

**Designs of Speech Audiometric Tests in  
Vietnamese – The Issues of Normative Values,  
Dialectal Effects, and Tonal Patterns**



**Inauguraldissertation**

Zur

Erlangung des Doktorgrades  
der Humanwissenschaftlichen Fakultät  
der Universität zu Köln

vorgelegt von

**Quoc-Dung Nguyen**

Aus Dong-Thap (Vietnam)

Köln 2017

**Designs of Speech Audiometric Tests in  
Vietnamese – The Issues of Normative Values,  
Dialectal Effects, and Tonal Patterns**

Inauguraldissertation

Zur

Erlangung des Doktorgrades  
der Humanwissenschaftlichen Fakultät  
der Universität zu Köln

vorgelegt von

**Quoc-Dung Nguyen**

Aus

Dong-Thap (Vietnam)

**Köln 2017**

Doktorvater  
**Prof. Dr. Ir. Frans Coninx**

**The work described in this thesis is supported by the Katholischer Akademischer Ausländer-Dienst (KAAD)**

1. **Berichterstatter: Prof. Dr. Ir. Frans Coninx (Köln)**
2. **Berichterstatter: Prof. Dr. Reinhold Greisbach (Köln)**

**Tag des Rigorosums: 18. Mai 2017**

## ACKNOWLEDGEMENTS

My special thanks go to Prof. Dr. Frans Coninx for supervising this thesis and especially for co-operating in the designs of the speech audiometric tests –AAST and NAMES for the Vietnamese language. Without his support and encouragement, this work would hardly have been possible to complete. I also appreciate his devotion of time, energy, and thoughtful insights into my work.

I sincerely thank Dr. Kiều-Phương Hà, Dr. Yaw Nyadu Offei, Dr. Sandra Nekes, and Jun.-Prof. Dr. Karolin Schäfer for the insightful comments and suggestions on this work.

My sincerely thanks also go to AuD. Lê Long Hải for inspiring me in the field of audiology. I have a chance to do my research in this interesting area.

I am grateful to all of my friends in Germany, the Netherlands and Vietnam who have supported me along the way. Especially, I thank to Dr. Anh-Chi Thái Huỳnh, Dr. Vân-Anh Trần Tử, Ms. Ánh-Nguyệt Phạm, Ms. Thị-Lưu Trần, Mr. Thái-Dương Vũ Hoàng and Mr. Hồ Nguyễn for their friendships.

I owe my thanks to all the participants who participated in our research in Vietnam and Germany.

This work was initiated in late 2013 with the financial support of KAAD. I would like to express my gratitude to KAAD for the support and encouragement. A part of data collection was supported by the hospital of Thái Hòa and the company of Truyền Thông Tương Lai; I also want to thank for providing me participants and sound booths.

Finally yet importantly, I would like to express my deepest gratitude to my parents for giving me life and nourishing my spirit.

Thank you, and God bless you all!

## ABSTRACT

**Background:** Dialectal variations and linguistic factors are considered to be the primary causes of misdiagnosis during audiological assessments of speech performances. For new speech audiometry materials, the evaluation of the effects of the listener's dialect or linguistic factors on speech recognition thresholds (SRTs) or supra-threshold phoneme recognition scores (PRSs) would be expected to give a valid and reliable audiometric measurement for clients.

**Purpose:** This thesis assessed the SRTs of native and non-native listeners of Southern Vietnamese regarding the dialectal effects; the effect of tonal patterns of syllables on speech perception of older adults; and the correlations between SRTs and duo-tone thresholds, between SRTs and PRSs.

**Research Design:** To attain the aforementioned objectives, two different types of speech audiometry materials were designed. The Adaptive Auditory Speech Test (AAST) consisted of five subtests of six simple two-syllable words each, which were used to measure an individual's SRT. NAMES included four lists of meaningless two-syllable structures (CV-CVC) containing 20 items each, which were used to determine the listener's supra-threshold PRS. The phoneme distribution in both speech materials was a duplicate of the phoneme distribution in Southern Vietnamese. All speech materials were recorded in Southern Vietnamese, and using female voices. With regard to AAST, the speech intelligibility functions were equalized among speech stimuli within a subtest and across the subtests during optimization.

**Study Sample:** The SRTs of AAST were collected from 435 normal-hearing listeners aged between four and 85 years, of whom 117 participants were non-native listeners of the dialect. The supra-threshold PRSs of NAMES were tested on a sample of 186 normal-hearing listeners aged between 15 and 85 years, including 38 non-native adult listeners.

**Results:** Age-dependent normative values for AAST Vietnamese were resemble to those in other languages in which the age-related difference was roughly 8 dB between younger children and adults, and 11 dB between adults and oldest adults. The supra-threshold PRSs of NAMES were relatively equal across the age groups (except for the listeners above 75). Their scores differed from those of the young listeners with roughly 3 percent correct phoneme scores.

Regarding the dialectal effects, the SRTs of non-native listeners significantly deteriorated compared with those of native listeners of Southern Vietnamese. Differences in the SRTs ranged from 2 to 6 dB for AAST in quiet and 1 to 5 dB for AAST in noise. In contrast, slight differences were found in the PRSs of native and non-native listeners in NAMES—a variation of 2 percent correctness score.

