Validation of Mandarin Speech Audiometry Materials in Singapore

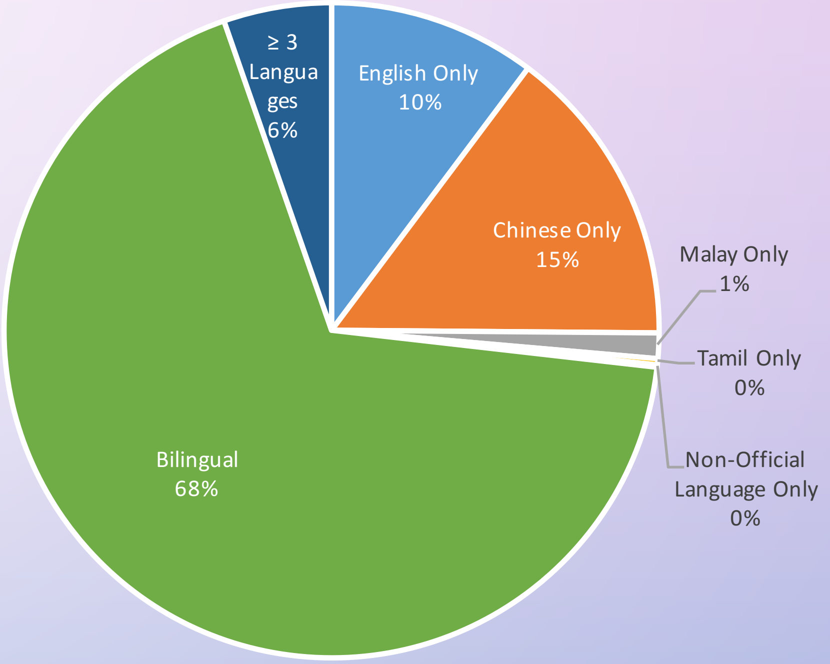
Co-investigator: Soh Wanxian (Kimberly)

Principal Investigator: Dr. Jenny Loo

Background

- Speech audiometry is not routinely conducted in Singapore clinics due to a lack of suitable speech audiometry materials.
- A significant portion of Singapore population, especially the elderly, is monolingual in Mandarin.
  - Mandarin-speaking population: 471,861 (in 2015)
  - 418,400 are ≥ 45 years old
- Speech audiometry materials developed in Standard Mandarin and Taiwan Mandarin are unsuitable
  - Lexical and pronunciation differences from Singapore Mandarin (Chua, 2003; Lee, 2010; Lock, 1989; Ng, 1985).
  - Heavy influence of other languages and dialects on Singapore Mandarin (Chen, 1986).
  - Speech audiometry scores can be adversely affected if it is not conducted in the native language or accent of an individual (Nissen et al., 2011; Weisleder & Hodgson, 1989).
- **Locally developed Mandarin speech audiometry materials** (Lee, Lee, Lim, Chang, & Lu, 2007) for speech recognition testing in Singapore clinics.
  - 10 lists of 10 disyllabic words recorded into a CD

Literate Resident Population in 2015 (≥ 15 y.o.)



Objective of Study

To validate the locally developed Mandarin speech audiometry materials (Lee et al., 2007) on **normal hearing** and **hearing impaired** native speakers of Singapore Mandarin.

Methodology

- 29 normal hearing participants – AC thresholds ≤25 dB HL
- 23 hearing impaired participants

Basic Hearing Test

- Otoscopy
- Tympanometry
- Pure tone audiometry

PTA / Fletcher Average (2FA)

- PTA = Average of 500Hz, 1kHz & 2kHz thresholds
- 2FA used in steeply sloping losses
  - Average of best 2 thresholds at 500Hz, 1kHz & 2kHz

SRT test

- ASHA method (1988) in 5 dB steps

Speech Recognition Test

- SRT
- SRT-10 dB
- SRT-5 dB
- SRT+5 dB
- SRT+10 dB
- SRT+20 dB
- SRT+30 dB
- SRT+40 dB

(SRT to SRT+10 dB: x2)

Results - Hypothesis #1

Hypothesis #1: PTA ≈ SRT ≈ dSRT

Pure Tone Average (PTA)	Speech Reception Threshold (SRT)	Derived SRT (dSRT)
• Obtained from audiogram • Average of 500Hz, 1kHZ & 2kHz AC thresholds	• ASHA Method (1988)	• Derived from 50% SRS point on P-I curves

