


Development of Psychometrically Equivalent **Singapore English** **Speech Audiometry Materials**

Presented by: Justina Tan (justinatanyuhan@gmail.com)
Research Supervisors: A/Prof Jenny Loo & Ms Kimberly Soh

CONTENT



1. Introduction
 2. Background
 3. Methods
 4. Results
 5. Discussion
 6. Conclusion
- 

Introduction

Speech Audiometry

- Test of auditory function using speech materials (recorded, live-voice)

Speech Recognition Threshold (SRT)

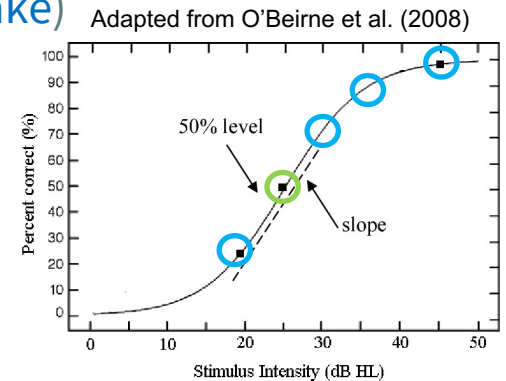
- Softest level at which speech stimuli can be detected 50% of the time
- Tested using *spondaic* words (e.g., blackboard, hotcake)

Word Recognition Score (WRS)

- Percentage of accuracy at a specified intensity
- Tested using *monosyllabic* words (e.g., run, spin)

Performance Intensity Function (P-I Function)

- Plotting percentage of accuracy (y-axis) against presentation level (x-axis)



Background

Problem: Existing English speech audiometry materials are not culturally appropriate

- Existing materials are recorded for and by monolingual English speakers
- Live speech presentation is used despite its significant drawbacks
- Previously recorded materials were not adequate (Soo, 2015; New, 2017)

Leads to:

- Minimal differences in performance compared to recorded American versions
- Lower scores than expected due to unfamiliarity with words/pronunciation rather than problem in audition

Research Aim

Develop **culturally-appropriate** and **psychometrically equivalent** speech test materials that can be used for testing the general Singaporean.

- 100 spondaic words for SRT testing
- 500 monosyllabic words for WRS testing



Source of language data: **National Speech Corpus (NSC)** by NTU & the Infocomm Media Development Authority (IMDA) (Koh et al., 2019).

- Created for natural language processing in speech recognition with a component containing 1000 hours of spontaneous speech between 250 pairs of speakers

Method



01

Phase I

Word Selection and
Familiarity Rating

02

Phase II

Recording, rating, selection,
acoustic standardization

03

Phase III

Perceptual Testing and
Equating Word Difficulty

04

Phase IV

Phonemic balancing,
validation etc...

Preparation



Word Selection

- 1000 hours of spontaneous conversation transcripts
- Data cleaning
- 17.5 million tokens (43,000 unique words)
- Sorted based on frequency of occurrence

Exclusion Criteria:

1. Contractions
2. Non-English words
3. Inappropriate or offensive terms

502 MS and 123 SD words were chosen.



PHASE I

Word Familiarity Rating

Age Group	No. of participants	Gender Proportion
21 to 39	4	2M, 2F
40 to 59	4	2M, 2F
60 to 79	2	1M, 1F

Based on proportion of age groups speaking English at home (Singapore department of Statistics, 2015).

- 10 Singaporeans (24 to 75 years old)
- Speakers of English and at least one other language (Chinese, Tamil, Malay etc.)
- 8 Chinese, 1 Malay, 1 Indian

Tasks:

1. Rate 502 MS/123 SD words on familiarity on a 5-point likert scale.
2. Complete a language background questionnaire.

494 MS and 110 SD words were chosen.

PHASE II

1. Speaker Selection
2. Recording of words
3. Word Intelligibility Rating
4. Acoustic Standardization

Speaker: 26 y/o Native Singaporean male speaker

Recording: Professional recording studio according to BSI guidelines (2012). Pronounced with NSC-determined IPA

Raters: 4 bilingual Singaporean raters (26 to 54 years old, 2M, 2F)