With respect to the effects of tonal patterns of syllables on speech recognition, older listeners obtained somewhat poorer results when the speech stimuli carried tones with low pitch levels and rising contours. In contrast, they performed better when the speech stimuli bore tones with high F0 level and rising contours. For AAST, average SRTs in quiet were comparable across the subtests in contrast to the average speech thresholds in noise, which were more heterogeneous. For example, the mean threshold value for AAST-a3 was significantly poorer than the other subtests. The threshold values in noise deteriorated because the speech stimuli of AAST-a3 carried tones with low F0 level and falling contours. Such speech stimuli were especially delivered to older listeners in masking noise. In NAMES, the older listeners performed slightly better when the stimuli carried a high-level tone *ngang* (A1) rather than low-falling tone (A2) and high-rising tone *sắc* (B1). Similar to young listeners, the older listeners frequently misidentified the tone *huyền* (A2) as the tone *nặng* (B2).

Regarding the relationships between speech materials and duo-tone, there were strong correlations between the SRTs and the duo-tone thresholds (0.5 and 4kHz). However, no or very weak association was found between the SRTs and PRSs, and between the PRSs and duo-tone thresholds.

**Conclusions:** The AAST and NAMES are valid speech audiometric tests to quantify speech recognition of listeners aged between four and 85 (AAST), between 15 and 85 (NAMES). The age-related normative values of AAST in Vietnamese are similar to those in German, Ghanaian, and Polish. The findings of the dialectal study indicate that dialectal variation has an impact on speech recognition. However, the extent of the effects depends on the speech materials being used for the measurement. In statistical terms, dialects significantly affect an individual's SRT if AAST with a semantic content is used in an adaptive procedure. The negligible differences in PRSs seem to indicate that the influences of dialects are minimal when NAMES is used without a semantic content at a constant presentation level. In other words, more effects of dialectal differences in “open speech tests with meaningful words” were found as compared to “closed speech test”. The findings on tonal pattern effects seem to implicate that the tonal patterns of syllables have a minor influence on speech perception of older adults, especially those above 75. Finally, the SRTs could be predicted using duo-tone thresholds. In contrast, the PRSs could not be predicted using either speech thresholds or duo-tone thresholds based on the correlations.

The two new speech audiometric tests provide reliable outcomes with the same properties in normal-hearing listeners as compared to the other AAST and nonsense syllable tests in the different languages. These two speech audiometric tests complement each other in evaluating hearing loss or language impairment. It is claimed that these speech tests will serve as an effective clinical tool for speech audiometric testing in Vietnam.

## ZUSAMMENFASSUNG

**Hintergrund:** Dialektale Variationen und linguistische Faktoren werden als Hauptgründe für Fehldiagnosen in audiologischen Bewertungen und Sprachverhaltengesehen. Für neue Audiometriematerialien, werden Evaluationen der Hörerdialekteffekte oder linguistische Faktoren für die Sprachverständnisschwellen oder Oberschwellenphonemerkennungsscores erwartet, um ein valides und verlässliches audiometrisches Maß Nutzern zu geben.

**Ziel:** Diese Doktorarbeit bewertet Sprachverständnisschwellen bei muttersprachlich-südvietnamesischen Hörern, Sprachverständnisschwellen bei nicht-muttersprachlichen Hörern bezüglich dialektalen Effekten, Sprachperzeption älterer Personen bei Effekten der tonalen Struktur von Silben, sowie Korrelationen zwischen Sprachverständnisschwellen und Reintonaudiometrien, zwischen Sprachverständnisschwellen und Phonemerkennungsscores.

**Forschungsdesign:** Um die bereits genannten Ziele zu erreichen, wurden zwei unterschiedliche Arten von Sprachaudiometrie-Materialien entwickelt. Der adaptiv-auditive Sprachtest (AAST) besteht aus fünf Untertests, jeweils sechs einfache zweisilbige Wörter, die die individuelle Sprachverständnisschwelle messen sollen. NAMES enthält vier Listen von bedeutungslosen Zweisilbern (KV-KVK) jeweils 20 Items, die benutzt werden, um die Oberschwellenphonemerkennungsscores zu bestimmen. Die Phonemverteilung beider Sprachmaterialien ist ein Doppel der Phonemverteilung im Südvietnamesischen. Alle Sprachmaterialien wurden von einer weiblichen Stimme im südvietnamesischen Dialekt aufgenommen. Was AAST betrifft, ist die Sprachverständlichkeit gleich zwischen Wortstimuli innerhalb eines Untertests und bei unterschiedlichen Untertests während des Optimierungsverfahrens.

**Stichprobenstudie:** Von 435 Normalhörenden zwischen 4 und 85 Jahren, davon 117 nicht-muttersprachlichen Hörern des Dialekts wurden Sprachverständnisschwellendes AASTs gesammelt. Oberschwellenphonemerkennungsscores von NAMES wurden mit einer Stichprobe von 186 Normalhörenden im Alter von 15 bis 85 Jahren, davon 38 nicht-muttersprachlichen Hörern erfasst.

