not included in the derivation. Munthuli et al. (2014) presented TU PB'14, a phonetically balanced word list reflecting a proper phoneme

distribution analysed from a large Thai text corpus called InterBest (Munthuli et al., 2014). A year later, Poonyaban et al. presented TU PB'15, a revision of TU PB'14 sets (Poonyaban et al., 2015). It claimed to present a more proper word list for speech perception test, as it reflects phoneme distribution of a spoken corpus called LOTUS-Cell instead of a written corpus as in TU PB'14. These word lists, however, contain some words which cannot be picturized and be easily recognized, especially by children, such as /mâak/ which means 'numerous'; /ŋâaj/ which means 'easy'; /júʔ/ which means 'incite'; /nùaŋ/ which means 'delay'; and /ráŋ/ which means 'hesitate or hold back'. In addition, some words in the lists would be highly unfamiliar to preschoolers, such as /bun/ which means 'merit'; /méɛw/ which is a name of an ethnic group in Thailand; and /nîi/ which means 'debt'. Recently in 2018, Tantibundhit et al. analysed and presented a test tool for TU PB'14 set, and its results confirm that the word lists still have room for improvement.

Since the Thai language is a tonal language, every single syllable of Thai has its own contrastive pitch or lexical tone. A coarticulation of two consecutive syllables can cause acoustic variability. *To minimize the coarticulation effect, a set of monosyllabic words reflecting all Thai acoustic phenomena is needed and must include words with obvious formant patterns of vowels and clear acoustic characteristics of initial and final consonants.*

The Thai phoneme inventory consists of 21 consonants (see Table 1), 21 vowels (see Table 2), and 5 tones. All consonants occur in onset position. Only voiceless unaspirated plosives /p, t, k, ʔ/, nasals /m, n, ŋ/, and glides /j, w/ are allowed in coda position. Thai has 18 monophthongs and 3 diphthongs. Vowel length is contrastive in monophthongs. The tonal system consists of 5 tones: High tone [45], Mid tone [33], Low tone [21], Falling tone [451], and Rising tone [214].

A set of words for preschoolers needs a proper resource of word sets. This research used Thai Textbook Corpus, TTC (Champaiboon & Aroonmanakun, 2016), built to study the

Table 1. Thai consonants.

	bilabial	labiodental	alveolar	post-alveolar	palatal	velar	glottal
plosive	p p^h b		t t^h d			k k^h	ʔ
nasal	m		n			ŋ	
trill			r				
fricative		f	s				h
affricate				tɕ $tɕ^h$			
approximant					j		
lateral approximant			l				

w voiced labial-velar approximant

Table 2. Thai vowels.

	monophthong					
	front		central		back	
	short	long	short	long	short	long
close	i	ii	ɨ	ɨɨ	u	uu
mid	e	ee	ə	əə	o	oo
open	ɛ	ɛɛ	a	aa	ɔ	ɔɔ

diphthong		
ia	ɨa	ua

development of word frequency in Thai textbooks. Even though the corpus contains all words from fundamental textbooks of elementary and high-level curricula (all 12 grades), this research used only the word set found in 1st grade textbooks (Champaiboon & Aroonmanakun, 2016).

Methods

Designing word lists

There are 2 main emphases for word selection: phonetic coverage and general usage of the words for a specific age range.

Phonetic coverage

The stimulus material comprised two word lists of Thai words intelligible to preschoolers: one for a vowel hearing test and the other consonants. Nouns were the most preferable in order to reduce the complication of word representation in a picture. For cases in which nouns with particular vowels were not familiar to children, verbs and adjectives were employed instead. All of the tokens were monosyllabic words with 3 syllable types: open syllables, plosive-final syllables, and nasal-final syllables. Monosyllabic words were chosen in order to minimize the linguistic redundancy cues in multisyllabic words as discussed by Kirk et al. (2000). Multisyllabic words have cues other than speech signal, such as phoneme duration reduction/extension, vowel tone modification for syllable stress, etc., to aid word recognition and have fewer phonetically similar words than monosyllabic counterparts. Subsequently, monosyllabic words were employed in this study to minimize the influence of other linguistic factors.

For tokens with final consonants, three places of articulation, i.e. bilabial, alveolar, and velar, were chosen, because they exhibit three patterns of second formant frequency (F2) transition. The F2 transition is significant for speech perception. For the vowel word list, all vowels in the Thai sound system, namely 9 long monophthongs, 9 short monophthongs, and 3 diphthongs, were included. Almost all of the initial consonants were bilabial consonants except for /ə/, with the initials being nasal and alveolar due to the limitation of children's word intelligibility. For the consonant word list, consonants were categorized based on manner and place of articulation. All words in this list were controlled to have /aa/. In summary, all Thai vowels with different three-dimension vowel qualities (tongue height, tongue advancement, and lip posture) were tested, while consonants, called 'consonant group' in this study, were grouped based on two out of three contrastive dimensions (place and manner of articulation).

General usage of the words for a specific age group

Usage was considered in both objective and subjective perspectives. We first objectively selected words based on their statistical appearances and usage derived from 1) Thai Textbook Corpus (TTC), 2) Thai National Corpus (TNC), 3) Zipf scores, and 4) part of speech. A 1st grade word list in TTC, the closest recorded corpus available in Thai, is representative of the word set used with preschoolers in this research. The TNC contains 33 million words collected from various resources (Aroonmanakun et al., 2009). TTC and TNC frequencies are, therefore, the number of occurrences of a word found in the studied corpora. For example, /mi̯i̯/ 'bear' has TTC and TNC frequencies of 77 and 682, respectively,

meaning that there are 77 and 682 occurrences of /mĭi/ in the TTC and TNC corpora, respectively. Zipf scoring is a logarithmic frequency score, ranging from 1 to 7, where values of 3 and lower indicate low-frequency words while values of 4 or higher indicate high-frequency words. The formula can be found in Van Heuven et al. (2014). The selection criteria were: 1) word frequency from TTC, 2) word frequency from TNC, 3) Zipf score, and 4) part of speech – with nouns being the most preferable, followed by verbs and adjectives.

Tables 3–5 show all words chosen in accordance with the aforementioned objective criteria, covering all vowels and consonant groups. Each vowel and consonant group (except /ŋ/), represented, respectively, by grey and white, has 2 words. From each of these pairs, the word achieving the best result from the listening test was then added to the final suggested word list. Consonants are separated in groups (the so-called consonant group) according to their manner and place of articulation. For example, /t/ and /t^h/ are in the same group even though they are different in aspiration. Since this work proposes a word list for Thai preschoolers, recognizability and picturizability of words within this group were emphasized. The TTC frequencies vary from 1 to 3005, and this research took these numbers, along with the TNC frequencies, into account if the corresponding words appeared at all in the corpora, and if the words would thus be familiar to children. In this way, words with Zipf larger than 4 (high-frequency words) were selected. Nouns were the most preferable as they are easier to picturize compared to verbs and adjectives and, hence, had a higher chance of being correctly recognized by children. For the subjective perspective, the listening test is described below in the Data collection section.