Results - Hypothesis #1

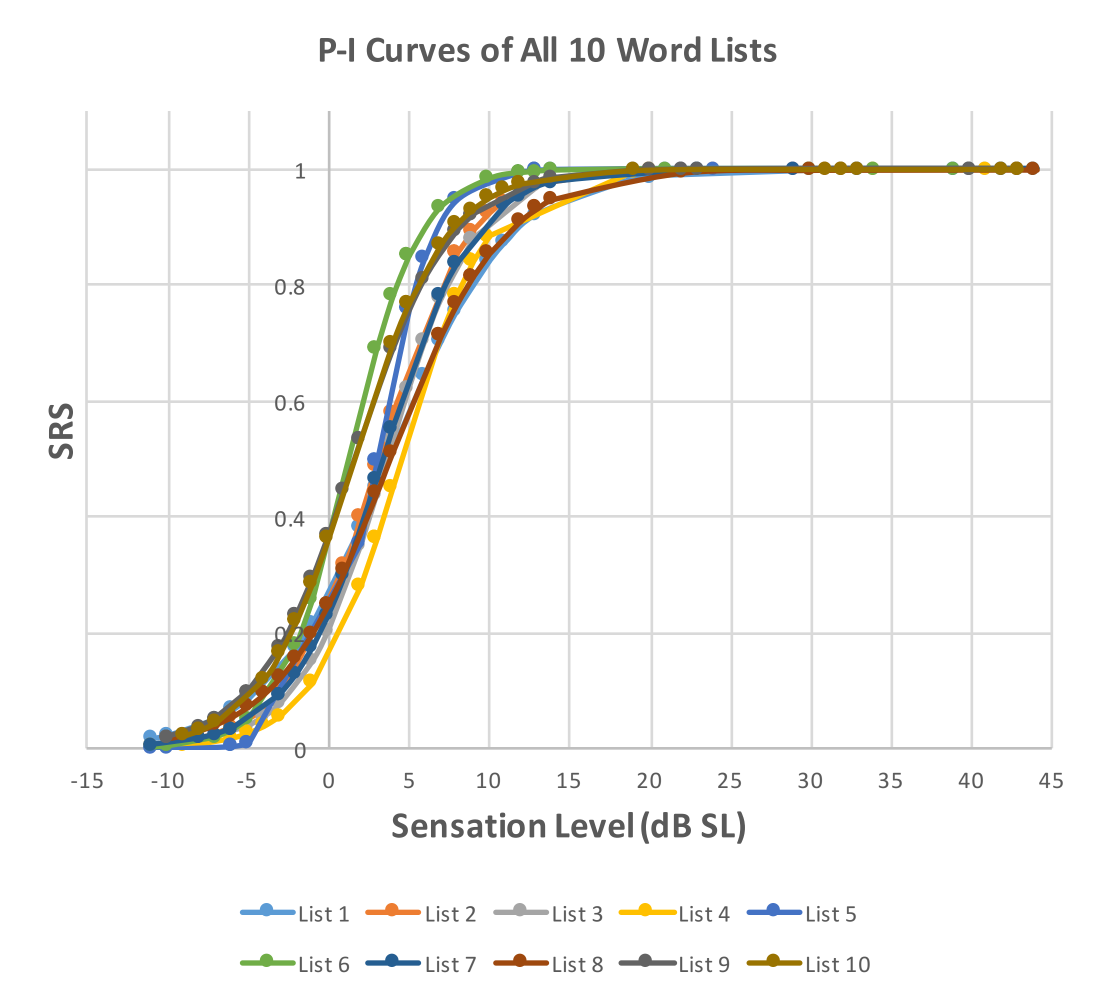
Hypothesis #1: PTA ≈ SRT ≈ dSRT

- Strong correlation and good agreement between PTAs and SRTs
  - **r_s = .84**
  - Mean difference of **6.03 dB** across all participants
  - ➔ **Within 10 dB and hence unlikely to falsely trigger suspicion of NOHL.**
- Strong correlation and very close agreement between SRTs and dSRTs
  - **r_s = .95**
  - Mean difference of **2.98 dB** across all participants
  - ➔ **SRT can be confidently used to predict intensity level at which 50% SRS will be scored.**