**Ergebnisse:** Bezüglich normativer Werte, sind altersabhängige normative Werte in vietnamesischen AAST ähnlich zu solchen in anderen Sprachen, in welchen der altersbezogene Unterschied ungefähr 8 dB zwischen jüngeren Kindern und 11 dB zwischen Erwachsenen und älteren Erwachsenen beträgt. Oberschwellenphonemerkennungsscores von NAMES sind relativ gleich unabhängig des Alters (außer die ältesten Hörer über 75 Jahren). Der Score des ältesten unterschied sich von den jüngeren Hörern mit ungefähr 3% des richtigen Phonemscores.

Was dialektale Effekte angeht, ist die Spracherkennungsschwelle von nicht-muttersprachlichen Hörern signifikant verschlechtert verglichen mit den muttersprachlich-südvietnamesischen Hörern.

Unterschiede in den Spracherkennungsschwellen reichen von 2 bis 6 dB für AAST im Ruhe - und von 1 bis 5 dB für AAST im Störgeräusch. Im Kontrast dazu, wurden nur kleine Unterschiede, die relevant für den Phonemerkennungsscore waren, zwischen muttersprachlichen und nicht-muttersprachlichen Hörern mit NAMES gefunden, eine Variation von 2% Richtigkeit des Scores.

Was Effekte der tonalen Struktur bei der Spracherkennung älterer Hörer betrifft, erzielten ältere Hörer ein ein wenig schlechteres Ergebnis, wenn Sprachstimuli Töne mit niedrigem Level tragen. Im Kontrast dazu, zeigen sie bessere Resultate, wenn Sprachstimuli Töne mit hohem F0 Level und steigender Kontur haben. Für AAST sind die durchschnittlichen Sprachverständnisschwellen in Ruhe vergleichbar bei den einzelnen Untertests von AAST im Kontrast zu den Durchschnittssprachschwellen bei Geräuschen, die heterogener sind, zum Beispiel ist der Durchschnittsschwellenwert AAST-a3 signifikant geringer als bei anderen Untertests. Schwellenwerte bei Geräusch sind schlechter, weil Sprachstimuli von AAST-a3 Töne mit tiefen F0 Level und fallenden Konturen, besonders bei solchen Stimuli, die älteren Hörern mit Geräuschen gegeben wurden, enthielten. Für NAMES hatten die älteren Hörer etwas bessere Resultate als die Stimuli einen hohen Tone (*ngang-A1*) erhielten, verglichen mit tief-fallend und hoch-steigenden Tönen (*huyền-A2* und *sắc-B1*). Vergleichbar mit den jungen Hörern misidentifizierten ältere Hörer Ton *huyền* (A2) sowie Ton *nặng* (B2).

Bezüglich der Beziehungen zwischen Sprachmaterialien und Reintonautonomie, gibt es eine starke Korrelation zwischen den Sprachverständnisschwellen und den Reintonautonomien (0,5 & 4kHz). Dennoch wird keine oder nur eine schwache Assoziation zwischen Sprachverständnisschwellen und Phonemerkennungsscore gefunden, zwischen Phonemerkennungsscores und Reintonautonomie.

**Schlussfolgerungen:** AAST und NAMES sind valide Sprachaudiometrietests um Spracherkennung für Hörer von 4 bis 85 Jahren (AAST) und 15 bis 85 Jahren (NAMES) zu quantifizieren. Die altersabhängigen normativen Werte von AAST sind vergleichbar mit denen des Deutschen, des Ghanaisch und des Polnischen. Die Ergebnisse der Dialektstudie zeigen, dass dialektale Variation einen Einfluss auf die Spracherkennung hat. Dennoch das Ausmaß der Effekte ist abhängig von den für die Hörmessungen verwendeten Sprachaudiometriematerialien. In diesem statistischen Sinne, beeinflussen Dialekte bei Verwendung von AAST mit semantischem Inhalt in einem angepasstem Prozess, die individuelle Sprachverständnisschwellen. Während zu vernachlässigende Phonemscores zu zeigen scheinen, dass die Dialekteinflüsse minimal sind, wenn man NAMES ohne semantischem Inhalt in einem konstanten Presentationslevel verwendet. Die Ergebnisse der Studie von Strukturtonen scheinen zu implizieren, dass die tonalen Struktur der Silben einen geringeren Einfluss auf die Sprachwahrnehmung auf ältere Erwachsene, besonders die ältesten Hörer über 75 Jahre. Schlussendlich können die Sprachverständnisschwellen durch die Reintonaudiometrie vorhergesagt werden. Im Gegensatz dazu können Phonemscores weder durch Sprachschwellen noch durch Reintonaudiometrie basiert auf Korrelation prognostiziert werden.

