SEALS 21

Perception Study of Consonants in Thai

C. Tantibundhit[1], C. Onsuwan[2], S. Thatphithakkul[3], P. Chootrakool[3], N. Thatphithakkul[3], K. Kosawat[3], N. Saimai[1], and T. Saimai[1]

[1]MedIntelligence and Innovation Laboratory, Thammasat University

[2]Department of Linguistics, Thammasat University

[3]National Electronics and Computer Technology Center (NECTEC)

tchartur@engr.tu.ac.th, consuwan@tu.ac.th

Purpose

- The purpose of this study is to develop an intelligibility testing of Thai speech based on the (English) diagnostic rhyme test (DRT) (Voiers, 1983). The Thai diagnostic rhyme test (TDRT) consists of 2 test sets: initials (21 phonemes) and final consonants (8 phonemes).
- The TDRT is an adaptation of several useful frameworks: DRT, new Chinese diagnostic rhyme test (NCDRT) (McLoughlin, 2008), and the analysis method of balanced confusion matrix (e.g., Miller and Nicely, 1955).
- TDRT's advantages are twofold. TDRT allows us to systematically evaluate percent intelligibility responses in each stimulus pair and to effectively obtain confusion matrices.

- Confusion patterns provide important information for the understanding of how speech signals are auditorily processed and transformed (as some parts of the signals will become more distinct while others suppressed (Stevens, 1981)).

- This insight is crucial for a number of areas in speech research, including speech recognition.

- To date, a very small number of studies on Thai have investigated confusion of Thai speech sounds, including tones (e.g., Gandour and Dardarananda (1983), Thubthong (2001)). This aspect of Thai consonants is not well understood.

Test for Initial Consonants

Speech materials:

- 21 monosyllabic rhyming words (CVV), each of which differs only in their initial consonant (commonly used words)
- filler words

Speaker: native male speaker of Thai

Recording procedure:

- Words were embedded in a sentence context and read 5 times.
- One token for each target word was selected based on impressionistic hearing evaluation and spectrographic inspection.

no.	transcription	Thai script	translation
1.	/pāa/	ปา	throw
2.	/p^{h}āa/	พา	bring
3.	/bāa/	บา	teacher
4.	/tāa/	ตา	eye
5.	/t^{h}āa/	ทา	paint
6.	/dāa/	ดา	advance along a wide front
7.	/tɕāa/	จา	talk
8.	/tɕ^{h}āa/	ชา	tea
9.	/kāa/	กา	crow
10.	/k^{h}āa/	คา	stick
11.	/ʔāa/	อา	uncle
12.	/fāa/	ฟา	F musical note
13.	/sāa/	ซา	lessen
14.	/hāa/	ฮา	laugh
15.	/māa/	มา	arrive
16.	/nāa/	นา	field
17.	/ŋāa/	งา	ivory
18.	/lāa/	ลา	donkey
19.	/rāa/	รา	fungus
20.	/wāa/	วา	2 meters (Thai unit)
21.	/jāa/	ยา	medicine

Table 1 Twenty one rhyming words differing in their initial consonants across 21 phonemes.