Sound recordings

Sound recordings were made in a silent (33 dBA of background noise) 6.2 m x 6.8 m room. A professional female speaker with 15 years' experience, a Bangkok accent, normal articulators, and voice talent enunciated with normal vocal effort at an approximately 65 decibel sound pressure level (dB SPL) into a high-sensitivity microphone (BSWA Tech Model MPA416) with a flat frequency response (±0.5 dB) at a range of 20 Hz to 10 kHz. The microphone was located approximately 15 cm away from the speaker and was connected to a microphone conditioning unit (BSWA Tech Model MC102). The signal was acquired with a studio-grade USB audio interface (M-Audio Fast Track Ultra 8 R) with a 44.1 kHz sampling frequency and a 24-bit amplitude resolution. The recorded sound files are 1.48 seconds in length on average, silences included. Speaker pitch ranged from 124.96 to 249.93 Hz. Average signal-to-noise ratio (SNR) of all the 73 recorded words was 47.1 dB with a standard deviation of 2.77 dB (see Figure 1).

Picture design

The pictures were designed to be simple, self-explanatory, and easily recognized by children aged 4–5 years old. For the purposes of simplicity and self-explanation, in most cases, only one object representing the word of interest in a picture was used (see Figure 2a–d). In some cases, however, an arrow had to be used in a picture in order to: 1) point to a body part associated with the word of interest (see Figure 3a-d) show an action or a transition in the state of things (see Figure 4a,b).

Table 3. Vowel word list for 9 long monophthongs and 9 short monophthongs. IPA stands for International Phonetic Alphabet, POS part of speech, n. noun, v. verb, adj. adjective, [v., adj.] verb or adjective, [v./adj.] verb for the first word and adjective for the second word. See the Data collection section for details about Classroom and the Results section for details about Tier-1st, Tier-3/3, and Overall Tier.

Vowel type	Word number	IPA	Transcription	Word	TTC	TNC	Zipf	POS	Class room	Tier −1st	Tier −3/3	Overall Tier
Long vowel	1	ii	**/mi̯i/**	**bear**	**77**	**682**	**4.31**	**n.**	**A**	**1**	**1**	**1**
	2	ii	/pi̯ik/	wing	147	1251	4.57	n.	D	3	3	3
	3	ee	/pe̯e/	rucksack	1	963	4.46	n.	C	6	6	6
	4	ee	**/me̯ek/**	**cloud**	**92**	**1222**	**4.56**	**n.**	**D**	**4**	**4**	**4**
	5	ɛɛ	**/mɛ̯ɛ/**	**mother**	**3005**	**30084**	**5.95**	**n.**	**B**	**2**	**2**	**2**
	6	ɛɛ	/pɛ̯ɛt/	eight	27	2482	4.87	n.	E	4	4	4
	7	aa	**/ma̯a/**	**dog**	**489**	**3753**	**5.05**	**n.**	**B**	**1**	**1**	**1**
	8	aa	**/ma̯a/**	**horse**	**209**	**4119**	**5.09**	**n.**	**D**	**1**	**1**	**1**
	9	ɨɨ	/mɨ̯ɨ/	hand	354	18626	5.75	n.	A	2	2	2
	10	ɨɨ	**/pɨ̯ɨn/**	**gun**	**86**	**3479**	**5.02**	**n.**	**E**	**1**	**1**	**1**
	11	əə	/pʰə̯əm/	add	60	21554	5.81	v.	C	5	6	5.5
	12	əə	/pə̯ət/	open	183	21352	5.81	v., adj.	E	5	6	5.5
	13	uu	**/pu̯u/**	**crab**	**158**	**2708**	**4.91**	**n.**	**C**	**1**	**1**	**1**
	14	uu	**/mu̯u/**	**pig**	**161**	**2635**	**4.90**	**n.**	**E**	**1**	**1**	**1**
	15	oo	**/bo̯o/**	**ribbon**	**37**	**1365**	**4.61**	**n.**	**A**	**1**	**1**	**1**
	16	oo	/pô̯oŋ/	boastful	3	339	4.01	adj.	B	1	2	1.5
	17	ɔɔ	**/mɔ̯̂ɔ/**	**pot**	**62**	**1188**	**4.55**	**n.**	**D**	**1**	**2**	**1.5**
	18	ɔɔ	/mɔ̯ɔn/	pillow	47	776	4.37	n.	C	2	2	2
Short vowel	19	i	**/bi̯n/**	**fly**	**656**	**4252**	**5.10**	**v.**	**A**	**1**	**1**	**1**
	20	i	/pi̯t/	close	123	8836	5.42	v., adj.	D	6	6	6
	21	e	/pe̯t/	duck	156	898	4.43	n.	C	2	3	2.5
	22	e	**/pʰe̯t/**	**diamond**	**15**	**4902**	**5.17**	**n.**	**E**	**1**	**1**	**1**
	23	ɛ	**/pʰɛ̯̂ʔ/**	**goat**	**219**	**336**	**4.00**	**n.**	**B**	**1**	**1**	**1**
	24	ɛ	/pɛ̯̂ʔ/	stick	4	573	4.23	v.	C	3	3	3
	25	a	/pʰa̯t/	fan	134	1811	4.73	n.	B	2	4	3
	26	a	**/pʰa̯k/**	**vegetable**	**298**	**3800**	**5.06**	**n.**	**D**	**1**	**3**	**2**
	27	ɨ	/mɨ̯k/	ink	2	701	4.32	n.	C	2	3	2.5
	28	ɨ	**/pʰɨ̯̂ŋ/**	**bee**	**116**	**438**	**3.54**	**n.**	**E**	**1**	**1**	**1**
	29	ə	**/ŋə̯n/**	**money**	**239**	**32452**	**5.99**	**n.**	**A**	**1**	**1**	**1**
	30	ə	/lə̯ʔ/	dirty	22	243	3.86	v., adj.	E	3	3	3
	31	u	**/pʰu̯ŋ/**	**belly**	**6**	**364**	**4.04**	**n.**	**A**	**1**	**2**	**1.5**
	32	u	/mu̯ /	mosquito net	15	337	4.00	n.	E	3	4	3.5
	33	o	**/pʰo̯m/**	**hair**	**455**	**90527**	**6.43**	**n.**	**D**	**1**	**1**	**1**
	34	o	**/mo̯t/**	**ant**	**47**	**656**	**4.29**	**n.**	**B**	**1**	**1**	**1**
	35	ɔ	/bɔ̯̂ʔ/	cushion	2	491	4.17	n.	D	3	4	3.5
	36	ɔ	**/pɔ̯ŋ/**	**bulge**	**9**	**224**	**3.83**	**v., adj.**	**C**	**3**	**3**	**3**

The final suggested words, in bold, can be downloaded from https://20to20k.nstda.or.th/Publications/NCU-20_ThaiMonosyllabicWordList.zip.