## TABLE OF CONTENTS

---

ACKNOWLEDGEMENTS.....	i
ABSTRACT.....	ii
ZUSAMMENFASSUNG.....	iv
LIST OF FIGURES.....	x
LIST OF TABLES.....	xii
ABBREVIATIONS.....	xiii
1. INTRODUCTION.....	1
2. REVIEW OF LITERATURE.....	5
2.1 Dialectal effects on speech audiometric testing.....	5
2.2 Vietnamese language and dialects.....	8
2.2.1 Vietnamese.....	8
2.2.2 Vietnamese dialects.....	9
2.3 Tonal perceptions of older listeners.....	19
2.4 Tones in Southern Vietnamese.....	22
2.5 Speech perception in older adults.....	23
2.5.1 Deficits in speech perception.....	23
2.5.2 Causes of speech perception deficits.....	24
2.6 Speech audiometry.....	25
2.6.1 Speech materials in Vietnamese.....	26
2.6.2 Adaptive Auditory Speech Test.....	27
2.7 Summary.....	28
3. DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE.....	31
3.1 Designs of a speech material of AAST.....	31
3.1.1 Phoneme frequencies in Southern Vietnamese—written text.....	32
3.1.2 Choices of disyllabic noun phrases.....	32
3.1.3 Phoneme frequencies.....	35
3.1.4 Picture drawings.....	37
3.1.5 Sound recordings.....	38
3.1.6 Preparations of pilot tests for AAST.....	39
3.2 Designs of the speech material of NAMES.....	49

3.2.1	Phoneme frequencies in Southern Vietnamese—in spoken text.....	49
3.2.2	Phoneme selections and nonsense disyllabic CV-CVC structures.....	49
3.2.3	Stimuli recordings .....	51
3.2.4	Phoneme categories in the speech material of NAMES .....	52
3.2.5	Preparations of the pilot test.....	53
3.3	Summary .....	54
4.	OBJECTIVES AND RESEARCH QUESTIONS .....	56
4.1	Objectives .....	56
4.2	Research questions and hypotheses .....	56
5.	METHODS.....	59
5.1	Speech test of AAST .....	59
5.1.1	Speech stimuli.....	59
5.1.2	Listeners .....	61
5.1.3	Test procedure.....	64
5.1.4	Data analyses .....	65
5.1.5	Exclusion of outliers .....	65
5.2	Speech test of NAMES .....	67
5.2.1	Speech stimuli.....	67
5.2.2	Listeners .....	68
5.2.3	Test procedures .....	69
5.2.4	Data analysis .....	69
6.	RESULTS.....	71
6.1	Results of the speech test AAST .....	71
6.1.1	Normative values .....	71
6.1.2	Effects of dialects on speech audiometry testing.....	75
6.1.3	Learning effects on speech recognition thresholds.....	79
6.1.4	Effects of F0 on the speech recognition thresholds .....	80
6.1.5	Response matrix and error analyses on AAST-a2.....	82
6.2	Results of the speech test NAMES.....	85
6.2.1	Normative values .....	85
6.2.2	Effects of response modes.....	90

6.2.3	Effects of dialects on speech audiometry testing .....	90
6.2.4	Effects of F0 on supra-threshold phoneme recognition scores .....	93
6.3	Correlation between speech audiometry materials and duo-tone audiometry .....	94
6.3.1	The correlation between SRTs and duo-tone thresholds .....	94
6.3.2	Correlation between SRTs (AAST-a2) and PRSs (NAMES).....	97
6.3.3	Correlation between PRS (NAMES) and duo-tone thresholds .....	97
6.4	Summary of the results.....	99
7.	DISCUSSION.....	101
7.1	Normative speech threshold values in Vietnamese AAST .....	101
7.2	Learning effects on AAST.....	102
7.3	Tonal effects on perception of speech in AAST and NAMES.....	103
7.4	Normative phoneme scores in Vietnamese NAMES.....	105
7.5	Effects of dialects on speech perception in relation to AAST and NAMES .....	107
7.6	Interdependencies of SRT, duo-tone threshold, and PRS .....	111
7.7	Summary of discussion.....	113
8.	APPLICATION STUDY OF AAST IN HEARING-IMPAIRED CHILDREN .....	115
8.1	Introduction.....	115
8.2	Methods.....	116
8.3	Results .....	117
8.4	Discussion.....	119
8.5	Conclusion .....	120
9.	GENERAL DISCUSSION .....	121
9.1	Evaluation of speech materials .....	121
9.2	Implications of the study.....	126
9.3	Limitations of the study.....	129
9.4	Future research .....	129
10.	SUMMARY AND CONCLUSION .....	131
11.	APPENDIX .....	134
Appendix A:	Frequency of occurrence of phonemes in Vietnamese .....	134
Appendix B:	Sublists of NAMES test with nonsense two syllables structures .....	136
Appendix C:	Descriptive statistics of normative values for native listeners.....	137

Appendix D: Statistical values of AAST-a2 for non-native listeners .....	139
Appendix E: Descriptive statistical values of SRT across subtests of AAST.....	141
Appendix F: Word confusion matrix of native listeners in AAST-a2.....	142
Appendix G: Word confusion matrix of non-native listeners in AAST-a2.....	143
Appendix H: Error rates and response matrix of native listeners for NAMES .....	144
Appendix I: Error rates and response matrix across dialectal groups.....	146
Appendix J: Distribution of individual values in SRTs and duo-tone thresholds .....	149
Appendix K: SRT values and duo-tone thresholds in different ambient noise levels .....	151
Appendix L: A comparison between AAST-Vietnamese and AAST-German .....	152
Appendix M: Children with the performances of AAST .....	153
12. REFERENCES .....	154