Complete word list: http://charturong.ece.engr.tu.ac.th/SEALS21/Initials.pdf

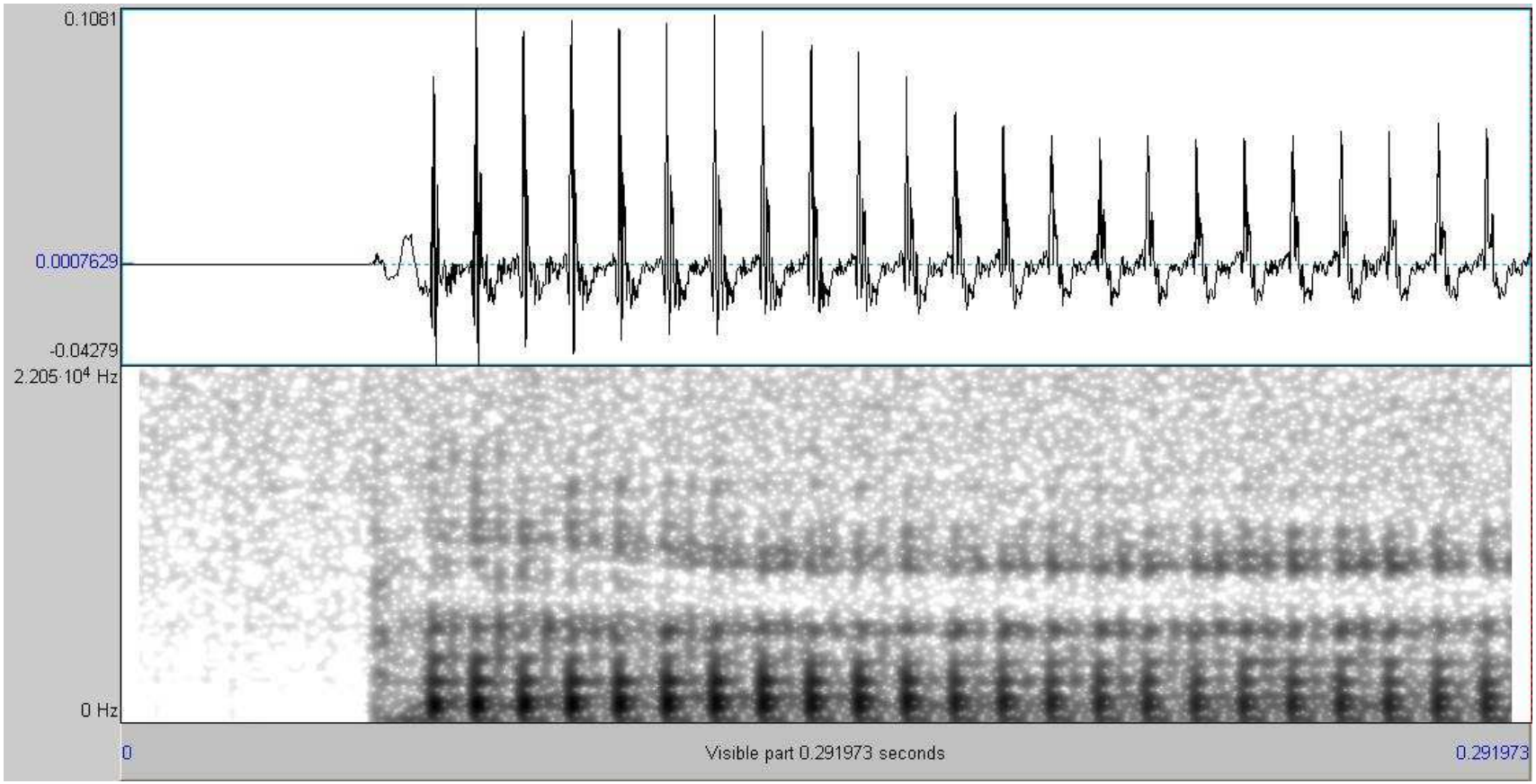
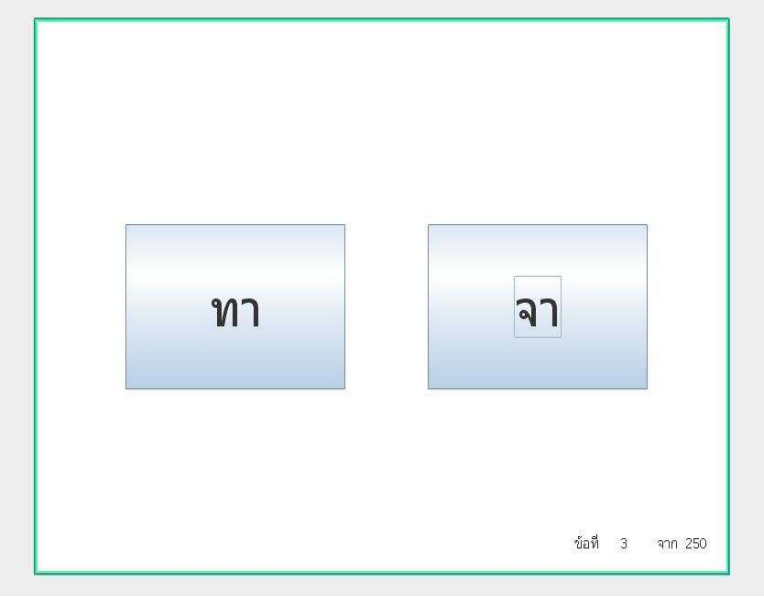