	Normal hearing	Hearing impaired
PTA and SRT		
Correlation, r	.40	.83
Paired t-test of PTA and SRT		
Mean Difference, M	6.17 dB	5.85 dB
95% CI	4.47 – 7.87	3.40 – 8.29
SRT and dSRT		
Correlation, r	.85	.94
Paired t-test of SRT and dSRT		
Mean Difference, M	2.97 dB	2.98 dB
95% CI	2.20 – 3.73	1.51 – 4.45

Results - Hypothesis #2

Hypothesis #2: The 10 word lists are perceptually equivalent.



dB Sensation Levels Required for Selected SRS								
	20%	30%	40%	50%	60%	70%	80%	90%
List 1	-1.40	0.60	2.24	3.76	5.27	6.92	8.93	11.95
List 2	-0.71	0.78	2.00	3.11	4.23	5.46	6.94	9.17
List 3	-0.07	1.37	2.57	3.65	4.74	5.91	7.36	9.55
List 4	0.79	2.24	3.43	4.52	5.60	6.79	8.24	10.43
List 5	0.63	1.56	2.32	3.02	3.73	4.49	5.43	6.83
List 6	-1.71	-0.54	0.41	1.27	2.14	3.09	4.26	5.98
List 7	-0.51	0.99	2.24	3.39	4.52	5.77	7.28	9.56
List 8	-1.00	0.87	2.40	3.81	5.22	6.74	8.63	11.45
List 9	-2.53	-0.92	0.41	1.61	2.83	4.15	5.76	8.18
List 10	-2.33	-0.82	0.44	1.60	2.74	3.99	5.52	7.81

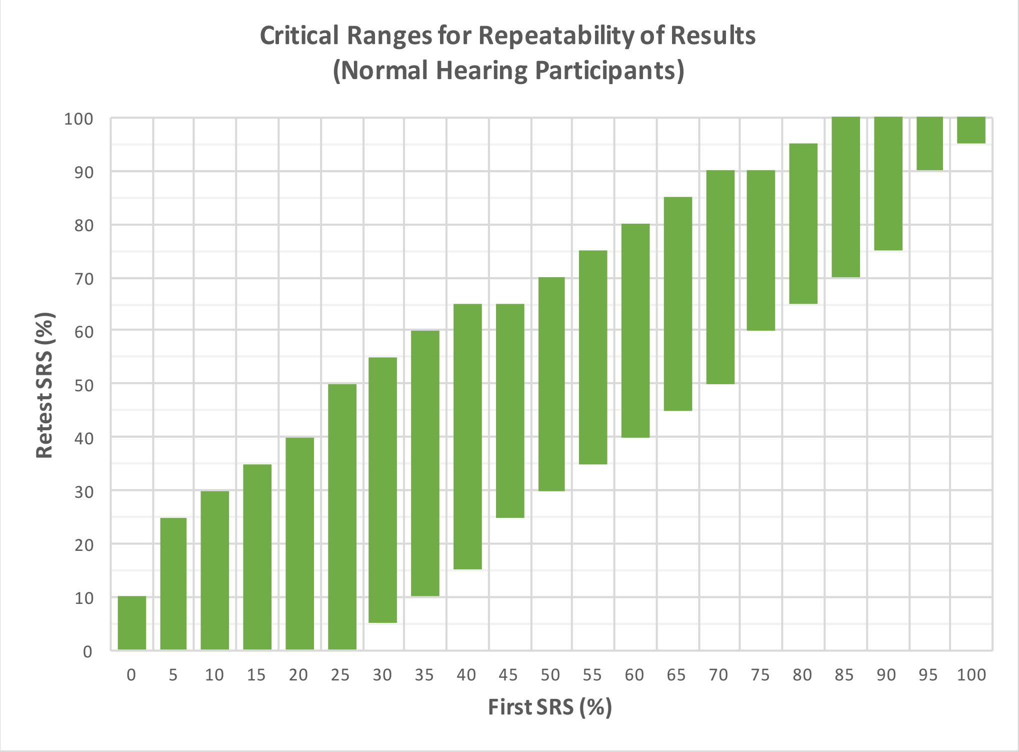
ANOVA on dB SL required to score selected SRS on each of the 10 word lists.