Table 4. Vowel word list for 3 diphthongs. ***The final suggested words are in bold.***

Vowel type	Word number	IPA	Transcription	Word	TTC	TNC	Zipf	POS	Class room	Tier −1st	Tier −3/3	Overall Tier
Diph thong	37	ia	**/pi̯ak/**	**wet**	**43**	**960**	**4.46**	**v., adj.**	**A**	**3**	**3**	**3**
	38	ia	**/bi̯at/**	**jostle**	**19**	**488**	**4.16**	**v.**	**B**	**3**	**3**	**3**
	39	ɨa	/pɨ̯an/	stain	17	1144	4.53	v.	D	3	3	3
	40	ɨa	**/pʰɨ̯an/**	**friend**	**612**	**23736**	**5.85**	**n.**	**C**	**1**	**2**	**1.5**
	41	ua	**/bu̯a/**	**lotus**	**101**	**1906**	**4.76**	**n.**	**B**	**1**	**1**	**1**
	42	ua	**/mu̯ak/**	**hat**	**120**	**921**	**4.44**	**n.**	**E**	**1**	**1**	**1**

Table 5. Consonant word list for 15 consonant groups. ***The final suggested words are in bold.***

Consonant type	Word number	IPA	Transcrip tion	Word	TTC	TNC	Zipf	POS	Class room	Tier −1st	Tier −3/3	Overall Tier
Stop/Affricate	43	p	/paa/	throw	11	3205	4.98	v.	A	1	4	2.5
	44	p	**/pàak/**	**mouth**	**238**	**10680**	**5.51**	**n.**	**C**	**1**	**1**	**1**
	45	p^h	**/p^hâa/**	**cloth**	**5**	**586**	**4.24**	**n.**	**D**	**1**	**1**	**1**
	46	t	**/taa/**	**eye**	**875**	**22543**	**5.83**	**n.**	**A**	**1**	**1**	**1**
	47	t^h	/t^haa/	paint	33	4246	5.10	v.	D	4	4	4
	48	d	/dàap/	sword	55	1260	4.58	n.	C	1	2	1.5
	49	tɕ	/tɕaan/	plate	64	2744	4.92	n.	B	2	2	2
	50	$tɕ^h$	**/tɕ^{h}áaŋ/**	**elephant**	**404**	**2878**	**4.94**	**n.**	**D**	**1**	**1**	**1**
	51	k^h	**/k^{h}ǎa/**	**leg**	**237**	**6472**	**5.29**	**n.**	**B**	**1**	**2**	**1.5**
	52	k	/kaa/	crow	238	5859	5.24	n.	E	2	2	2
	53	ʔ	**/ʔàaŋ/**	**basin**	**7**	**841**	**4.40**	**n.**	**C**	**1**	**1**	**1**
	54	ʔ	**/ʔâa/**	**open/agape**	**37**	**1079**	**4.51**	**v./adj.**	**E**	**1**	**1**	**1**
Fricative	55	f	**/fáa/**	**sky**	**137**	**6329**	**5.28**	**n.**	**A**	**1**	**1**	**1**
	56	f	**/fǎa/**	**lid**	**45**	**1110**	**4.52**	**n.**	**D**	**1**	**1**	**1**
	57	s	**/sǎam/**	**three**	**407**	**17959**	**5.73**	**n.**	**A**	**1**	**1**	**1**
	58	s	/sàat/	splash	29	736	4.34	v.	C	3	4	3.5
	59	h	**/hǎaŋ/**	**tail**	**172**	**1607**	**4.68**	**n.**	**B**	**1**	**1**	**1**
	60	h	/hàat/	beach	72	1259	4.58	n.	E	2	2	2
Nasal	61	m	**/măa/**	**dog**	**489**	**3753**	**5.05**	**n.**	**A**	**1**	**1**	**1**
	62	m	**/máa/**	**horse**	**209**	**4119**	**5.09**	**n.**	**C**	**1**	**1**	**1**
	63	n	**/nàa/**	**face**	**651**	**44148**	**6.12**	**n.**	**D**	**2**	**2**	**2**
	64	n	/naa/	rice field	330	12420	5.57	n.	E	3	3	3
	65	ŋ	**/ŋaa/**	**ivory**	**73**	**1187**	**4.55**	**n.**	**B**	**4**	**4**	**4**
Glide	66	w	**/waaŋ/**	**put**	**171**	**11263**	**5.53**	**v.**	**A**	**1**	**1**	**1**
	67	w	**/wâat/**	**draw**	**78**	**3276**	**4.99**	**v.**	**D**	**1**	**1**	**1**
	68	j	/jaa	drug	217	18103	5.73	n.	B	1	2	1.5
	69	j	**/jaaŋ/**	**rubber band**	**26**	**2050**	**4.79**	**n.**	**C**	**1**	**1**	**1**
Trill	70	r	**/raa/**	**fungus**	**32**	**6200**	**5.27**	**n.**	**A**	**1**	**3**	**2**
	71	r	/raaŋ/	rail	28	823	4.39	n.	B	2	3	2.5
Lateral	72	l	**/laa/**	**donkey**	**205**	**6301**	**5.28**	**n.**	**A**	**1**	**3**	**2**
	73	l	/lâak/	drag	55	2106	4.80	v.	B	3	4	3.5

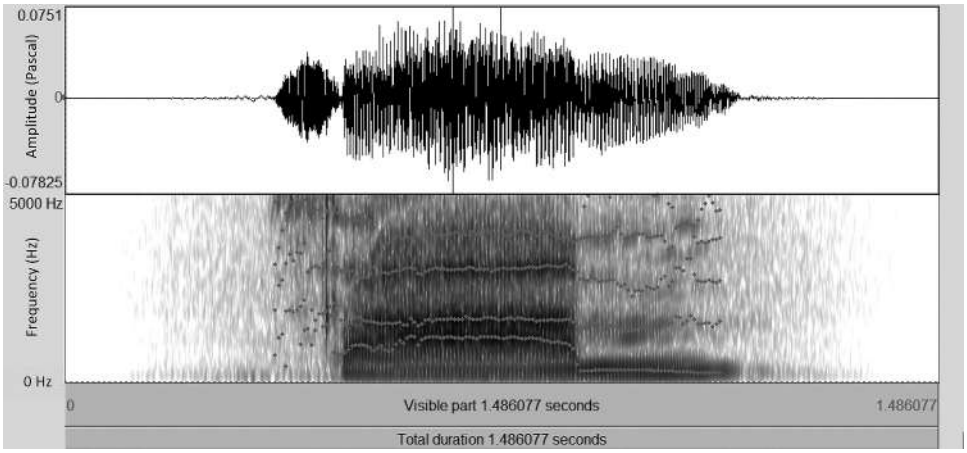

Figure 1. Spectrogram of recorded /s-ăa-m/. The top window shows its waveform shape. The SNR of the recorded word was 42.78 dB. The bottom shows the spectrogram, while its formants are shown as grey dots (plotted using Praat).

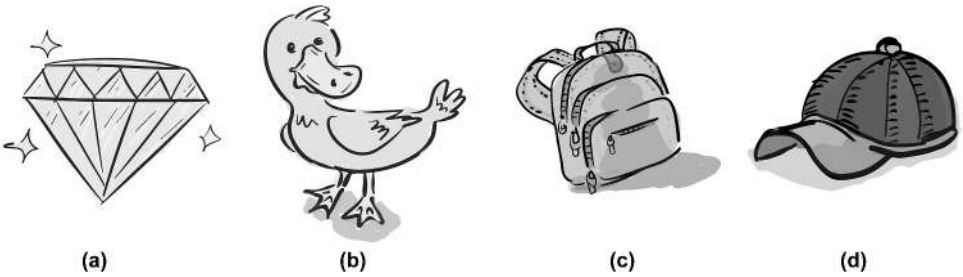

Figure 2. Examples of pictures containing one object each. (a) /p^{h}ét/ 'diamond. (b) /pèt/ 'duck'. (c) /pêe/ 'rucksack'. (d) /mùak/ 'hat'. Nouns were most preferred.