Figure 1 Waveform display (top panel) and wide-band spectrogram (bottom panel) of the token /pāa/ (without added noise).

Testing procedure:

- There were 420 trials for initial consonants and 80 trials for filler words. In each trial, a target stimulus was randomly presented and visually appeared with another rhyming word (20 others differing in initial phoneme). Listeners were asked to choose what they heard between the 2 rhyming words (A/B forced choice).
- All trials were corrupted by each of 4 signal-to-noise ratio (SNR) levels of additive white Gaussian (AWG) noise: −6, −12, −18, and −24dB.



Participants: 28 Thais

Peceptual Results for Initials

Initial Consonant	SNR (dB)			
	-6dB	-12dB	-18dB	-24dB
P_e	93.06%	87.14%	77.35%	24.08%

Table 2 Average percent intelligibility for initial consonants.

$$P_e = \frac{N_r - N_w}{T} \times 100\% \ ,$$

where *Pe*, *Nr*, *Nw*, and *T* are percent intelligibility score, numbers of correct responses, numbers of wrong responses, and total numbers to stimuli, respectively (Voiers, 1983).

- Percent intelligibility scores are decreasing as increasing level of noise. The SNR level of −18dB is the most interpretable (subject's performance at SNR levels of −6dB and −12dB is near-perfect, at −24dB it is near-chance).

Stimulus	Response																				
	/p/	/p^h/	/b/	/t/	/t^h/	/d/	/ʨ/	/$ʨ^h$/	/k/	/k^h/	/ʔ/	/f/	/s/	/h/	/m/	/n/	/ŋ/	/l/	/r/	/w/	/j/
/p/	133	0	0	4	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
/p^h/	1	121	0	1	6	0	1	4	0	2	1	0	1	2	0	0	0	0	0	0	0
/b/	0	0	137	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
/t/	0	3	2	116	2	0	4	1	3	0	0	5	3	0	0	0	0	0	1	0	0
/t^h/	4	3	1	2	112	1	3	4	2	2	0	2	3	1	0	0	0	0	0	0	0
/d/	1	0	0	0	0	132	3	0	4	0	0	0	0	0	0	0	0	0	0	0	0
/ʨ/	1	0	0	4	1	1	115	5	1	0	0	3	5	0	0	1	0	1	0	1	1
/$ʨ^h$/	2	2	0	1	2	1	1	126	0	2	1	0	1	0	0	0	0	1	0	0	0
/k/	1	0	0	3	0	0	2	0	130	0	0	0	1	0	0	0	1	1	0	0	1
/k^h/	0	3	1	2	6	0	1	2	1	117	5	0	0	2	0	0	0	0	0	0	0
/ʔ/	0	1	0	0	1	0	0	0	1	1	128	0	0	4	1	1	2	0	0	0	0
/f/	5	0	1	2	1	2	0	0	0	0	0	126	1	0	0	0	0	1	0	1	0
/s/	5	0	1	4	1	1	1	1	1	1	1	5	116	0	0	0	0	1	1	0	0
/h/	1	4	0	0	2	0	0	0	0	2	4	0	0	123	1	0	2	1	0	0	0
/m/	0	0	0	0	0	0	0	1	1	0	0	0	0	1	130	2	3	0	1	1	0
/n/	1	0	0	1	0	0	0	0	0	1	3	1	1	4	2	122	2	0	2	0	0
/ŋ/	0	0	0	0	0	1	0	1	0	0	3	0	0	1	1	3	129	1	0	0	0
/l/	1	0	0	0	0	1	2	2	1	1	0	0	0	0	1	1	1	124	1	2	2
/r/	1	1	4	3	1	5	5	3	6	2	0	1	2	1	1	1	4	4	91	0	4
/w/	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	139	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	140