## LIST OF FIGURES

Figure 1: The average F0 contour of Northern and Southern Vietnamese tones.....	15
Figure 2: The frequency of consonants in each subtest of six AAST words compared with those in the written language .....	35
Figure 3: The frequency of vowel phonemes in each subtest of AAST in comparison with their counterpart in the written language.....	36
Figure 4: Test screens of the subtests of AAST, a1 to a4.....	37
Figure 5: Test screen of the subtest aTP .....	38
Figure 6: Psychometric curves of the subtests of AAST in quiet before the intensity adjustments; slope 7.5%/dB. ....	42
Figure 7: Psychometric curves of subtests of AAST in noise before intensity balances; slope at 50%=9.8 %/dB.....	43
Figure 8: Psychometric curves of the subtests of AAST in quiet after intensity adjustments; slope=8.2%/dB. ....	44
Figure 9: The psychometric curves of the subtests of AAST after intensity corrections in noise. The slope of 50% is roughly 8.4%/dB.....	45
Figure 10: The confusion analyses of the subtests a1 and a2.....	47
Figure 11: The confusion analyses of the subtests of a3, a4, and aTP .....	48
Figure 12: The flowchart illustrating the steps to create meaningless CVCVC syllables.....	50
Figure 13: Distribution of energies (total power) in the stimuli of NAMES.....	52
Figure 14: A screenshot of the NAMES test .....	53
Figure 15: Amplitude waveforms and F0 of speech stimuli, a1 to a3.....	60
Figure 16: Amplitude waveforms and F0 of the speech stimuli, a4 and aTP.....	61
Figure 17: Speech recognition threshold values in quiet condition across the different age groups of the southern listeners .....	74
Figure 18: Speech recognition threshold values in a noisy condition across the different age groups of the Southern listeners.....	74
Figure 19: The speech recognition threshold values presented for the non-native listeners and the native listeners in quiet, including groups of Northern.....	77
Figure 20: The speech recognition threshold values in noise illustrated for the non-native and native listeners.....	78

Figure 21: The reaction time examined only in the correct responses of the three groups of dialects 79

Figure 22: The speech threshold values in quiet (left) and noise (right) under learning effects on six-year-old children ..... 80

Figure 23: Speech recognition threshold values in quiet condition across the speech materials of AAST ..... 81

Figure 24: Speech recognition threshold values in noise condition across the speech materials of AAST ..... 82

Figure 25: The proportion of errors among the stimulus words for the native listeners ..... 84

Figure 26: Correlations between SRTs and duo-tone thresholds ..... 96

Figure 27: Correlations between the SRTs and PRSs ..... 97

Figure 28: The correlations between PRS and the duo-tone threshold ..... 98

Figure 29: Age-related norm values in AAST across languages: German, Ghanaian, Polish ..... 102

Figure 30: SRT values for hearing-impaired children, N=8 (N=15 ears) ..... 117

Figure 31: SRT values of AAST in quiet, correlated to aided duo-tone threshold of HA users ..... 118

Figure 32: SRT values of AAST in quiet, correlated to unaided PTA of HA users (Coninx, 2005) . 118

Figure 33: Psychometric curves for hearing-impaired children with hearing aids ..... 118

Figure 34: Effects of top-down lexical semantic and bottom-up acoustic processing of AAST and NAMES ..... 126

## LIST OF TABLES

Table 1: The 26 selected words in four subtests of AAST in Vietnamese, with six words each .....	34
Table 2: The frequency of consonants (percent) on the subtests of AAST compared with those in SVN .....	35
Table 3: The average and total RMS powers for the stimuli of AAST.....	39
Table 4: Mean SRT's for all five subtests of AAST .....	46
Table 5: Proportion of consonant phonemes in NAMES test.....	51
Table 6: Proportion of vowel phonemes in NAMES .....	51
Table 7: Number of outliers (ears) in the normative data of AAST-a2.....	66
Table 8: Number of outliers in dialect datasets .....	66
Table 9: Numbers of outliers across subtests of AAST.....	67
Table 10: Word confusion matrix for the groups of the native listeners, N=6000 stimuli.....	83
Table 11: Descriptive statistical values of the Southern listeners, N=268.....	86
Table 12: The response matrix and proportion of errors for the onset by the native listeners .....	88
Table 13: The response matrix and proportion of errors for vowel phonemes .....	88
Table 14: Error rates and response matrix for tone identification.....	89
Table 15: Error rates and response matrix for coda recognition.....	89
Table 16: Descriptive statistical values of PRS for effect of test responses .....	90
Table 17: Descriptive statistical values for three listener groups.....	91
Table 18: Descriptive statistical values of PRS in tonal identification task.....	93

