Malaysian Association
of Applied Linguistics

Certificate of Appreciation

The certificate is presented to:

Wan Aslynn Salwani Wan Ahmad

in recognition of participation as a Presenter at the

Malaysian Association of Applied Linguistics Asia Pacific Conference

(MAAL APC 2022)

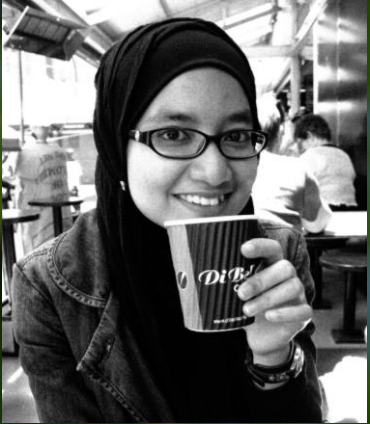*The Role of Languages in the Pursuit of Global Sustainability*

10-11 August 2022 • KLCC Convention Center, Kuala Lumpur, Malaysia.

Prof. Dr. Shameem Rafik-Galea
MAAL APC 2022 Conference Chair
President, Malaysian Association of Applied Linguistics (MAAL)

DEVELOPING A SPOKEN MALAY CORPUS FOR PAEDIATRIC MALAY MATRIX SENTENCE TEST (PaedMalayMST) MATERIAL

Nur 'Azzah Zakaria
azzah@iium.edu.my

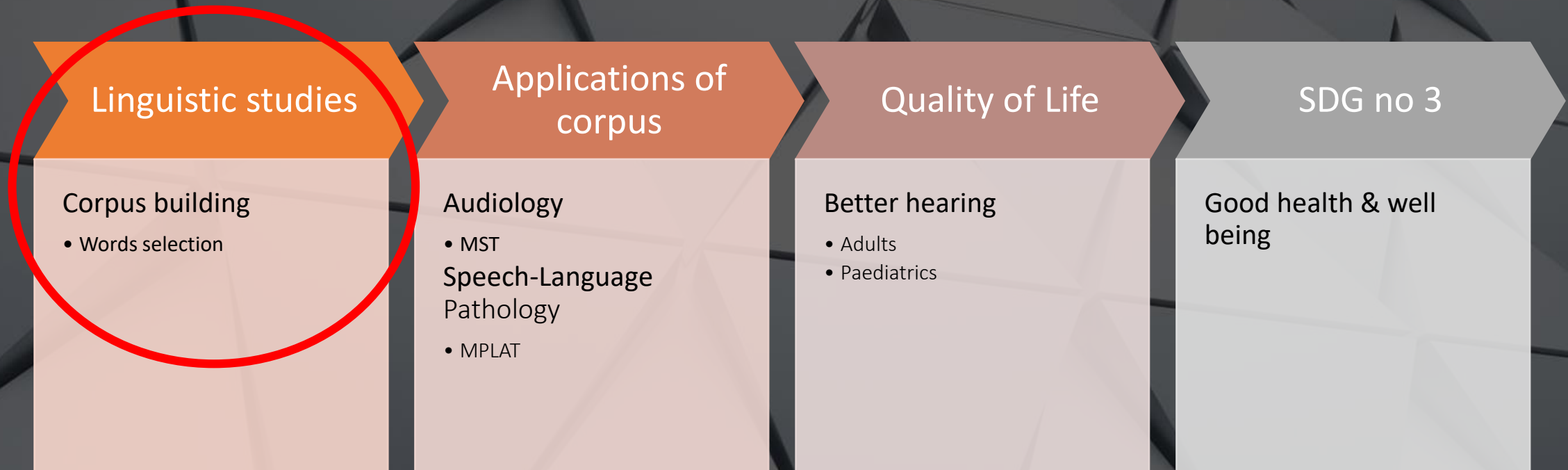Wan Aslynn Wan Ahmad, PhD
wanaslynn@iium.edu.my