Experimental design

Splitting the words into 5 groups

The 73 proposed words were split into 5 groups, where recognizability (the ability to associate sounds with the corresponding pictures) of the words in a group would then be tested out among all the children in a classroom. The number of words in each of the Groups A to D was 15, and Group E was 13. Although different groups of children judged different items, they were tested with words from both consonant and vowel sub-lists, balanced in terms of phonetic qualities, i.e. front/central/back vowels and obstruent/sonorant consonants. It must be noted that words with the same test phoneme, either vowel or consonant, were not put into the same group.

Designing answer sheets

Pictures corresponding to the words in a designated word group were used to form a set of A4-sized answer sheets (5 sets in total, Sets A–E). Since the recognizability test for each

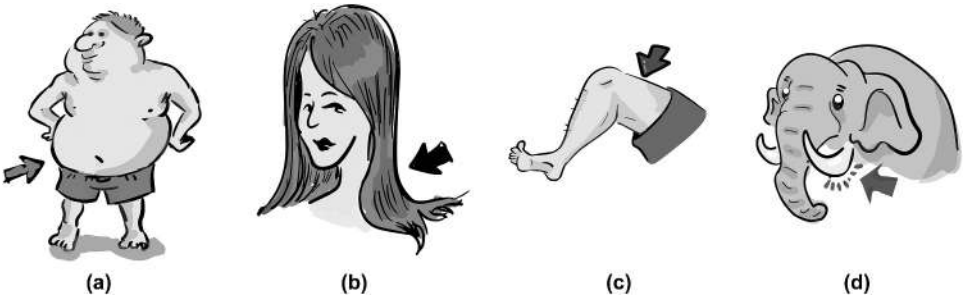

Figure 3. Examples of figures using an arrow. (a) /pʰuŋ/ 'belly'. (b) /pʰŏm/ 'hair'. (c) /kʰăa/ 'leg'. (d) /ŋaa/ 'ivory'. An arrow was used to point to a body part of interest.

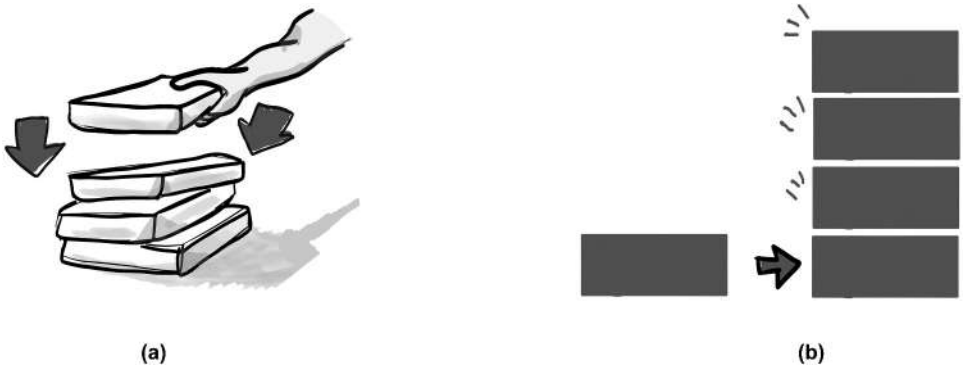

Figure 4. An arrow was used to show an action or a transition. (a) /waaŋ/ 'put'. (b) /pʰə̂əm/ 'add'.

word was to be repeated 3 times, the total number of answer sheets was 45 for Sets A–D (15 words x 3 repetitions = 45 answer sheets) and only 39 for Set E (13 words x 3 repetition = 39 answer sheets).

On each answer sheet, the ('target') picture corresponding to the sound presented during our experiment (see the Data collection section) was positioned on either the left, the middle, or the right of the answer sheet, together with 2 other incorrect, non-corresponding pictures ('non-targets') occupying the 2 remaining positions. Two non-target pictures for each answer sheet were randomly selected from a total of 72 non-corresponding pictures with the following criteria:

(1) 2 non-target pictures on an answer sheet were not the same;
(2) for each complete repetition of testing of the 15/13 target pictures, a non-target picture was to be used only 3 times at most (9 times in total for all 3 repetitions) in order to prevent overuse of any non-target and the potential for a biased test result towards the overused non-target.

The following criteria were also imposed on designing our answer sheets:

(1) For a single repetition, positions of target pictures (left, middle, or right) on consecutive answer sheets were allowed to remain the same no more than 3 times consecutively in order to prevent a person taking the test from falsely anticipating the next target picture based on an often repeated position of previous targets;
(2) for different repetitions, positions of the same target picture on the corresponding 3 answer sheets were changed to prevent a person taking the test from memorizing the position of the target during a previous repetition;
(3) pictures which were visually similar or had similar meanings, such as 'cloud' and 'sky', were not used in the same answer sheet in order to avoid confusion in children's graphical interpretation. Similarly, in future application of final word lists in interactive speech audiometry, such visually similar words should not be tested against one another to avoid graphical confusion, but rather to focus solely on assessing children's hearing capability.

Data collection

We recruited 130 children from 5 classrooms (Classrooms A to E) aged between 4 and 5 years with no medical record of hearing impairment. Parental consent was obtained before the test date. One student from Classroom B was unable to finish the test due to illness and was subsequently excluded. The number of participants who completed the test, therefore, was 129: 26, 28, 25, 27, and 23 children from Classrooms A to E, respectively.

A recognizability test of words in a designated group was carried out among children from a single classroom – words in Groups A to E with children in Classrooms A to E, respectively. A test among all children from the same classroom was carried out on the same day (5 days in total) in a 15 × 20 metre hall equipped with 6 loudspeakers located on the side walls (3 on each) 3 metres above the floor, facing approximately 45 degrees downwards (see Figure 5). There were 12 people conducting data collection on each day: 1 main instructor standing in front of the room and controlling the data collection, 1 person on a stage playing the sound and showing a corresponding answer sheet on a project screen to help guide the children, and 10 adults looking after the children (1 adult per 2–4 children, depending on the seating arrangement).

The children were seated approximately 1.5 metres away from one another. Each child was given a pencil, together with 45 or 39 answer sheets (45 for children in Classrooms A to D, 39 for Classroom E). The recorded sound of each word was played only once through the loudspeakers at approximately 60–70 dBA measured at the centre of the room. After hearing the sound, the children were instructed to not verbally repeat it so as not to distract the others and to mark on a corresponding answer sheet the picture corresponding to the sound. Immediately upon completion of the first repetition of the test (15/13 sounds were played, and 15/13 answer sheets were marked), taking approximately 10 minutes, the second and third repetitions followed, as it was difficult to gain the children's attention on the task; a break between repetitions would otherwise disrupt their attention and interrupt the 'flow' of the test.

Figure 5. Experimental setup for data collection. The subjects were seated approximately 1.5 metres away from one another, with 12 people in total carrying out the experiment: 1 main instructor, 1 person playing media on the stage, and 10 caregivers.