## ABBREVIATIONS

AAST	Adaptive Auditory Speech Test
a1	AAST-a1
a2	AAST-a2
a3	AAST-a3
a4	AAST-a4
Atp	AAST-aTP
A1	High-level tone ( <i>ngang</i> )
A2	Low-falling tone ( <i>huyền</i> )
B1	High-rising tone ( <i>sắc</i> )
B2	Creaky falling ( <i>ngã</i> )
C	dipping ( <i>hỏi-ngã</i> )
BELLS:	Battery for the Evaluation of Listening and Language Skills
CVN:	Central Vietnamese (languages)
NVN:	Northern Vietnamese
SVN:	Southern Vietnamese
HL:	Hearing Level
dB:	Decibels
dBFS	Decibels relative to full scale
HA:	Hearing Aids
Hz:	Hertz
kHz:	Kilo Hertz
NAMES:	Nonsense syllable test
PRS:	Phoneme Recognition Score
PTA:	Pure-tone Audiometry
SNR:	Signal to Noise Ratio
SPL:	Sound Pressure Level
SRT:	Speech Recognition Threshold

## 1. INTRODUCTION

In hearing assessments, the dialectal variation among clients is a huge challenge for audiologists or speech pathologists. Dialectal differences can lead them to wrongly classify the levels, types, and configurations of hearing loss. In this sense, a good understanding of the client's linguistic background can help the clinician choose the appropriate speech materials to precisely evaluate the level of hearing loss as well as avoid misidentification.

Recently, researchers have shown an increased interest in issues of dialectal effects on speech audiometry testing (Nissen et al., 2013; Shi & Canizales, 2012; Le et al., 2007; Adank & McQueen, 2007; Schneider, 1992; Crews, 1990; Weisleder & Hodgson, 1989). In general, the literature on dialectal effects offers paradoxical findings regarding whether dialect significantly affects speech audiometric testing. Shi and Canizales (2012), Le et al. (2007), Adank and McQueen (2007), and Weisleder and Hodgson (1989) found a deterioration in speech performances and a delay in reaction time if the speech material was not in the listener's dialect. From this evidence, they argued that dialectal variations significantly influence speech audiometric testing (speech recognition thresholds, word recognition scores, and word processing). In contrast, Nissen et al. (2013), Schneider (1992), and Crews (1990) found little or no difference in the results when speech materials were derived from non-regional dialects. Eventually, they asserted that the dialect of listeners or speakers did not affect speech audiometry testing too much due to negligible differences in speech threshold/word scores among the native and non-native listeners of a particular dialect. Besides the contradictory findings, these previous studies have also shown several limitations. For example, they focused on a small number of listeners with a narrow age range (Schneider, 1992; Crews, 1990; Nissen et al., 2013). Therefore, a large number of listeners with a big age range are necessary to inspect the issue of dialectal effects on speech audiometry testing. Although this issue has been studied for decades in several languages, particularly in Spanish, Mandarin, and English, it is yet to be examined in Vietnamese.

In tonal languages, like Vietnamese, Cantonese, and Thai, listeners identify words depending on tonal patterns and phonetic structures. The identification of each tone depends on the change in the pitch height (F0 level) and the pitch contour (F0 direction) in which the tone appears in an individual syllable. The major cues to tonal identification are pitch height and pitch contour. The question of how the fundamental frequency (F0) affects the listener's speech perception as the tones' pitch levels change received considerable attention decades ago. Many studies on the lexical tone perception in

Vietnamese have focused primarily on young adults (Brunelle, 2013; Kirby, 2009; Brunelle & Jannedy, 2007 and 2009; Vũ, 1981 and 1982). The authors have suggested that the low tone with a falling contour is the most confused one among the six tones in Northern Vietnamese (NVN) and five tones in Southern Vietnamese (SVN). In addition, the speech stimuli carrying low-falling tones were recognized as relatively poorer than speech stimuli carrying high-rising tones. For older listeners, unfortunately, no reference data on tonal perceptions was collected. In other tonal languages, the findings of the perceptual study of tones in Cantonese (Varley & So, 1995), and Mandarin (Yang, 2015) showed that (1) the older listeners scored relatively worse in speech than the younger listeners, and (2) the low tones were more confusable than the high tones. In contrast to Varley and So (1995), Li and Thompson (1977) found that high-rising tones were more confusable in both production and perception than the falling tones. Their finding came from observing tonal processing in Mandarin-speaking children.

Speech audiometry materials have become a basic tool to evaluate hearing loss. In conjunction with pure-tone audiometry, speech materials can support the classification of the degree and type of hearing impairment. In Vietnamese, speech audiometry, unfortunately, is not available to measure hearing. A standard audiological assessment only includes otoscopic inspection, tympanometry, and pure-tone audiometry. However, the hearing ability of individuals relies not only on the ability to detect the auditory stimuli but also on the capacity to discriminate and identify speech stimuli. Therefore, the lack of speech audiometry materials in audiology assessments in Vietnamese is considered to be unfair for hearing-impaired listeners because they do not know how well they can identify speech stimuli at the lowest level. The present work developed the speech materials of AAST and NAMES firstly for the research purpose of measuring the hearing ability of normal subjects, and secondly, for their applicability for hearing-impaired children.