Department of Audiology & Speech-Language Pathology
Kulliyyah of Allied Health Sciences
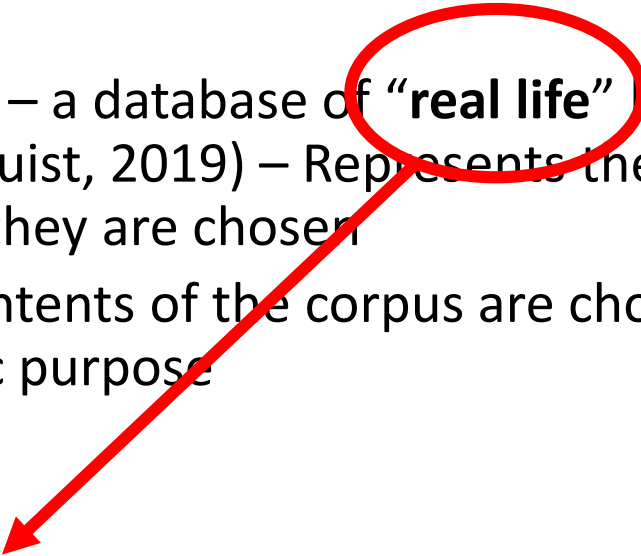International Islamic University Malaysia

Overview

- The big picture
- A corpus
- Matrix sentence test
- Objectives of the study
- Methodology
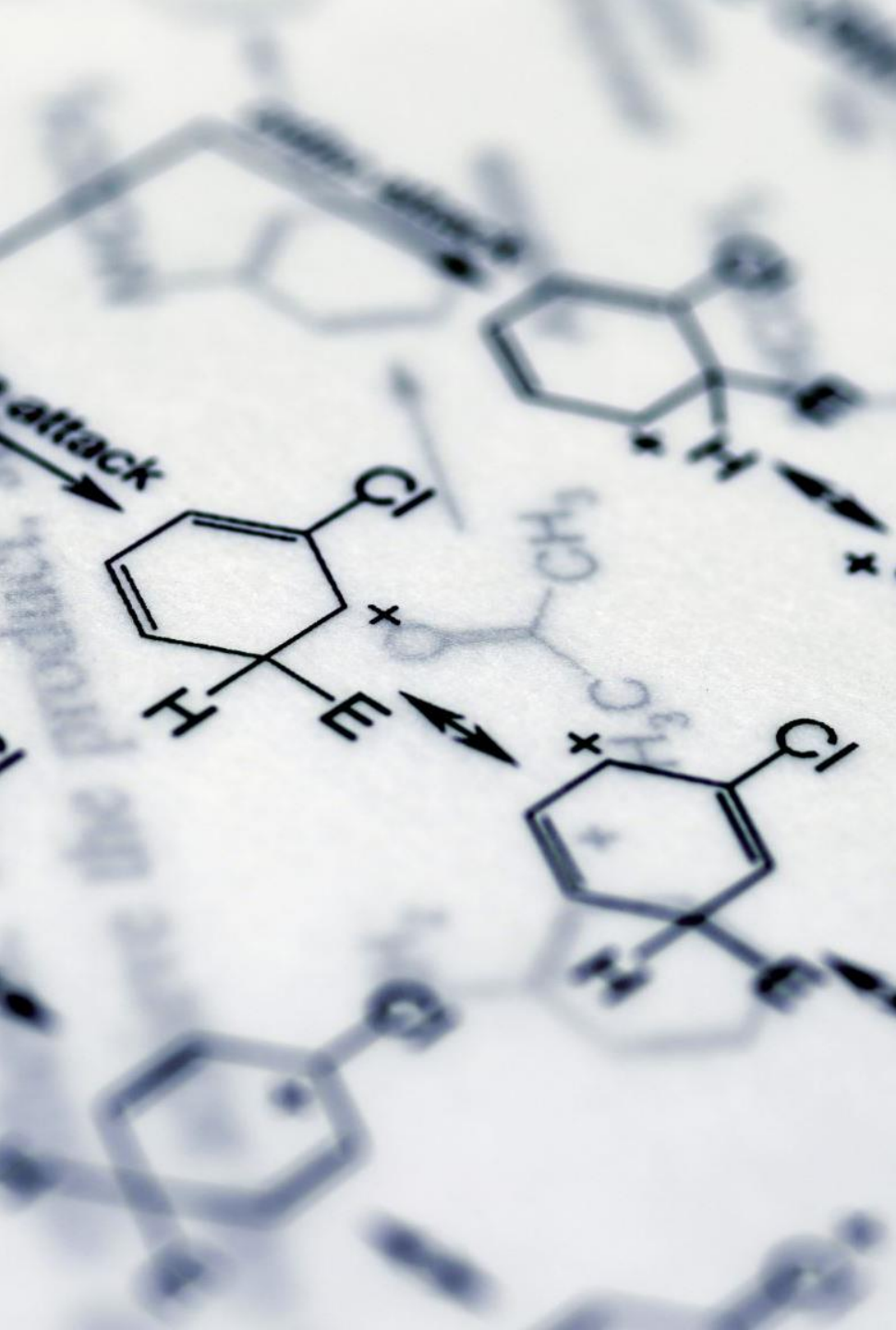- Results
- Discussion
- Conclusion