Results

Overall accuracy of each word was obtained by adding the number of subjects who correctly gave the correct answers for the word, from all of the 3 iterations, and dividing the sum by the number of all subjects in the classroom multiplied by 3, given by

$$P_m = \frac{\sum_{i=1}^{3} n_{m,i^{th}}}{N*3} * 100,$$

where $n_{m,i^{th}}$ is the number of subjects providing the correct answer for word m for the i^{th} iteration, and N is the number of subjects in the classroom in which the word was tested ($N = 26$, 28, 25, 27, and 23 for Classrooms A to E, respectively). Average overall accuracy for a given number of words, Q_M, from a set of words, W, is given by

$$Q_M = \frac{\sum_{m \in W} P_m}{M} * 100,$$

where M is the number of words from set W. Figure 6 shows the overall accuracy of each word. The average overall accuracy for the whole set of 73 words, Q_{73} and $W = \{1, 2, 3, \ \ldots \ , 73\}$, was 93.19% with a standard deviation (std.) of 11.28%. However, there were 4 words for which the accuracy was remarkably low: /pêe/ 'rucksack' ($P_3 = 62.67\%$, see arrow 1), /pʰə̂əm/ 'add' ($P_{11} = 49.33\%$, see arrow 2), /pə̀ət/ 'open' ($P_{12} = 52.17\%$, see arrow 3), and /pìt/ 'close' ($P_{20} = 39.51\%$, see arrow 4). Excluding these 4 words increased the average overall accuracy for the remaining 69 words, Q_{69} and $W = \{1, 2, 3, \ \ldots \ , 73\} - \{3, 11, 12, 20\}$, to 95.64% with a standard deviation (std.) of 4.42%. From the remaining 69 words, the word with the lowest accuracy was /n̲aa/ 'ivory' ($P_{65} = 82.14\%$).

The percentage of subjects who correctly gave the correct answer for each word for each iteration, $P_{m,i}th$, is given by

$$P_{m,i^{th}} = \frac{n_{m,i^{th}}}{N} * 100$$

It can be seen from Figure 7 that words /pêe/ 'rucksack' ($P_{3,1^{st}} = 24\%$), /pʰə̂əm/ 'add' ($P_{11,1^{st}} = 44\%$), /pə̀ət/ 'open' ($P_{12,1^{st}} = 52.17\%$), and /pìt/ 'close' ($P_{20,1^{st}} = 29.63\%$) were exceptionally hard to recognize in the first iteration (see arrows 5a–8a), although the

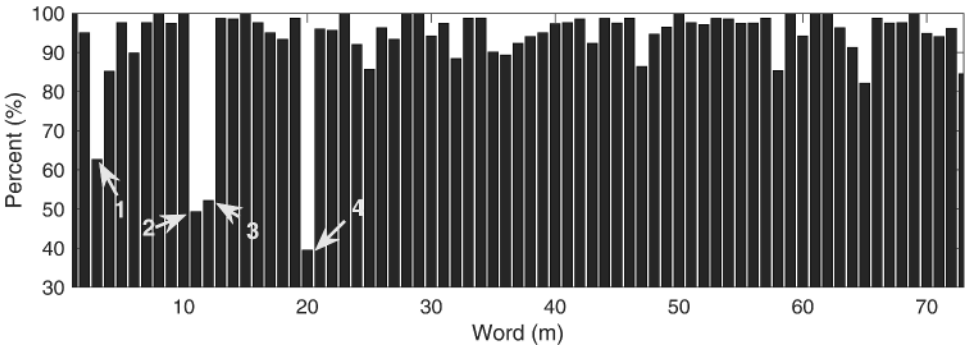

Figure 6. Overall accuracy. 69 out of 73 words achieved over 80% accuracy, except for numbers 3 (/pêe/ 'rucksack'), 11 (/pʰâəm/ 'add'), 12 (/pàət/ 'open'), and 20 (/pìt/ 'close').

percentages peaked in the last iteration at $P_{3,3^{rd}} = 84\%$ (/pêe/ 'rucksack'), $P_{11,3^{rd}} = 76\%$ (/pʰâəm/ 'add'), $P_{12,3^{rd}} = 56.52\%$(/pàət/ 'open'), and $P_{20,3}rd = 62.96\%$ (/pìt/ 'close'), see arrows 5b–8b. The words /pʰâəm/ 'add', /pàət/ 'open', and /pìt/ 'close', all verbs exhibited extremely poor recognizability in the second iteration ($P_{11,2^{nd}} = 28\%$, $P_{12,2^{nd}} = 47.83\%$, and $P_{20,2^{nd}} = 25.93\%$, respectively, see arrows 6c–8 c). Without these 4 words, average accuracies for the first, second, and third iterations, $Q_{M,i^{th}} = \frac{\sum m \in WP_{m,i^{th}}}{M} * 100$, of the 69 words, $M = 69$, were $Q_{69,1}st = 93.73\%$ (std. = 7.13%), $Q_{69,2^{nd}} = 96.91\%$ (std. = 4.97%), and $Q_{69,3^{rd}} = 96.28\%$ (std. = 4.62%), respectively. With these 4 words, average accuracies of the 73 words for the first, second, and third iterations decreased to $Q_{73,1^{st}} = 90.64\%$

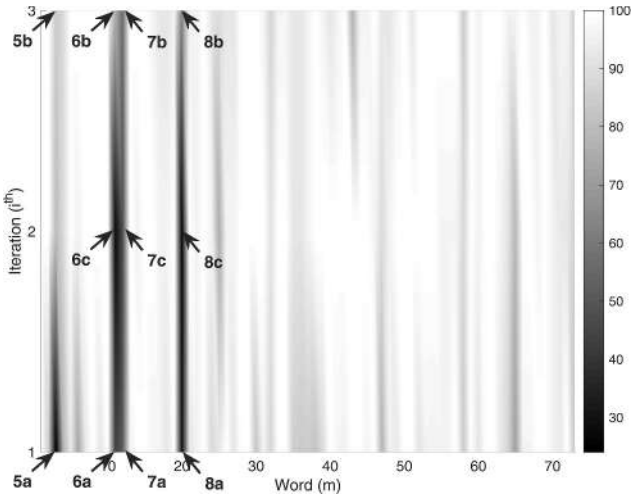

Figure 7. The percentage of subjects giving the correct answer for each iteration. The colour bar on the right represents percentage, where the brightest colour (white) represents the highest value (100) and the darkest colour (black) represents the lowest value. A high percentage of subjects giving the correct answer for the first iteration suggests that the word was easily recognized at first glance. ***A table containing values in the figure can be downloaded from https://20to20k.nstda.or.th/Publications/NCU-20_ThaiMonosyllabicWordList.zip***.

(std. = 14.88%), $Q_{73,2^{nd}} = 94.08\%$ (std. = 13.73%), and $Q_{73,3^{rd}} = 94.83\%$ (std. = 7.95%), respectively.

The percentage of subjects giving k out of 3 correct answers for each word m is given by

$$P_{m,k/3} = \frac{n_{m,k/3}}{N} * 100,$$

where $n_{m,k/3}$ is the number of subjects giving k out of 3 correct answers for the word m.