Acoustic Standardization: Equate RMS levels across test words

452 MS and 110 SD words were acoustic standardized.

PHASE III

PARTICIPANTS

- 10 Singaporean bilinguals
- 4 males and 6 females
- Aged 21 to 25 years old
- $M = 24$, $SD = 1.49$

Pure tone audiometry

- 250Hz to 8000Hz
- PTA: -2 to 18 dB HL
- $M = 10.7$ dB HL, $SD = 5.34$

PROCEDURES

- Language background questionnaire
- Ascending presentation of words (2-dB steps)
- Plotting of P-I functions

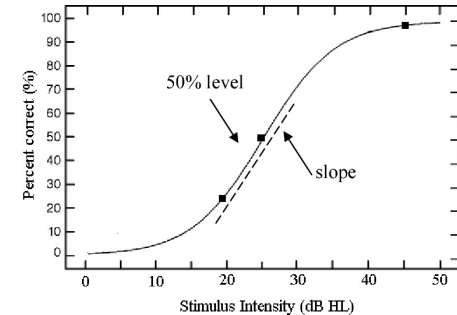
270 MS and 110 SD words were used in the speech identification task.

Results

Phase III – Perceptual Testing and Equating Word Difficulty

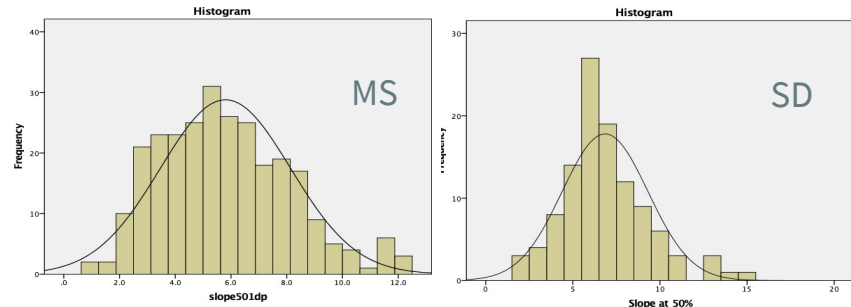
Logistic Regression

- Logistic regression done to obtain regression slopes and intercepts
- Obtained slopes at 50% and 20-80% performance
- P-I functions plotted for every word



Test of normality

- MS: Normal distribution
- SD: Leptokurtik normal distribution

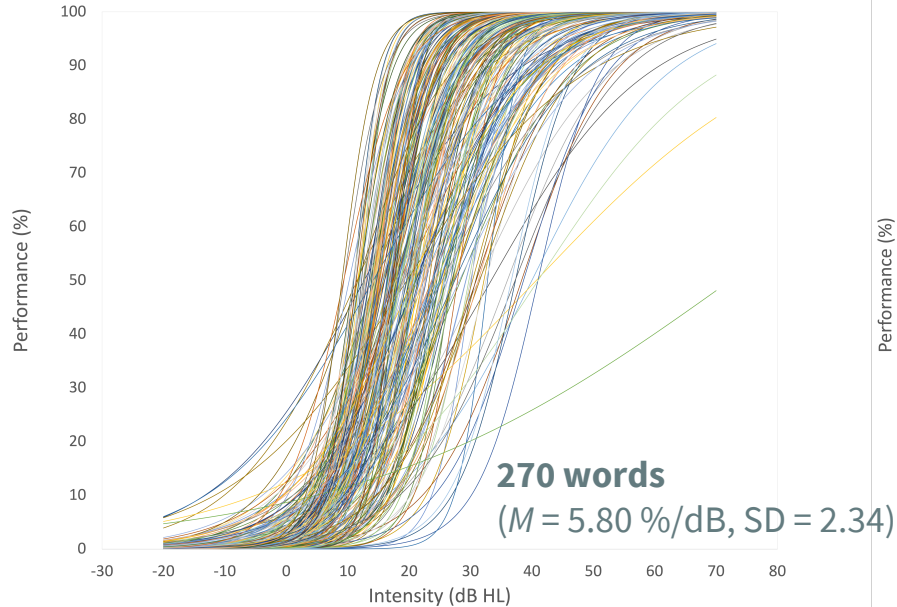


Cut off: 10th to 90th percentile
(2.86%/dB to 8.98%/dB)