The BIG picture

Linguistic studies

Corpus building
- Words selection

Applications of corpus

Audiology
- MST

Speech-Language Pathology
- MPLAT

Quality of Life

Better hearing
- Adults
- Paediatrics

SDG no 3

Good health & well being

What is a corpus?

- Definition:
  - Corpus – a database of "**real life**" language (Nordquist, 2019) – Represents the language from which they are chosen
  - The contents of the corpus are chosen to support a specific purpose

  - Written

  - Spoken

  - Genre
  - Varieties of the language

Examples of corpora

https://dgd.ids-mannheim.de/DGD2Web/jsp/Welcome.jsp

http://www.natcorp.ox.ac.uk

Overview of The Corpus of Contemporary American English(COCA)

Genre	# texts	# words	Explanation
Spoken	44,803	127,396,932	Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Oprah)
Fiction	25,992	119,505,305	Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books 1990-present, and fan fiction.
Magazines	86,292	127,352,030	Nearly 100 different magazines, with a good mix between specific domains like news, health, home and gardening, women, financial, religion, sports, etc.
Newspapers	90,243	122,958,016	Newspapers from across the US, including: USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. Good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.
Academic	26,137	120,988,361	More than 200 different peer-reviewed journals. These cover the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business
Web (Genl)	88,989	129,899,427	Classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages (by Serge Sharoff). Taken from the US portion of the GloWbE corpus.
Web (Blog)	98,748	125,496,216	Texts that were classified by Google as being blogs. Further classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages. Taken from the US portion of the GloWbE corpus.
TV/Movies	23,975	129,293,467	Subtitles from OpenSubtitles.org, and later the TV and Movies corpora. Studies have shown that the language from these shows and movies is even more colloquial / core than the data in actual "spoken corpora".
	485,179	1,002,889,754	

What about Malay?

- Accessible public corpora for Malay – Korpus DBP & MyBaca however the words in these corpora are collections from written texts ONLY
- To date, there is **no** known **spoken corpus** for Malay

http://www.mybaca.org

http://sbmb.dbp.gov.my/korpusdbp/SelectUserCat.aspx

MATRIX SENTENCE TEST (MST)

It is a speech recognition tests with sentence-based test material

The sentences are generated by a 50-word base matrix

The sentence has a fixed grammatical structure but semantically unpredictable

Different word lists for adult and paediatric populations

MST for both populations are available in languages like Russian, German and Polish,

For Malay, only the list for adult population is available (Jamaluddin, 2016)

What is MST for?

-To **assess** and **diagnose** peripheral and central hearing impairments;

-To **verify** and **monitor** the rehabilitation outcomes of amplification and;

-For **research** applications

Malay MST

- An example of the matrix for adults

- These words can be randomly combined to create 100 000 sentences (10^5).

- During the test, patients listen to these sentences (varied intensity) and repeat them to determine their hearing level.