Table 3 Confusion matrix for initial consonants at SNR -18dB.

Result patterns:

- At the SNR level of −18dB, /r/is the most confusable initial consonant and it is mostly misperceived as/k/,/d/, and/tɕ/. The least confusable consonants are/j/and/w/.

- At the SNR level of −18dB, investigation of listeners' misidentified responses reveals that the listeners favor/t/ and disfavor/w/and/r/over other consonants.

Test for Final Consonants

Speech materials: • 84 monosyllabic pairs of ($C_iV(V)C_f$), each of which differs only in their final consonant phoneme and the tone in each pair is identical

C_i = /p/,/t/,/k/,/p^h/,/t^h/or/k^h/

V = 9 short vowels VV = 9 long vowels

C_f = /p/,/t/,/k/,/m/,/n/, /ŋ/,/w/or/j/

• pairs of filler words
(all are commonly used words)

Speaker: native male speaker of Thai (who took part in the initial set)

Recording procedure:

- Words were embedded in a sentence context and read 5 times.
- One token for each target word was selected based on impressionistic hearing evaluation and spectrographic inspection.

Table 4 (1)
84 pairs of rhyming words differing in their final consonants across 8 phonemes.

pair no.	transcription	Thai script	translation	-	transcription	Thai script	translation
1.	/tàp/	ตับ	liver	-	/tàt/	ตัด	cut
2.	/kòp/	กบ	frog	-	/kòt/	กด	press
3.	/pɔ̀ɔp/	ปอบ	ogre	-	/pɔ̀ɔt/	ปอด	lung
4.	/tòp/	ตบ	slap	-	/tòk/	ตก	fall
5.	/kàp/	กับ	and	-	/kàk/	กัก	confine
6.	/tɔ̀ɔp/	ตอบ	answer	-	/tɔ̀ɔk/	ตอก	hammer
7.	/t^{h}úp/	ทุบ	pond	-	/t^{h}úm/	ทุ้ม	bass
8.	/k^{h}âap/	คาบ	hold in the mouth	-	/k^{h}âam/	ข้าม	skip
9.	/k^{h}áp/	คับ	tight	-	/k^{h}ám/	ค้ำ	prop up