Results

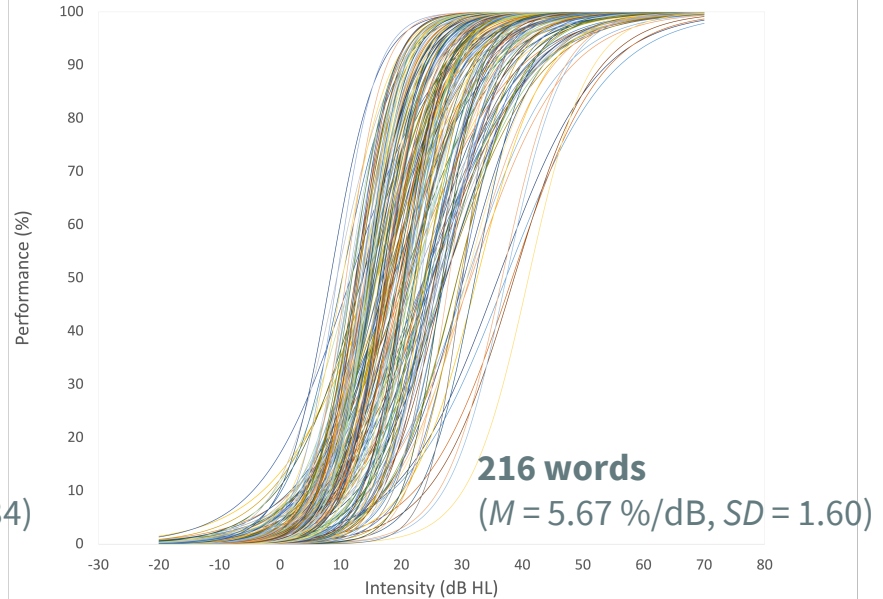
Before cut-off

Monosyllabic Words



After cut-off

Monosyllabic Words

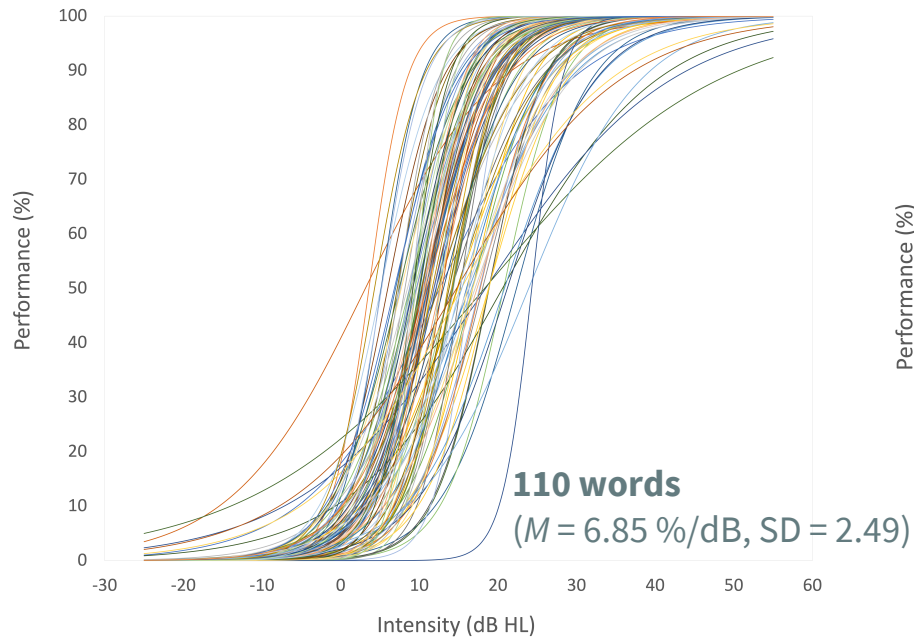


Cut off: 20th to 80th percentile
(4.76%/dB to 8.57%/dB)