SUBJECT	VERB	NUMBER	OBJECT	ADJECTIVES
Saya I	bagi give(s)	satu one	bola ball	baru new
Kita We	ada have/has	dua two	buku book	besar big
Dia He or she	dapat receive(s)	banyak many	baju shirt	lama old
Kami Us	perlu need(s)	semua all	lampu lamp	kecil small
Ibu Mom	beri give(s)	tiga three	meja table	merah red
Abang Brother	ambil take(s)	empat four	kotak box	hitam black
Ayah Dad	mahu want(s)	lima five	kunci key	putih white
Adik Younger sibling	suka like(s)	enam six	pisau knife	hijau green
Kakak Sister	nampak see(s)	tujuh seven	mangkuk bowl	mahal expensive
Nenek Grandfather	minta asked for	lapan eight	topi hat	cantik beautiful

Objectives of this study

To produce a spoken Malay corpus for paediatric population

To identify spoken Malay words suitable for PaedMalayMST from this new corpus

Methodology

OBJECTIVE 1: To produce a database for spoken Malay corpus suitable for paediatric population

OBJECTIVE 2: To identify spoken Malay words for PaedMalayMST

| Identify sources | Transcription of dialogue into text | Analysis of compiled text | Transcribe the words phonemically | Identify most frequent 1000 words | Selection for PaedMalayMST |

Popular Malay animated movies and television series

Two-stage transcription by 2 researchers

Ranked the words based on frequency of occurrence

- A Malay root word
- Disyllabic word
- Words from these categories: **number**, **object** and **adjective**
- Words that can be illustrated

The sources

Movies

- *Geng: Pengembaraan Bermula*
- *Boboiboy: The Movie*

TV series

- *Upin dan Ipin*
- *Ejen Ali*

RESULTS

Objective 1

To produce a database for spoken Malay corpus suitable for paediatric population

A total of **1900 words** were catalogued

Words were ranked **according to the frequency of** occurrence

Transcription of compiled words

- Analysed Phonetic and phonemic distribution

Malay	English tr		Bilabial			Labio-dental		Alveolar							Post-alveolar			Palatal		
			Nasal	Plosive	Plosive	Fricative	Fricative	Nasal	Plosive	Plosive	Fricative	Fricative	Approximant	Lateral	Affricate	Affricate	Fricative	Nasal	Approximant	Nasal
			m	p	b	f	v	n	t	d	s	z	r	l	tʃ	dʒ	ʃ	ɲ	j	ŋ
355 siang		sɪjeŋ									1								1	
356 sorang		soʀeŋ									1		1							
357 sorry		sorɪ									1		1							
358 suka		sukɐ									1									
359 susah		suseh									2									
360 tentang		tənteŋ						1	2											
361 tinggal		tɪŋgel							1					1						
362 tinggalkan		tɪŋgelken						1	1					1						
363 tipu		tɪpu		1					1											
364 trex		tɪrɛks							1		1		1							
365 tumbukan		tumbuken	1		1			1	1											
366 uh		uh																		
367 yeay		jɛej																	2	
368 adudu		ɐdudu								2										
369 aiyoyo		ɐɪjojo																	2	
370 appa		ɐppɐ		2																
371 atuk		ɐtoʔ							1											
372 ayuh		ɐjuh																	1	
373 bahaya		bɐhejɐ			1														1	
374 bang		beŋ			1															
375 belakang		bəlɐkeŋ			1									1						
376 belum		bəlum	1		1									1						
377 bersedia		bəsədɪɐ			1					1	1								1	
378 damai		dɐmeʀ	1							1										