10.	/p^{h}óp/	พบ	meet	-	/p^{h}ón/	พ้น	pass
11.	/k^{h}óp/	คบ	associate	-	/k^{h}ón/	ค้น	seek
12.	/k^{h}áp/	คับ	tight	-	/k^{h}án/	คั้น	squeeze
13.	/kèp/	เก็บ	keep	-	/kèŋ/	เก่ง	excellently
14.	/k^{h}âap/	คาบ	hold in the mouth	-	/k^{h}âaŋ/	ข้าง	side
15.	/tàp/	ตับ	liver	-	/tàŋ/	ตั่ง	stool
16.	/tàp/	ตับ	liver	-	/tàw/	เต่า	turtle
17.	/t^{h}áp/	ทับ	overlay	-	/t^{h}áw/	เท้า	foot
18.	/k^{h}áp/	คับ	tight	-	/k^{h}áw/	เค้า	outline
19.	/p^{h}âap/	ภาพ	picture	-	/p^{h}âaj/	พ่าย	lose
20.	/kàp/	กับ	and	-	/kàj/	ไก่	chicken
21.	/k^{h}âap/	คาบ	hold in the mouth	-	/k^{h}âaj/	ค่าย	camp
22.	/pàt/	ปัด	sweep	-	/pàk/	ปัก	stab down
23.	/pàat/	ปาด	slice off	-	/pàak/	ปาก	mouth
24.	/tàt/	ตัด	cut	-	/tàk/	ตัก	scoop
25.	/k^{h}òt/	ขด	coil	-	/k^{h}òm/	ข่ม	oppress
26.	/k^{h}át/	คัด	select	-	/k^{h}ám/	ค้ำ	prop up
27.	/k^{h}út/	ดุด	curl	-	/k^{h}úm/	ดุ้ม	protect
28.	/kɔ̀ɔt/	กอด	hug	-	/kɔ̀ɔn/	ก่อน	before
29.	/k^{h}ùt/	ขุด	dig	-	/k^{h}ùn/	ขุ่น	be turbid
30.	/t^{h}àat/	ถาด	tray	-	/t^{h}àan/	ถ่าน	charcoal
31.	/pèt/	เป็ด	duck	-	/pèŋ/	เป่ง	be ripe
32.	/tìt/	ติด	close	-	/tìŋ/	ติ่ง	protrusion
33.	/t^{h}ít/	ทิด	man who resumes secular life	-	/t^{h}íŋ/	ทิ้ง	discard
34.	/pàt/	ปัด	sweep	-	/pàw/	เป่า	blow
35.	/k^{h}àt/	ขัด	rub	-	/k^{h}àw/	เข่า	knee
36.	/k^{h}ít/	คิด	think	-	/k^{h}íw/	คิ้ว	eyebrow
37.	/k^{h}âat/	คาด	anticipate	-	/k^{h}âaj/	ค่าย	camp
38.	/kàt/	กัด	bite	-	/kàj/	ไก่	chicken
39.	/t^{h}àat/	ถาด	tray	-	/t^{h}àaj/	ถ่าย	take a picture
40.	/t^{h}úk/	ทุกข์	suffering	-	/t^{h}úm/	ทุ้ม	bass
41.	/k^{h}úk/	ดุก	prison	-	/k^{h}úm/	ดุ้ม	protect
42.	/t^{h}òk/	ถก	discuss	-	/t^{h}òm/	ถ่ม	spit