Results

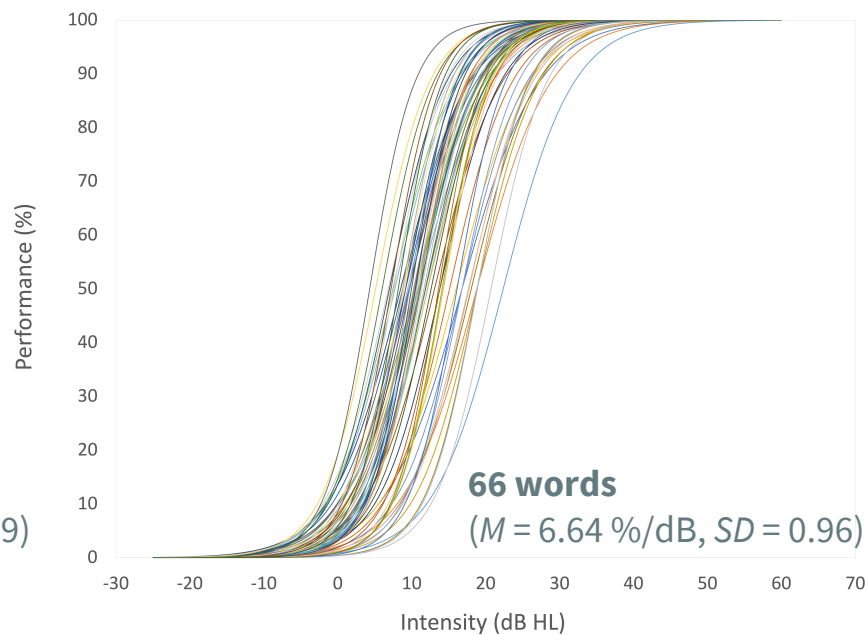
Before cut-off

Spondaic Words



After cut-off

Spondaic Words



Results

Phase III – Perceptual Testing and Equating Word Difficulty

Language Background Questionnaire

- 2 English-Tamil bilinguals, 8 English-Mandarin Chinese bilinguals
- All participants reported English as predominant language
- All participants lived in Singapore since birth
- All rated English proficiency from "very well" to "extremely well"

8) How well do you **speak** the following languages?

	Extremely well	Very well	Moderately well	Slightly well	Not familiar at all
English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mother Tongue	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Discussion

Steepness of slope

- The slope steepness indicates the maximum rate that performance increases as stimulus intensity increases (McPherson & Akeroyd, 2014)
- Psychometric equivalence can be achieved through intensity adjustment of words with similar slopes to the mean PTA of the participants (e.g., Lee & Lee, 2020)

Current study (Slope at 50%):

Spondees – 1.70 to 15.47%/dB ($M = 6.64$)

Monosyllables – 0.81 to 12.28%/dB ($M = 5.67$)

Discussion

Cut-off criteria selection

- Monosyllables vs Spondees
- Differences in SRT and WRS requirements (i.e., test-list homogeneity vs test-item homogeneity)

Words with high error rates

- “While” – “wow”, “wild” (90%)

Word-initial or word-final fricatives:

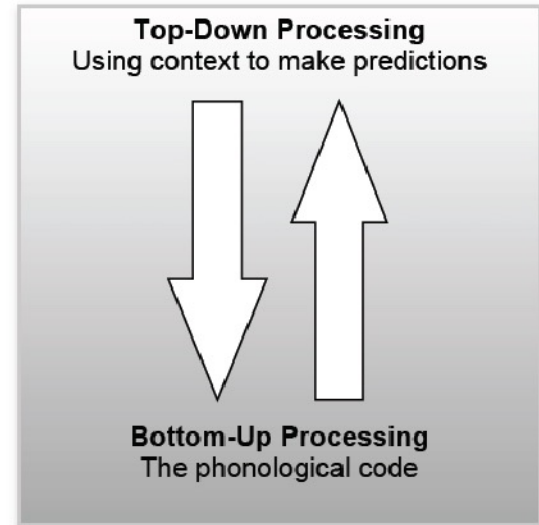
- “**th**an” → “van”, “zen” (70%)
- “worth” → “worse”, “worst” (50%)
- “**f**it” → “sit” (40%)
- “**th**ese” → “bees”, “vees” (20%)



Discussion

Phoneme discrimination

- Individual differences
- Contextual cues and top-down processing in spondees vs. monosyllables
- Role of pronunciation familiarity
- ? Standard Singaporean pronunciation