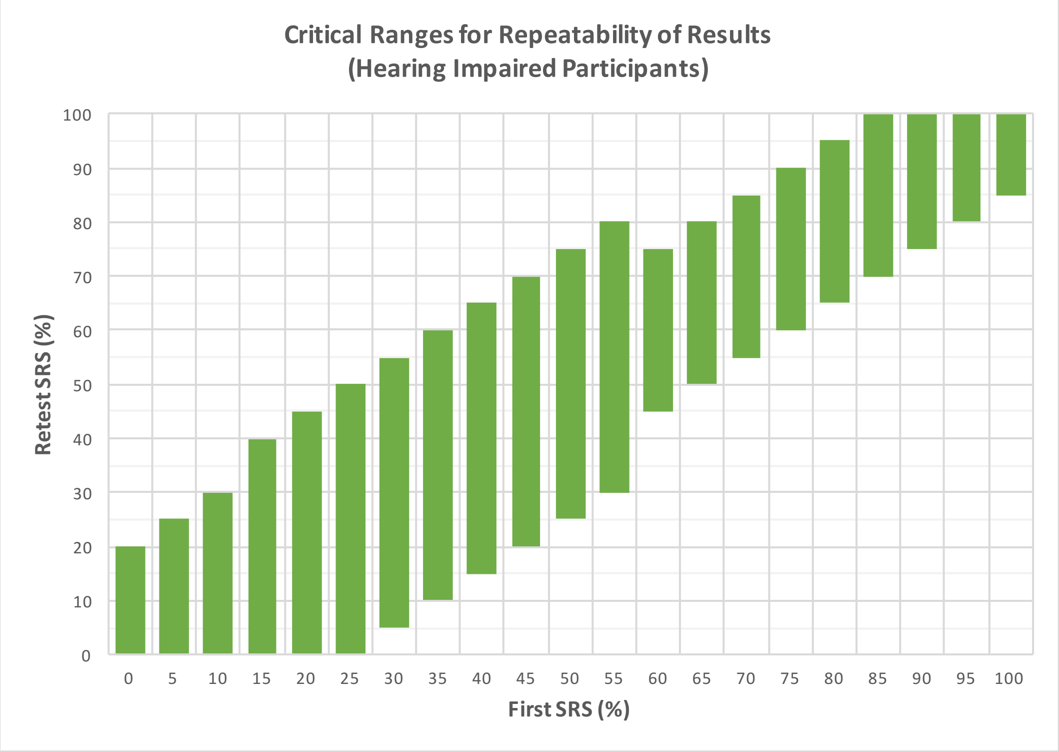
Source of variation	SS	df	MS	F	P-value	F crit
Between Groups	97.79	9	10.86	0.96	0.48	2.02
Within Groups	791.11	70	11.30			
Total	888.84	79				

Results - Hypothesis #3

Hypothesis #3: Repeatability of Results (Test-retest reliability)

- Strong correlation between first and retest SRS across all participants
    - Spearman's rank correlation, r_s = .93
- Critical ranges calculated to determine if a retest SRS is really different from the first SRS





Results – Hypothesis #3

Hypothesis #3: Repeatability of Results (Test-retest reliability)

- List combinations with poor repeatability of results

First List \ Second List	1	2	3	4	5	6	7	8	9	10
1					30 (1)	20 (1)	21.7 (3)	25 (1)		
2	20 (1)								35 (1)	
3				25 (1)	22.5 (2)			25 (1)		25 (2)
4			25 (1)			43.8 (4)		15 (1)		
5		37.5 (2)				17.5 (2)				20 (1)
6				25 (1)	10 (1)			10 (1)		
7	25 (1)									26.3 (4)
8							35 (1)		50 (1)	
9	25 (1)		20 (1)		32.5 (2)			32.5 (2)		
10					20 (1)			35 (1)	45 (1)	

Cases where retest SRS exceeded critical range of first SRS. Values within the cells indicate the average difference between first and retest SRS, while numbers in parentheses show the number of such occurrences for the particular list combination.

	Mean (dB SL)
List 4	5.26
List 1	4.78
List 8	4.76
List 3	4.39
List 7	4.16
List 2	3.87
List 5	3.50
List 9	2.44
List 10	2.37
List 6	1.86

Mean dB SL required to be perceived

Results - Hypothesis #4

Hypothesis #4: Maximum SRS decreases with increased severity of sensorineural hearing loss.