Table 4 (2)
Complete word list:
http://charturong.ece.engr.tu.ac.th/SEALS21/Finals.pdf

pair no.	transcription	Thai script	translation	-	transcription	Thai script	translation
43.	/pʰók/	พก	carry	-	/pʰón/	พ้น	pass
44.	/kʰôok/	โคก	mound	-	/kʰôon/	โค่น	fall
45.	/pòk/	ปก	cover	-	/pòn/	ป่น	powdered
46.	/tàak/	ตาก	air	-	/tàaŋ/	ต่าง	differ
47.	/kʰòok/	โขก	knock	-	/kʰòoŋ/	โข่ง	Pila (gastropod)
48.	/kʰàak/	ขาก	spit	-	/kʰàaŋ/	ข่าง	spinning top
49.	/pʰàk/	ผัก	vegetable	-	/pʰàw/	เผ่า	tribe
50.	/pʰàak/	ผาก	parched	-	/pʰàaw/	ผ่าว	scorching
51.	/pàk/	ปัก	stick	-	/pàw/	เป่า	blow
52.	/kúk/	กุ๊ก	cook	-	/kúj/	กุ๊ย	thug
53.	/tàak/	ตาก	air	-	/tàaj/	ต่าย	rabbit
54.	/kàak/	กาก	garbage	-	/kàaj/	ก่าย	rest on
55.	/tùm/	ตุ่ม	pimple	-	/tùn/	ตุ่น	mole
56.	/kâam/	ก้าม	claw	-	/kâan/	ก้าน	stem
57.	/tām/	ตำ	pound	-	/tān/	ตัน	clog
58.	/tɔ̄ɔm/	ตอม	swarm	-	/tɔ̄ɔŋ/	ตอง	banana leaf
59.	/pɔ̂m/	ป้อม	fortress	-	/pɔ̂ŋ/	ป้อง	cover up
60.	/tēm/	เต็ม	full	-	/tēŋ/	เต็ง	favorite
61.	/kām/	กำ	grasp	-	/kāw/	เกา	scratch
62.	/tām/	ตำ	pound	-	/tāw/	เตา	stove
63.	/kâam/	ก้าม	claw	-	/kâaw/	ก้าว	step
64.	/tɤ̄ɤm/	เติม	add	-	/tɤ̄ɤj/	เตย	screw pine
65.	/tāam/	ตาม	follow	-	/tāaj/	ตาย	die
66.	/pāam/	ปาล์ม	palm	-	/pāaj/	ปาย	Pai district
67.	/kɛ̀n/	แก่น	core	-	/kɛ̀ŋ/	แก่ง	islet
68.	/kōon/	โกน	shave	-	/kōoŋ/	โกง	cheat
69.	/tɛ̄ɛn/	แตน	wasp	-	/tɛ̄ɛŋ/	แตง	melon
70.	/kān/	กัน	keep out	-	/kāw/	เกา	scratch
71.	/pân/	ปั้น	mold	-	/pâw/	เป้า	target
72.	/tān/	ตัน	clog	-	/tāw/	เตา	stove
73.	/pɔ̄ɔn/	ปอน	sloppy	-	/pɔ̄ɔj/	ปอย	tuft
74.	/tùn/	ตุ่น	mole	-	/tùj/	ตุ่ย	puffy
75.	/pâan/	ป้าน	obtuse	-	/pâaj/	ป้าย	plate
76.	/pʰêŋ/	แพ่ง	civil	-	/pʰêw/	แผ้ว	clear
77.	/tī̂ŋ/	ติ่ง	admonish	-	/tī̂w/	ติ่ว	cram for an examination
78.	/tâŋ/	ตั้ง	establish	-	/tâw/	เต้า	breast
79.	/kɔ̂ŋ/	ก้อง	echo	-	/kɔ̂j/	ก้อย	little finger
80.	/tàaŋ/	ต่าง	differ	-	/tàaj/	ต่าย	rabbit
81.	/kōoŋ/	โกง	cheat	-	/kōoj/	โกย	shovel
82.	/kàw/	เก่า	old	-	/kàj/	ไก่	chicken
83.	/kʰâaw/	ข้าว	rice	-	/kʰâaj/	ค่าย	camp
84.	/kʰāaw/	คาว	fishy	-	/kʰāaj/	คาย	spit out

Testing procedure:

- There were 168 trials for final consonants and 32 trials for filler words. In each trial, a target stimulus was randomly presented and visually appeared with other word in the pair. Listeners were asked to choose what they heard between the 2 rhyming words (A/B forced choice).

- All trials were corrupted by each of 4 signal-to-noise ratio (SNR) levels of additive white Gaussian (AWG) noise: −6, −12, −18, and −24dB.

Participants: 28 Thais (who also took part in the initial set)