The distribution of Malay phonemes
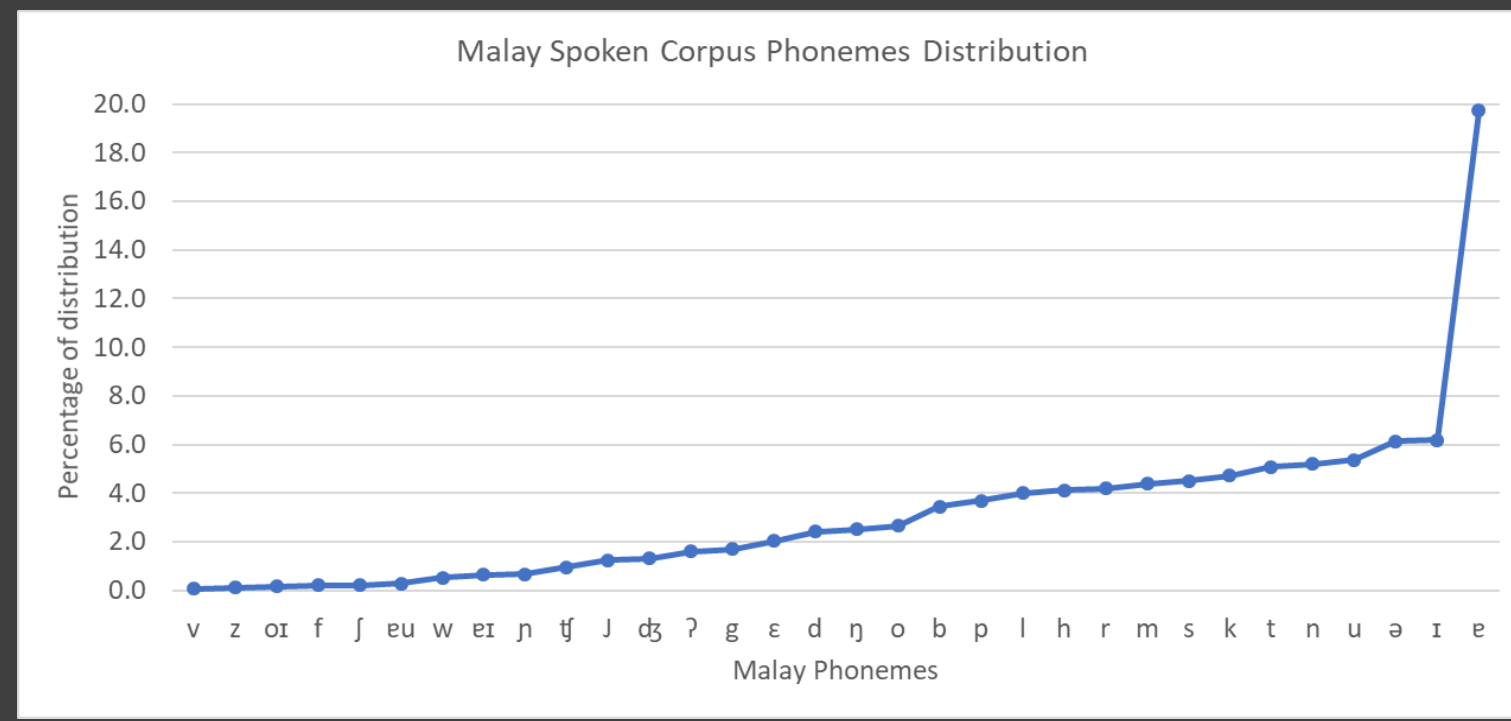
To ensure the frequency of phonemes in the corpus is represented equally in the PaedMalay MST

Objective 2
To identify spoken Malay words for PaedMalayMST

Inclusion criteria	Categories		
	Number	Object	Adjectives
First 1000 words	7	58	85
Disyllabic words	4	48	70
Words that were difficult to illustrate	0	2	69
Final words selected	4	46	1

Examples of the illustrations of words selected from the corpus

Number	Object	Adjective
1		
2		
3		
5		
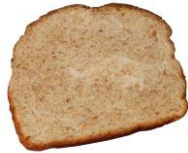
Discussion

1. Linguistic scenario

- More mix-code
- Frequently used phonemes
- Fluidity of language
- Shed new lights on the study of Malay
- Contribute to the literature of Malay linguistics

2. The PaedMalayMST

- First step for the purpose of this test
- Selection of words
- The use of illustration
- Comfortability
- Focus on the hearing test- for children with speech impairment

3. Good health and wellbeing (SDG 3)

- Appropriate management and rehabilitation

- Late detection and intervention for paediatric patients may lead to cognitive issues which may hamper their learning

- If the children with hearing impairment are given proper rehabilitation, it will improve their speech language, reading and cognitive ability. (Mendel, 2008)

- Better quality of life leads to good health and well being

Recommendation

This corpus can benefit many fields, not only limited to clinical use

Education-language can make use of the words compiled for classroom teaching

Psychology

The possibility is endless!