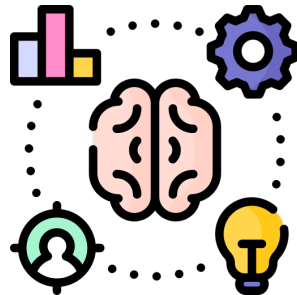
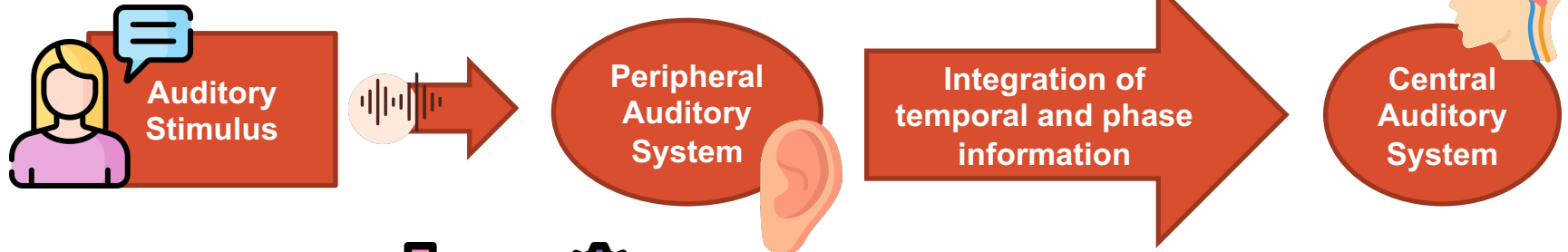


Taken from Zoghbor (2016)

Discussion

Test scoring

- Accurate information transmission between listener and speaker
- Influenced by prior knowledge and sensitivity to phonemic contrasts



i:	ɪ	ʊ	u:	ɪə	eɪ	e	ə	ɜ:	ɔ:	ʊə
see	sit	book	too	here	day	men	about	word	sort	tour
ɔɪ	əʊ	æ	ʌ	ɑ:	ɒ	eə	aɪ	aʊ	p	b
boy	go	cat	but	part	not	wear	my	how	pig	bed
t	d	tʃ	dʒ	k	g	f	v	θ	ð	s
time	do	church	judge	kilo	go	five	very	think	the	six
z	ʃ	ʒ	m	n	ŋ	h	l	r	j	w
zoo	short	casual	milk	no	sing	hello	live	read	yes	we

Limitations & Recommendations



Limited resources (time, sample size)



Participant fatigue



Possible tester bias



Longer data collection, larger sample size



Digital scaling, scoring alternatives



Validation studies

References

- British Standards Institution. (2012). BS EN ISO 8253-3:2012: Acoustics – Audiometric test methods. Part 3: Speech Audiometry.
- Beattie, R. C., Edgerton, B. J., & Svihovec, D. V. (1977). A comparison of the Auditec of St. Louis cassette recordings of NU-6 and CID W-22 on a normal-hearing population. *Journal of Speech and Hearing Disorders*, 42(1), 60-64.
- Honu-Mensah, C. M. (2015). Development and Evaluation of Psychometrically Equivalent Trisyllabic Words for Speech Audiometry in Fante (Doctoral dissertation, University Of Ghana).
- Koh, J. X., Mislan, A., Khoo, K., Ang, B., Ang, W., Ng, C., & Tan, Y. Y. (2019). Building the singapore english national speech corpus. *Malay*, 20(25.0), 19-3.
- Lee, G. J. C., & Lee, S. L. H. (2020). Development of SC-10: A psychometrically equivalent Singapore Mandarin disyllabic word list for clinical speech audiometry use. *World Journal of Otorhinolaryngology-Head and Neck Surgery*.
- MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, 18, 2331216514537722.
- Nissen, S. L., Harris, R. W., Jennings, L. J., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin: Palabras trisilábicas psicométricamente equivalentes para evaluar umbrales de recepción del lenguaje en Madarín. *International Journal of Audiology*, 44(7), 391-399.
- O'Beirne, G., Sincock, B.P., & Starfinger, C. (2008). Clinical Applicability of Adaptive Speech Testing: A comparison of the administration time, accuracy, efficiency and reliability of adaptive speech tests with conventional speech audiometry.
- Zoghbtor, W. S. (2016). A Model for Speech Processing in Second Language Listening Activities. *English Language Teaching*, 9(2), 13-19.



THANK YOU!