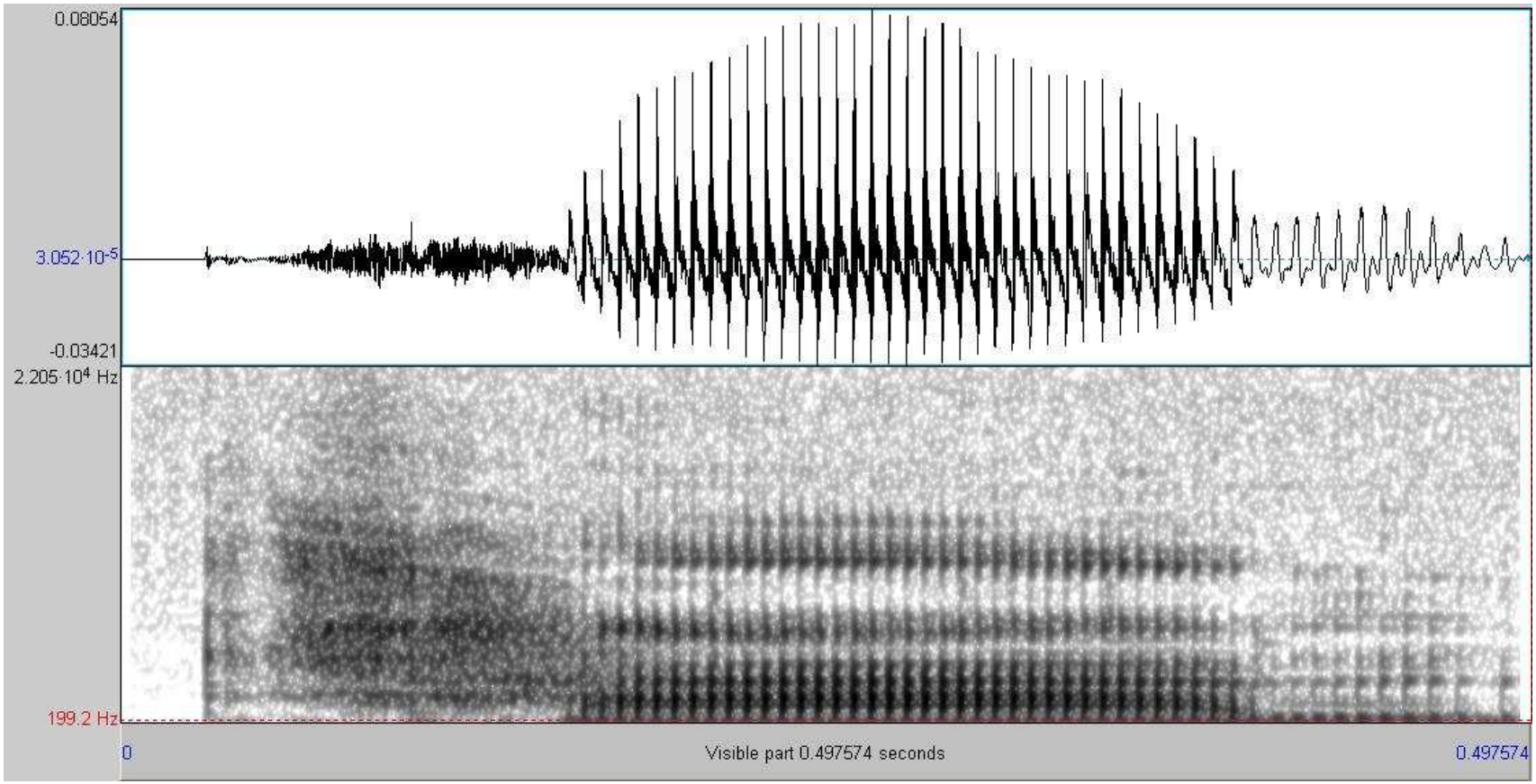


Figure 2 Waveform display (top panel) and wide-band spectrogram (bottom panel) of the token /kʰâam/(without added noise).

Peceptual Results for Finals

Final Consonant	SNR (dB)			
	-6dB	-12dB	-18dB	-24dB
P_e	91.67%	84.01%	67.35%	27.21%

Table 5 Average percent intelligibility for final consonants.

- Similar to initial consonants, percent intelligibility scores are decreasing as increasing level of noise. The SNR level of −18dB is the most interpretable.

Stimulus	Response							
	/p/	/t/	/k/	/m/	/n/	/ŋ/	/w/	/j/
/p/	143	1	2	0	0	0	0	1
/t/	3	141	2	0	1	0	0	0
/k/	1	8	133	0	3	1	0	1
/m/	0	0	1	143	1	1	1	0
/n/	0	2	0	0	136	6	0	3
/ŋ/	0	1	0	3	2	140	0	1
/w/	0	0	0	1	0	0	146	0
/j/	1	0	0	0	1	0	0	145

Table 6 Confusion matrix for final consonants at SNR -6dB.