Conclusion

Familiarity of words among Malay speaking paediatrics

Words extracted were analysed further for the list

The three-category matrix for the PaedMalayMST consists of at least four words in each category except for only one word in adjective category.

Inclusion of more words from other database

References

Abd Samad, A., & Iman Arshad, N. (2017). The Sequence of Acquisition of Personal Pronoun Case and Person Reference among 6 Year Old Children in Two Selected Malaysian Kindergartens. International Journal of Applied Linguistics and English Literature, 6(3), 273. https://doi.org/10.7575/aiac.ijalel.v.6n.3p.273

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. Journal of the American Academy of Audiology, 12(10), 514–522.

Bullard, J. (2010). Why is the Environment Important for Children's Learning ? Parent Guides by Grade. In Pearson Allyn Bacon Prentice Hall (pp. 12–14). Pearson Education Inc.

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. Wiley Interdisciplinary Reviews: Cognitive Science, 1(5), 629–647. https://doi.org/10.1002/wcs.63.Speech

Ciscare, G. K. S., Mantello, E. B., Fortunato-Queiroz, C. A. U., Hyppolito, M. A., & Dos Reis, A. C. M. B. (2017). Auditory speech perception development in relation to patient's age with cochlear implant. International Archives of Otorhinolaryngology, 21(3), 206–212. https://doi.org/10.1055/s-0036-1584296

Coene, M., Van Der Lee, A., & Govaerts, P. J. (2015). Spoken Word Recognition Errors in Speech Audiometry: A Measure of Hearing Performance? BioMed Research International, 2015. https://doi.org/10.1155/2015/932519

Fei, J., Aiting, C., Yang, Z., Xin, X., & Dongyi, H. (2010). Development of a script of phonemically balanced monosyllable lists of Mandarin–Chinese. Journal of Otology, 5(1), 8–19. https://doi.org/10.1016/S1672-2930(10)50003-5

FINAS. (2016). Malaysia film festival 28.

Forthcoming, O. F. (1980). September 21-22 Sep tern ber 26-30 October 20-23, 2, 1980–1982.

Gervain, J., & Werker, J. F. (2008). Infant Speech Perception Contributes to Language Acquisition, 6, 1149–1170. https://doi.org/10.1111/j.1749-818X.2008.00089.x

Hearing abilities measured with the Hearing in Noise Test (HINT) Studies on normal hearing and cochlear implant users Marte Myhrum Oslo University Hospital. (2016).

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. International Journal of Audiology, 51(7), 536–544. https://doi.org/10.3109/14992027.2012.670731

Hörtech Oldenburg. (2016). International Matrix Tests: Reliable speech audiometry in noise, 1–20. Retrieved from http://www.hoertech.de/images/hoertech/pdf/mp/Internationale-Matrixtests.pdf

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., Verschuure, H., & Dreschler, W. A. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. International Journal of Audiology, 53(10), 760–763. https://doi.org/10.3109/14992027.2014.920111

References

Jamaluddin, S. A. (2016). Development Of The Uc Auditory- Visual Matrix Sentence Test. University of Canterbury.

Jerger, S., Lewis, S., Hawkins, J., & Jerger, J. (1980). Pediatric speech intelligibility test. I. Generation of test materials. International Journal of Pediatric Otorhinolaryngology, 2(3), 217–230. https://doi.org/10.1016/0165-5876(80)90047-6

Ji, F., Xi, X., Chen, A. T., Ying, J., Wang, Q. J., & Yang, S. M. (2011). Development of a Mandarin monosyllable test material with homogenous items (I): homogeneity selection. Acta Otolaryngologica, 131(9), 962–969. https://doi.org/10.3109/00016489.2011.574646

Jin, L., A. Razak, R., Wright, J., & Song, J. (2010). Issues in developing grammatical assessment tools in Chinese and Malay for speech and language therapy, 1974(April), 145–156. https://doi.org/10.1017/CBO9781139084642.016

Khalil, N. A. (2017). A pilot study: developing Malay speech audiometry materials for clinical use in Singapore.

Kirk, K. I., Diefendorf, A. O., Pisoni, D. B., & Robbins, A. M. (1995). Assessing Speech Perception in Children.