- 22 out of 23 hearing impaired participants achieved 100% SRS within the tested intensity ranges
- The remaining hearing impaired participant achieved a maximum of 95% SRS.
- Possible reasons:
  - SNHL not severe enough to impact speech discrimination – highest PTA 52.5 dB HL
  - Good low-mid frequency hearing → good perception of Mandarin finals (vowels and/or nasals) → additional cues from tone

Classification of SNHL	Average dB SL required to achieve max SRS
Mild	17.2
Mild to Moderate	19.5
Mild to Moderately-Severe	26
Mild to Severe	26

→ Testing at SRT+30 dB or SRT+40 dB would be appropriate for determining optimal speech recognition performance of someone with such a hearing profile.

Results – Commonly Misheard Words

List No.	Word Pair Wrong Response (No. of occurrences)	
List 1	旅 lǚ	游 yóu
	lǐ (7)	
List 2	变 biàn	化 huà
	diàn (29)	
	哀āi	伤 shāng
	ān (7)	
List 3	重 zhòng	点 diǎn
	zhèng (8)	
	发 fā	展 zhǎn
		zhǎng (7)
List 5	恩ēn	人 rén
	ān (8)	
List 8	城 chéng	市 shì
		shí (8)

Wrong responses provided by participants ≥10% of the time. Numbers in parentheses indicate number of times the particular response was given.

Issues in Scoring

1. Responses influenced by Chinese dialects
2. Word pairs containing double third tones

Issues in Scoring

1. Responses influenced by Chinese dialects

<u>Example:</u>

- Hokkien as first acquired dialect
- Difficulty pronouncing /ü/, tendency to replace it with /i/
- 月亮 (moon) yùe liàng → yè liàng
    - Non-existent word pair in Mandarin
- 旅游 (travel) lǚ yóu → lǐ yóu 理由 (reason)
    - Perceived 旅游 correctly, but unable to pronounce accurately?
    - Perceived incorrectly as 理由, responded exactly what s/he heard?
- Not useful to provide forced choice ("was it 旅游 or 理由"?)
    - Repeating the stimulus
    - Might suggest the correct response to them
    - They might tell you the same thing again because they cannot pronounce it properly

→ Requested them to express the meaning in another language / dialect (which I understand).

Issues in Scoring

2. Word pairs containing double third tones

Example:

Stimulus: 产品 chǎn3 pǐn3 (pronounced as “chán2 pin3”)

Response: chǎn3

- Perceived “chán2” wrongly as “chǎn3”, and responded with what was wrongly heard?
- Perceived 产 correctly, pronounced it in its original third tone?

➔ Asked them to repeat what they heard

➔ Benefit of doubt given. Scored as 5%.

Summary & Recommendations

Hypothesis #1: PTA ≈ SRT ≈ dSRT ✓

Hypothesis #2: The 10 word lists are perceptually equivalent. ✓

Hypothesis #3: Repeatability of Results (Test-retest reliability) ✓

Hypothesis #4: Maximum SRS decreases with increased severity of SNHL. ✗

Recommendations:

- Do a retest to check reliability of obtained SRS against the critical range
- Avoid using Lists 4 and 6 in conjunction on the same individual
- Leave out the word lists with word pairs that both bear the third tone (Lists 2, 5, 6 and 10)

Conclusion

- Appropriate for use on native speakers of Singapore Mandarin who have
  - Normal hearing; or
  - SNHL that is not worse than mild to severe downward sloping presbycusis
- The materials from List 1, 3, 4, 7, 8 and 9 have been deemed suitable for obtaining both SRT and SRS in the clinic.
- Testers need to be fully aware of the limitations in testing, scoring and interpretation of results before using the materials.

Future Studies

- Patients with conductive hearing loss
- Patients with more severe SNHL (higher PTAs and flatter audiograms at severe or profound HL ranges)
- Patients already diagnosed with retrocochlear lesions to test for rollover effect
- Development of Mandarin sentence tests and speech in noise tests

Acknowledgements

- Dr. Jenny Loo – for her patience in providing invaluable guidance and support
- Mr. Gary Lee (NTFGH) – for generously sharing the Singapore Mandarin speech audiometry materials that he had previously developed with his team
- All participants of the study – who spared time and effort, and even referred other participants to me
- Classmates who graciously volunteered their time to participate in the study
- Every person who has played a part in the MSc Audiology course

And most importantly, God, whose enabling grace saw me through the past two years.