From Figure 8, it is observed that less than half of the subjects were able to give 3 out of 3 correct answers for the same 4 words (/pêe/ 'rucksack', /pʰâəm/ 'add', /pàət/ 'open', /pìt/ 'close') – $P_{3,3/3} = 24\%$, $P_{11,3/3} = 16\%$, $P_{12,3/3} = 39.13\%$, and $P_{20,3/3} = 11.11\%$ (see arrows 9, 10a–12a). Two out of 3 correct answers for the words /pʰâəm/ 'add', /pàət/ 'open', and /pìt/ 'close' were also obtained from fewer than half of the subjects – $P_{11,2/3} = 48\%$, $P_{12,2/3} = 47.83\%$, and $P_{20,2/3} = 29.63\%$ (see arrows 10b–12b).

Average number of correct answers for each word m is given by

$$R_m = \frac{\sum_{j=1}^{N} c_{m,j}}{N},$$

where $c_{m,j}$ is the number of correct answers Subject j gave for the word m for all 3 iterations. The data were verified as not normally distributed, with distribution of the number of correct answers of the 69 non-problematic words all negatively skewed – the numbers of correct answers were predominantly 3 and occasionally 0, 1, or 2. The 4 problematic words were also considered, with the distributions of one being negatively skewed (/pêe/ 'rucksack'; mostly 2 correct answers) and the distributions of the remaining three being positively skewed (/pʰâəm/ 'add', /pàət/ 'open', and /pìt/ 'close'; mostly 1, 0, and 1 correct answers, respectively). Figure 9 shows the average numbers of correct answers for all 73

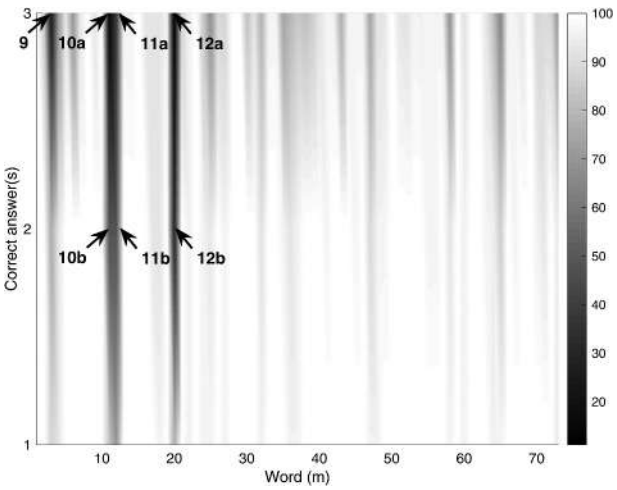

Figure 8. The percentage of subjects giving k out of 3 correct answers. The colour bar on the right represents percentage, where the brightest colour (white) represents the highest value (100) and the darkest colour (black) represents the lowest value. A high percentage of subjects giving 3 out of 3 correct answers suggests that the word was recognized with a high level of certainty. ***A table containing values in the figure can be downloaded from https://20to20k.nstda.or.th/Publications/NCU-20_ThaiMonosyllabicWordList.zip.***

words as well as their 95% confidence intervals calculated by employing the bootstrap technique with 1,000 resamplings. This particular technique was used because: 1) it provides inference about a population from sample data; 2) it does not require assumption of normality of the data; 3) it is a non-parametric technique, requiring no analytical form to help estimate the distribution of the data; and 4) it is a straightforward way to derive estimates of confidence intervals. Using 1,000 resamples was sufficiently large for the estimation of 95% confidence intervals (Davidson & MacKinnon, 2000). Observe that the same 4 words (/pêe/ 'rucksack', /pʰâəm/ 'add', /pàət/ 'open', and /pìt/ 'close') exhibited average numbers of correct answers lower than 2 with low values of the 95% confidence interval: $R_3 = 1.88$ (95% CI 1.48–2.20), $R_{11} = 1.48$ (95% CI 1.08–1.84), $R_{12} = 1.57$ (95% CI 1.04–2.09), and $R_{20} = 1.19$ (95% CI 0.89–1.56). Except for these 4 words, lower bounds of the confidence intervals of the remaining 69 words were all over 2.00.

Each of the 73 words was next classified into 3 tiers addressing different purposes: 1) 'Tier-1st', 2) 'Tier-3/3′, and 3) 'Overall-Tier'. Tier-1st was defined as *the level of difficulty*, ranging from 1 (easiest) to 6 (hardest), *in recognizing a word at first glance* by taking into account the statistical value $P_{m,1^{st}}$ (the percentage of subjects giving the correct answer for word *m* for the first iteration). Tier-3/3 was defined *as the level of certainty*, ranging from 1 (high certainty) to 6 (low certainty), *in which a word would be correctly recognized*, reflected by the statistical value $P_{m,3/3}$ (the percentage of subjects giving 3 out of 3 correct answers for a word). *A word to be used in interactive speech audiometry should be easily recognized at first glance with a high level of certainty*. We, therefore, for each word *m*, averaged values of these 2 Tiers to obtain *the final verdict, referred to as the Overall Tier. A word with a lower value in the Overall Tier, compared to another word with a higher value, is suggested for use in interactive speech audiometry*. Table 6 shows different levels of Tier-1st and Tier-3/3 values, classified through values of $P_{m,1^{st}}$ and $P_{m,3/3}$, respectively. Tables 3–5 show Tier-1st, Tier-3/3, and Overall Tier values for each word. Within a vowel or a consonant group, the word(s) with the lowest value in the Overall Tier is suggested for interactive speech audiometry and is marked in bold. Note that neither of the words /pʰâəm/ 'add' and /pàət/ 'open', with an Overall Tier value of 5.5 for the vowel /əə/, are recommended due to their poor recognizability as previously described.

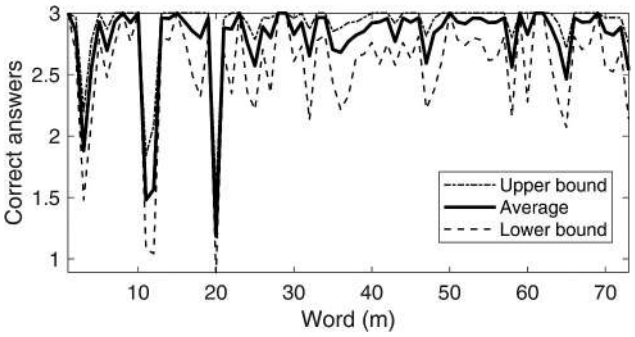

Figure 9. Average number of correct answers for each word, and lower and upper bounds of the 95% confidence interval. Only 4 words (*m* = 3,11, 12, and 20) exhibited an average number of correct answers lower than 2 and low values of 95% confidence interval.