Stimulus	Response							
	/p/	/t/	/k/	/m/	/n/	/ŋ/	/w/	/j/
/p/	139	3	4	0	0	0	0	1
/t/	7	126	8	0	1	1	0	4
/k/	6	12	117	2	1	5	0	4
/m/	0	0	0	141	1	1	0	4
/n/	0	5	0	2	139	1	0	0
/ŋ/	0	0	3	5	5	131	0	3
/w/	0	1	0	0	0	1	145	0
/j/	0	1	0	0	0	2	0	144

Table 7 Confusion matrix for final consonants at SNR -12dB.

Stimulus	Response							
	/p/	/t/	/k/	/m/	/n/	/ŋ/	/w/	/j/
/p/	119	4	14	4	0	2	0	4
/t/	11	118	4	0	7	0	1	6
/k/	6	12	117	0	3	6	0	3
/m/	1	0	1	119	6	11	3	6
/n/	0	9	0	10	117	3	0	8
/ŋ/	0	1	7	4	3	132	0	0
/w/	0	0	0	3	2	0	142	0
/j/	0	0	1	6	11	9	0	120

Table 8 Confusion matrix for final consonants at SNR -18dB.

Stimulus	Response							
	/p/	/t/	/k/	/m/	/n/	/ŋ/	/w/	/j/
/p/	108	9	12	4	0	5	0	9
/t/	18	95	11	4	6	4	2	7
/k/	12	12	92	4	7	10	4	6
/m/	8	4	9	94	7	10	4	11
/n/	3	14	5	11	89	15	2	8
/ŋ/	6	5	11	11	9	90	5	10
/w/	7	8	11	9	14	7	79	12
/j/	5	8	5	9	8	11	0	101

Table 9 Confusion matrix for final consonants at SNR -24dB.

Result patterns:

- At the SNR level of −18dB, /k/and /n/are the most confusable final consonants. /k/is mostly misperceived as /t/and/n/as/m/. The least confusable consonant is/w/.

- At the SNR level of −18dB, investigation of listeners' misidentified responses reveals that the listeners favor/n/ and/ŋ/and disfavor/w/over other consonants.

Conclusions

- Listeners' performance is lower in the final consonant than in the initial set, with listeners' percent intelligibility scores increase with SNR levels.
- Preliminary analysis, at the level of −18dB, shows that place of articulation errors predominate among the initial and final consonants and that voicing is little affected by noise. (in line with reports of English consonants (Miller and Nicely, 1955; Benkí, 2003)).
- The biases of listeners' misidentified responses could be explained in light of frequency of phoneme occurrences found in a Thai BEST corpus (approximately 9 million words; each word was annotated for its pronunciation):

- of the 21 initial consonants, /w/ is among consonants of lowest occurrence, which include/$tɕ^h$/,/h/,/ʔ/,/b/,/ŋ/ ,and/f/.

- of the 8 final consonants, /n/ has the highest occurrences while/w/the lowest.

References

Benkí, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica* 60, 129-157.

Gandour, J. and Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language* 18(1), 98-114.

Human Language Technology Laboratory, BEST. *http://www.hlt.nectec.or.th/best/*

McLoughlin, I.. (2008) Subjective intelligibility testing of Chinese speech. *IEEE Transactions on Audio, Speech and Language Processing* 16(1), 23-33.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338-352.

Stevens, K. N. (1981). Contraints imposed by the auditory system on the properties used to classify speech sounds: data from phonology, acoustics, and psychoacoustics. In T. Myers, J. Laver, and J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 61-74). Amsterdam: North -Holland.

Thubthong, N. (2001). *A study of various linguistic effects on tone recognition in Thai continuous speech.* Doctoral dissertation, Chulalongkorn University, Bangkok.

Voiers, W. D. (1983). Evaluating processed speech using the diagnostic rhyme test. *Speech Technology* 1, 30-39.