In clinical audiology, speech materials are usually used along with pure-tone audiometry. Therefore, it is necessary to know the extent to which speech thresholds or phoneme (word) scores can be forecast from pure-tone thresholds. In addition, the relations among pure-tone audiogram, speech recognition thresholds (SRTs), and phoneme (or word) recognition scores (PRSs) may illustrate the extent to which hearing impairment can be judged from pure-tone thresholds.

The themes we have mentioned above have so far not been at the centre of clinical investigations in Vietnam. This thesis focuses on five main topics: (1) effects of dialectal variations on speech perceptions, (2) effects of fundamental frequency (F0) of syllables on older native listeners' speech

perception, (3) the normative values of speech materials of AAST (in children, young adult, adults, and older adults) and NAMES (in young adults, adults, and older adults), (4) learning effects on the speech material of AAST, and (5) correlations among speech recognition thresholds (SRTs), phoneme recognition scores (PRSs), and duo-tone thresholds.

### **Significances of the Study**

The researchers of this study developed two speech materials to screen the speech threshold. Firstly, for research purposes and clinical needs, the two speech tests can be used to clinically measure the hearing ability (speech thresholds, or phoneme scores) of an individual, which audiologists and speech pathologists expect from audiology. The use of speech materials will improve the quality and practice of educational audiology in Vietnam. Secondly, the findings will reflect on the field of speech perception because dialect plays a significant role in speech recognition or speech discrimination in hearing assessments. This will help audiologists become more aware of the appropriateness of a speech material to test the speech threshold or phoneme score of their clients. Lastly, the findings of the current research offer some important insights into the linguistic areas of Vietnamese, for example, differences in speech perceptions across Vietnamese dialects, and the tonal and phoneme identification in older native listeners, which have got little attention so far.

### **Outline of the thesis**

The study has 10 sections, including this introductory one. The next section (Section 2) provides brief overviews of (1) the dialectal variations in Vietnamese, (2) the effects of dialects on speech recognitions, (3) the identification of tones by older adults in tonal languages, such as in Mandarin, Cantonese, Thai, and Vietnamese, and (4) features of speech perceptions by older adults. In this section, the phonemic and tonal features of Southern Vietnamese are compared with those in Northern and Central Vietnamese.

Section 3 gives a concise delineation of establishing the speech materials (AAST and NAMES), including the choice/development of lists of words, recording speech stimuli, and preparation of pilot tests. After outlining the research questions in Section 4, Section 5 deals with the methods that were used in the two experiments of AAST and NAMES. Section 6 provides the results of the current study in AAST and NAMES. This section focuses on the following five key themes: (1) the normative values in the SRTs and PRS scores in the two speech materials by the native listeners of Southern Vietnamese based on the listener's age, (2) the learning effects, (3) the SRTs and PRS

scores by the non-native listeners in the study of dialectal effects, (4) the SRTs and PRS scores in examining the tonal identification of older native listeners, (5) the association among SRTs, phoneme scores, and duo-tone thresholds. In addition, the analyses of the response matrix (AAST) and error patterns of phoneme identification (NAMES), and the effect of respond modes on phoneme scores are also provided in this section.

In line with the findings in Section 6, Section 7 provides a brief discussion of the implications of the findings to the six research questions raised in the “Objective” and “Research Question” sections. A brief critique of the findings is also provided with respect to the literature. Section 8 is not a central part of this thesis. It provides extra information regarding the application of AAST in hearing-impaired children. The study estimates the validity of AAST in terms of speech threshold values, speech intelligibility, and correlations between SRTs and duo-tone thresholds in a free-field condition by hearing-impaired children with hearing aids.

Section 9 discusses several related issues regarding the outcomes and addresses the contributions of the thesis to research issues within the areas of audiology and linguistics. Finally, Section 10 gives a brief summary and conclusion of the current work.

## 2. REVIEW OF LITERATURE

### 2.1 Dialectal effects on speech audiometric testing

The effects of dialects on speech recognition testing have been considerably examined so far (Weisleder & Hodgson, 1989; Schneider, 1992; Le et al., 2007; Adank et al., 2007; Shi & Canizales, 2012; Nissen et al., 2013). However, there are still two controversial opinions on the dialectal effects of auditory speech testing. Some authors—Schneider (1992), Crew (1990), and Nissen et al. (2013)—have found that dialectal varieties have a negligible impact on audiometric evaluation and might not alter clinical interpretations.