Observe from Figure 10 that for 35 out of 36 vowels and consonant groups in the Thai language, use of words belonging to Overall Tier 1–4 is suggested. These 35 vowels and consonant groups contain 45 (out of 73) words in total (words in bold in Tables 3–5) as some vowels or consonant groups have 2 words belonging to the same Tier, thus both are suggested. The only vowel for which both words belong to Overall Tier higher than 4 (5.5) is /əə/ (/pʰə̂əm/ 'add' and /pə̀ət/ 'open'). It is, therefore, *not suggested* or must be used cautiously for interactive speech audiometry. The 2 vowels with suggested words in Overall Tier 3 were /ɔ/ (/pɔ̀ŋ/ 'bulge') and /ia/ (/pìak/ 'wet' or /bìat/ 'jostle'). The 2 suggested words in Overall Tier 4 were /mêek/ 'cloud' (vowel /ee/) and /ŋaa/ 'ivory' (consonant /ŋ/). For vowel /i/, the word /pìt/ 'close' belongs to Overall Tier 6, hence /bin/ 'fly' (Overall Tier 1) is suggested.

Discussion

According to selection criteria, the proposed word list contains a total of 45 words from 35 vowels and consonants. Although the final word list is a subset of the original list, it is still phonemically balanced. The monophthongs representing all vowel qualities, namely tongue height, tongue advancement, and lip posture are suggested in the list. As all of the Thai diphthongs are similarly opening, the suggested diphthongs do represent the diphthong type in Thai. Moreover, the final word list includes consonants with all places and manners of articulation, adhering to the criterion. For these reasons, all 35 vowels and consonant

Table 6. Tier-1st and Tier-3/3. The levels were classified through corresponding values of $P_{m,1^{st}}$ and $P_{m,3/3}$. Values for both Tiers range from 1 (easiest, most certain) to 7 (hardest, least certain).

$P_{m,1^{st}}$,$P_{m,3/3}$(%)	Tier-1st, Tier-3/3
95–100	1
90–95	2
80–90	3
60–80	4
40–60	5
0–40	6

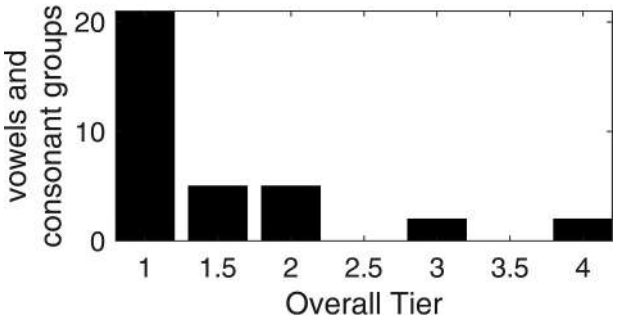

Figure 10. Number of Thai vowels and consonant groups with suggested words in each Overall Tier. A vowel or a consonant group can have more than 1 suggested word. For instance, for vowel /ua/ both / bua/ 'lotus' and /mùak/' hat' are in Overall Tier 1 and are both suggested.

groups can be used in interactive speech audiometry, with words belonging to Overall Tier 1 being the most preferable.

The ambiguous cases can be addressed as follows. The word /pêe/ 'rucksack' (Overall Tier 6, see Figure 2c) achieved an overall accuracy of only 62.67%, with 1.88 correct answers on average (95% CI 1.48–2.20), and was recognized 3 out of 3 times by only 24% of the subjects, even though it is a noun. This could be because, in the Thai language, the word is less common than /krapăw/ when referring to a rucksack.

In the first iteration of testing of the word /pʰêəm/ 'add' (Overall Tier 5.5, see Figure 4b, a verb) 32% of the subjects did not give an answer, and in the second iteration, 44% of the subjects associated the picture of the word /lɤ́ʔ/ 'dirty' with the sound of the word /pʰêəm/ 'add'. This could be due to several reasons: 1) the picture of the word /pʰêəm/'add' was ambiguous; 2) the word is a verb and hence was not straightforward to interpret or required complex interpretation to an extent; and 3) the picture of the word /lɤ́ʔ/ 'dirty' was somewhat abstract, causing confusion among the subjects.

A primary reason that the word /pɛ̀ɛt/ 'open' (Overall Tier 5.5, see Figure 11a; a verb or an adjective in Thai language) achieved an overall accuracy of only 52.17%, with 1.57 correct answers on average (95% CI 1.04–2.09), and that only 47.83% and 56.52% of the subjects correctly associated the sound of the word /pɛ̀ɛt/ 'open' with its corresponding picture in the second and third iterations, respectively, could be that the word was unintentionally tested against the picture of the word /pɨ̌n/ 'gun'. The sounds of the words /pɛ̀ɛt/ 'open' and /pɨ̌n/ 'gun' were rather difficult to differentiate in citation forms without the context, as the words have the same initial consonant /p/, the final consonants are produced with the same place of articulation (alveolar), and the vowels /ɛɛ/ and /ɨ̌/ are both central vowels with unrounded lips. The shared phonetic characteristics of the two words could trigger hesitation and confusion.

The rest of the subjects in the second (52.17%) and third (43.48%) iterations, those who did not correctly match the sound of the word /pɛ̀ɛt/ 'open' with its corresponding picture, associated the picture of the word /pɨ̌n/ 'gun' with the sound of the word /pɛ̀ɛt/ 'open'. It should also be noted that the word /pìt/ 'close' is a verb and that the picture representing

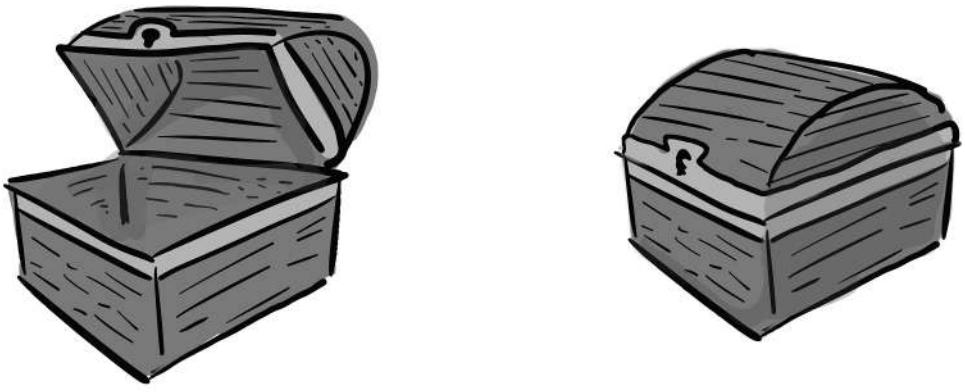

Figure 11. Two pictures representing two different actions or states. (a) /pɛ̀ɛt/ 'open'. (b) /pìt/ 'close'.

this word was an opened chest (see Figure 11a). This could have possibly caused the subjects to recognize the picture as a chest, a noun, instead.

The word /pìt/ 'close' (Overall Tier 6, see Figure 11b), a verb or an adjective in the Thai language), was inadvertently tested against 2 non-target pictures with the same initial consonant /p/, /pìak/ 'wet' in the first iteration and /pèt/ 'duck' in the second iteration. We hypothesize this was the primary cause of confusion for the subjects. In the first iteration, 33.33% of the subjects mistook the picture of the word /pìak/ 'wet' for the sound of the word /pìt/ 'close', while 29.63% correctly matched the sound of the word /pìt/ 'close' with its corresponding picture. In the second iteration, 70.37%% of the subjects mistook the picture of the word /pèt/ 'duck' for the sound of the word /pìt/ 'close', while only 25.93% correctly matched the sound of /pìt/ 'close' with its corresponding picture. This consequently led to its low overall accuracy of 39.51% and only 1.19 correct answers on average (95% CI 0.89–1.56). The confusion around the word /pìt/ and /pèt/ can be phonetically explained. The vowels /i/ and /e/, both front and unrounded vowels, are preceded and followed by identical consonants, /p_t/. Moreover, the tones of the two words are both falling pitch. Another reason could be that the word /pə̀ət/ 'open' is a verb (also an adjective in Thai); hence, the picture representing the word (see Figure 11b) was mistakenly recognized as a chest, a noun, instead.