References

American Speech-Language-Hearing Association. (1988). Determining threshold level for speech [Guidelines]. https://doi.org/10.1044/policy.GL1988-00008

Chen, C. Y. (1986). Salient segmental features in Singapore Mandarin 新加坡华语里声母与韵母的特征. *Journal of Chinese Linguistics*, 114–151.

Chua, C. L. (2003). *The emergence of Singapore Mandarin: A case study of language contact*. University of Wisconsin--Madison.

Fletcher, H. (1950). A method of calculating hearing loss for speech from an audiogram. *Acta Oto-Laryngologica*, *38*(sup90), 26–37.

Han, D., Wang, S., Zhang, H., Chen, J., Jiang, W., Mannell, R., … Zhang, L. (2009). Development of Mandarin monosyllabic speech test materials in China. *International Journal of Audiology*, *48*(5), 300–311.

Kim, J., Lee, J., Lee, K. W., Bahng, J., Lee, J. H., Choi, C.-H., … Park, J. (2015). Test-Retest Reliability of Word Recognition Score Using Korean Standard Monosyllabic Word Lists for Adults as a Function of the Number of Test Words. *Journal of Audiology & Otology*, *19*(2), 68–73.

Lee, J. C. G., Lee, L. H. S., Lim, S. H. N., Chang, Y. S., & Lu, Y. M. (2007). *Development and clinical evaluation of Singapore Mandarin speech audiometry materials*. Temasek Polytechnic.

Lee, L. (2010). The tonal system of Singapore Mandarin. In *22nd North American Conference on Chinese Linguistics and the 18th Annual Meeting of the International Association of Chinese Linguistics* (pp. 20–22).

Lock, G. (1989). Aspects of variation and change in the Mandarin Chinese spoken in Singapore. *Australian Journal of Linguistics*, *9*(2), 277–294.

Ng, B. C. (1985). A study of the variable /sh/ in Singapore Mandarin. *Pacific Linguistics. Series A. Occasional Papers*, (67), 31–37.

Nissen, S. L., Harris, R. W., Channell, R. W., Conklin, B., Kim, M., & Wong, L. (2011). The development of psychometrically equivalent Cantonese speech audiometry materials. *International Journal of Audiology*, *50*(3), 191–201.

Nissen, S. L., Harris, R. W., & Dukes, A. (2008). Word recognition materials for native speakers of Taiwan Mandarin. *American Journal of Audiology*, *17*(1), 68–79.

Nissen, S. L., Harris, R. W., Jennings, L.-J., Eggett, D. L., & Buck, H. (2005a). Psychometrically equivalent mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *International Journal of Audiology*, *44*(7), 379–390.

References (cont'd)

Nissen, S. L., Harris, R. W., Jennings, L.-J., Eggett, D. L., & Buck, H. (2005b). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *International Journal of Audiology*, *44*(7), 391–399.

Nissen, S. L., Harris, R. W., & Slade, K. B. (2007). Development of speech reception threshold materials for speakers of Taiwan Mandarin. *International Journal of Audiology*, *46*(8), 449–458.

Singapore Department of Statistics. (2016). Table 25 Resident Population Aged 15 Years and Over by Language Literate In, Age Group, Sex and Ethnic Group. *General Household Survey 2015*.

Thornton, A. R., & Raffin, M. J. M. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech, Language, and Hearing Research*, *21*(3), 507–518.

Wang, S., Mannell, R., Newall, P., Zhang, H., & Han, D. (2007). Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. *International Journal of Audiology*, *46*(12), 719–731.

Weisleder, P., & Hodgson, W. R. (1989). Evaluation of four Spanish word-recognition-ability lists. *Ear and Hearing*, *10*(6), 387–393.

Wilson, R. H., & Carter, A. S. (2001). Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology*, *12*(1), 7–14.

Wilson, R. H., & Strouse, A. (1999). Psychometrically equivalent spondaic words spoken by a female speaker. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1336–1346.

Young, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech, Language, and Hearing Research*, *25*(4), 586–593.

Thank you