Kuhl, P. K. (2007). Cracking the speech code: How infants learn language. Acoustical Science and Technology, 28(2), 71–83. https://doi.org/10.1250/ast.28.71

Lancioni, G. E., & Hoogland, G. A. (1980). Full-length Papers HEARING ASSESSMENT IN YOUNG INFANTS BY MEANS OF A CLASSICAL CONDITIONING PROCEDURE, 2.

Lee, L. W., & Low, H. M. (2011). Developing an online Malay Language Word Corpus for primary schools. International Journal of Education and Development Using Information and Communication Technology (IJEDICT), 7(3), 96–101.

Lee, L. W., & Low, H. M. (2014). The development and application of an online malay language corpus-based lexical database. Kajian Malaysia, 32, 151–166.

Locke, J. L. (1971). Phoneme perception in two- and three-year-old children. Perceptual and Motor Skills, 32(March 1971), 215–217 ST–Phoneme perception in two–and three. https://doi.org/10.2466/pms.1971.32.1.215

Martin, F. N., Champlin, C. a, & Perez, D. D. (2000). The question of phonetic balance in word recognition testing. Journal of the American Academy of Audiology, 11(9), 489–493; quiz 522.

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. International Journal of Audiology, 47(9), 546–553. https://doi.org/10.1080/14992020802252261

Mode, P., & Format, R. (2012). General purposes of Speech Audiometry Tests : Presentation Mode : Speech Audiometry Tests : Speech Detection Threshold (SDT).

Mukari, S. Z., & Said, H. (1991). The development of Malay speech audiometry. Medical Journal of Malaysia, 46(3), 262–268.

References

Neumann, K., Baumeister, N., Baumann, U., Sick, U., Euler, H. A., & Weißgerber, T. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. International Journal of Audiology, 51(3), 157–163. https://doi.org/10.3109/14992027.2011.633935

Ozimek, E., Kutzner, D., & Libiszewski, P. (2012). Speech intelligibility tested by the Pediatric Matrix Sentence test in 3-6 year old children. Speech Communication, 54(10), 1121–1131. https://doi.org/10.1016/j.specom.2012.06.001

Phoon, H. S., Abdullah, A. C., Lee, L. W., & Murugaiah, P. (2014). Consonant acquisition in the Malay language: A cross-sectional study of preschool aged malay children. Clinical Linguistics & Phonetics, 28(5), 329–345. https://doi.org/10.3109/02699206.2013.868517

Razak, R. A., Jin, L., & Hui, L. (2016). Profiling malay children's syntactic development : A Malay-LARSP Profi ling Malay Children ' s Syntactic Development : A Malay-LARSP, (January).

Razak, R. A., Madison, C. L., Siow, Y. K., & Aziz, M. A. A. (2010). Preliminary content validity and reliability of a newly developed Malay preschool language assessment tool. Asia Pacific Journal of Speech, Language, and Hearing, 13(4), 217–234. https://doi.org/http://dx.doi.org/10.1179/136132810805334985

Salehuddin, K. (2009). The acquisition of numeral classifiers by Malay children, (September), 308.

Sarudin, A., Janan, D., Osman, Z., Khair, A., & Noor, M. (2016). Penguasaan Bahasa Melayu Kanak-kanak Prasekolah Berdasarkan Inventori Penilaian Kemahiran Terancang, 4, 93–106.

Sidi. (2011). Malay Interrogative Knowledge Corpus. American Journal of Economics and Business Administration, 3(1), 171–176. https://doi.org/10.3844/ajebasp.2011.171.176

Theunissen, M., Swanepoel, D. W., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. International Journal of Audiology, 48(11), 743–757. https://doi.org/10.3109/14992020903082088

Thompson, G. (2005). A child's place: Why environment matters to children. Ecos, 26(1), 9–12.

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. The Journal of the Acoustical Society of America, 107(3), 1671–1684. https://doi.org/10.1121/1.428451

Wallace, I. F., Berkman, N. D., Watson, L. R., Coyne-Beasley, T., Wood, C. T., Cullen, K., & Lohr, K. N. (2015). Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review. Pediatrics, 136(2), e448–e462. https://doi.org/10.1542/peds.2014-3889