Regarding the low amplitudes of the recorded words, due to the high SNRs of all the words (47.1 dB on average), playing back these words through the loudspeakers did not hinder the experiment by any means; the children were able to recognize almost all the words, except for some words, which was primarily due to the aforementioned graphical and linguistic reasons.

It should be noted that 38 out of 45 suggested words are nouns. This strongly suggests that nouns allow for straightforward recognition, require less interpretation of the pictures, and are, therefore, best suited for interactive speech audiometry in preschoolers.

Based on the low-accuracy cases, it is advisable to be aware of the way in which the audiometric testing is delivered for children. A target word should be a high-frequency word and denote a low level of abstractness. The phonetic characteristics of non-target words are also critical. Both target and non-target words should be minimal pairs or near minimal pairs in order to assess children's phoneme distinguishability and speech comprehension. The test should employ words for which the corresponding pictures are very familiar to children and can be instantly and easily recognized without any uncertainty so that they can focus solely on listening tasks, and upon hearing the sound of the test word, they can immediately associate the sound with the correct picture without any graphical hesitation.

Conclusions

We proposed a monosyllabic word and picture list designed to minimize the effect of phonetic environments. Both long and short monophthongs as well as diphthongs were included in the vowel word list, controlled with some places of articulation of the adjacent consonants. The consonant word list comprised the consonant groups which were based on manner of articulation, with the consonants occurring in onset position and followed by /aa/. The word and picture list was verified through both objective (TTC and TNC frequencies, and Zipf scores) and subjective listening measures in preschoolers with normal hearing as a pilot study towards the ultimate goal of using the verified word and picture list for interactive speech audiometry in

preschoolers. Three qualitative ranking systems referred to as Tier-1st, Tier-3/3, and Overall Tier were also introduced in order to evaluate the effectiveness of the designed word and picture list in the familiarity and recognizability test and to consequently obtain a final suggested word list comprising 45 words covering all vowels and consonant groups in the Thai language, except for the vowel /əə/. Vowels and consonant groups having 2 suggested words with the same Overall Tier value of 1 are /aa/, /uu/, /o/, /ua/, /p/, /p^h/, /ʔ/, /f/, /m/, /w/. One vowel, /ia/, had 2 suggested words with an Overall Tier value of 3.

The final word and picture list can be used in various forms of interactive speech audiometry, namely a PC-/online-based system with manual responses as well as a tablet-based system with automatic responses retrieved from brain signals. In future work, when using the final word list in interactive speech audiometry, one must be aware that assessment of ability to distinguish phonemes and comprehend speech is a crucially important aspect of the test and that words which are (near) minimal pairs should be tested against one another. Such applications for hearing-impaired children are yet to be validated.

Acknowledgments

The authors would like to thank the Kindergarten of Thammasat University under the patronage of Her Royal Highness Princess Galyani Vadhana for their help with this study.

ORCID

Apit Hemakom http://orcid.org/0000-0003-0143-6191
Pasin Israsena http://orcid.org/0000-0002-1393-4844

Declaration of interest statement

We declare we have no competing interest.

References

Amatayakul, P. (1968). *Introduction to audiology (hearing sciences)*. Mahidol University.

Aroonmanakun, W., Tansiri, K., & Nittayanuparp, P. (2009). *Thai national corpus: A progress report*. Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009 (pp. 153–160). Association for Computational Linguistics, Suntec, Singapore.

Chaix, B., & Lewis, R. (2016, September 20). *Speech audiometry*. Cochlea. Retrieved April 14, 2020, from http://www.cochlea.eu/en/audiometry/subjective-measure/speech-audiometry

Champaiboon, T., & Aroonmanakun, W. (2016). Complexity of nominalized constructions as text difficulty indicator: A case study on Thai Textbooks, 1st–3rd grade. *Humanities Journal*, *23*(2), 148–177. https://so04.tci-thaijo.org/index.php/abc/article/view/73908

Davidson, R., & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Economic Reviews*, *19*(1), 55–68. https://doi.org/10.1080/07474930008800459

Eisenberg, L. S., Martinez, A. S., & Boothroyd, A. (2007). Assessing auditory capabilities in young children. *International Journal of Pediatric Otorhinolaryngology*, *71*(9), 1339–1350. https://doi.org/10.1016/j.ijporl.2007.05.017

Kirk, K. I., Hay-McCutcheon, M., Sehgal, S. T., & Miyamoto, R. T. (2000). Speech perception in children with cochlear implants: Effects of lexical difficulty, talker variability, and word length. The Annals of Otology, Rhinology & Laryngology. *Supplement*, *185*, 79–81. https://doi.org/10.1177/00034894001095S1234

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech evoked potentials: From the laboratory to the clinic. *Ear and Hearing*, *29*(3), 285–313. https://doi.org/10.1097/AUD.0b013e3181662c0e

Morikawa, K., Kozuka, K., & Adachi, S. (2012). A proposed speech discrimination assessment methodology based on event-related potentials to visual stimuli. *International Journal of E-health and Medical Communications*, *3*(2), 19–35. https://doi.org/10.4018/jehmc.2012040102

Munthuli, A., Tantibundhit, C., Onsuwan, C., Kosawat, K., & Klangpornkun, N. (2014). *Thai phonetically balanced word recognition test: Test-retest reliability and error analysis*. Proceedings of the Fifteenth Australasian International Speech Science and Technology (SST), pp. 1–4. ASSTA, Christchurch, New Zealand.

National Association of the Deaf in Thailand (2018, October 31). *Report on disability situation in Thailand 2018*. Retrieved April 14, 2020, from https://nadt.or.th/en/stat61.html

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: A call to action. *Bulletin of the World Health Organization*, *92*(5), 367–373. https://doi.org/10.2471/BLT.13.128728

Poonyaban, S., Aungsakulchai, P., Tantibundhit, C., Onsuwan, C., Tiravanitchakul, R., Kosawat, K., & Munthuli, A. (2015). Phonetically balanced and psychometrically equivalent monosyllabic word lists for word recognition testing in Thai. *Acoustical Society of America*, *25*, 050003. https://doi.org/10.1121/2.0000496

Tanaka, C., Taniguchi, L. D., & Lew, H. L. (2018). *Geriatric rehabilitation*. Elsevier.

van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEXUK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Yimtae, K., Israsena, P., Thanawirattananit, P., Seesutas, S., Saibua, S., Kasemsiri, P., Noymai, A., & Soonrach, T. (2018). A tablet-based mobile hearing screening system for preschoolers: Design and validation study. *JMIR mHealth and uHealth*, *6*(10), e186. https://doi.org/10.2196/mhealth.9560