Schneider (1992) examined the effect of three Spanish dialects on SRTs in a group of 12 Spanish children aged between six and seven. The speech stimuli included 12 Spanish words recorded in Castilian, Caribbean, and Mexican dialects. Before the hearing test, the children were familiarized with the test to make sure that they all knew well the speech stimuli being used. In the test, the children had to point out one of four pictures that represented the speech stimuli. The result showed no significant differences in SRTs between the dialectal groups. In particular, four children had the same speech thresholds in the three dialects, and eight had different speech thresholds across the dialects. However, these differences were negligible (2 dB or less). Similar to Schneider's conclusion, Crews (1990) stated in his work that the dialects of speakers did not significantly influence listeners' speech performances. In the experiment, 20 children who spoke Montana dialect—all seven years old with normal hearing sensitivity—were recruited. The speech materials encompassed a list of 35 monosyllabic words (Phonetically Balanced Kindergarten—PBK) and a list of 25 items of nonsense syllables (Nonsense Syllable Test—NST). The speech materials were recorded by two female speakers. One spoke Montana dialect, and the other spoke the General Southern dialect. The children were not familiarized with the speech stimuli. The speech stimuli were presented via a monitored live voice at 55 dB HL (a normal conversational level). In the PBK test, the listeners scored similar results for the different dialects of the speakers—96.2% correct responses in Montana and 94.6% correct responses in the General Southern dialect. No significant differences were found in the functions of the dialectal effect of speakers. Interestingly, in the NST test, the Montana children performed significantly worse when the stimuli were presented in their own dialect. They performed significantly well when the stimuli were presented in the General Southern dialect. Based on these findings, Crews stated that differences in word recognition score (WRS) were not necessarily a result of dialect but stemmed from other factors. One was the pronunciation of the

speaker who presented the speech stimuli in Montana dialect. It made the stimuli less intelligible than the stimuli presented in the General Southern dialect.

If the research of Schneider and Crews focused primarily on the dialect of speakers, recent studies conducted by Nissen and his colleagues (2013) examined the effects of the dialects not only of the speakers but also of the listeners. Thirty-two Mandarin listeners, aged between 18 and 50 years, were recruited in the study. Half of the listeners spoke Mainland Mandarin while the other half spoke Taiwan Mandarin. All of them lived in the United States at the time of the study. The speech materials included SRT and word recognition (WR) speech tests recorded in both the Mandarin dialects. The SRT material included trisyllabic words whereas the WR included disyllabic words. The listeners were asked to repeat the speech stimuli as they heard those. For the SRT test, the listeners were familiarized with the speech materials but not with the WR material. The results showed that there were statistically significant differences for the dialectal effects relevant to speakers and listeners in the speech materials of both SRT and WR. Although the authors realized that a statistically significant difference existed, these differences were small (less than 2 dB). Hence, the authors concluded that such differences have an insignificant effect on clinical measurements.

These researchers found no significant effect of dialects on audiometric speech tests. The insignificant effect was probably due to some limitations in these studies. In the Crews research, speech stimuli were presented via a monitored live voice. So, the listener's scores might have been affected by factors other than linguistics, for example, speaker variability, speaker loudness, and articulation errors. As Mendel and Owen (2011) recommended, the stimuli of the speech materials should be recorded for reliable and accurate measurements. In Schneider's work, the sample size was limited (12 children). It may not have reflected fully and precisely what the author expected. In Nissen's work, he and his colleagues conducted the research in the United States. So, the listeners might have been reciprocally influenced by either Taiwan Mandarin or Mainland Mandarin as a result of language contact within a close community.

In contrast to the above-mentioned findings, several fascinating studies (Shi & Canizales, 2012; Le et al., 2007; Adank et al., 2007; Weisleder & Hodgson, 1989) have found that dialects play a major role in speech audiometric measurements. Weisleder and Hodgson (1989) validated the Spanish WR test and also examined the dialectal effects on the speech materials. Their study included four lists of 50 disyllabic Spanish words each. The listeners were 16 Spanish-speaking subjects aged between 20 and 49 years. The listeners came from different countries, such as Mexico (9), Panama (2), Venezuela (2),

Spain (1), Honduras (1), and Columbia (1). The speech stimuli were presented to the listeners at three different hearing levels (8, 16, 24, and 32 dB HL) via recorded materials using the voice of a male native Spanish speaker. The listeners were requested to repeat (or guess) the responses as the speech stimuli were presented. Regarding the dialectal effect, the authors found that the Mexican listeners performed significantly better than the others at the low-intensity hearing levels. However, at the high-intensity levels, there was no difference among the listeners. The authors stated that regional variations in Spanish might have contributed to the different performances among the listeners at the soft hearing level.

Le and her colleagues (2007) examined the effects of word frequencies and dialects on spoken word recognitions. Twenty-one psychology students aged between 18 and 39 years—all native listeners of Australian English—participated in the experiment. The speech materials included 18 monosyllabic and 18 disyllabic words. The speech stimuli were recorded in the three following dialects: Australian English, South African English, and Jamaican Mesolect English. Before the experiment, all the listeners were screened individually to make sure that all of them had minimal experiences of the non-Australian dialects being used. Le and her colleagues found significant differences in the listeners' WR performances. The native listeners of Australian English did progressively worse as auditory stimuli differed from their own dialects. Based on the finding, the authors suggested that phonological and phonetic varieties of a different dialect affect the word recognition process of native listeners more significantly.

Similarly, Shi and Canizales (2012) conducted a research with 40 native Spanish-speaking people who lived in New York City at the time of the study. They spoke one of two different Spanish dialects. Twenty listeners, who came from Colombia, Ecuador, Bolivia, and Peru, spoke the Highland dialect. The other 20, who came from Dominican Republic, Puerto Rico, and the coastal areas of Colombia, spoke the Caribbean/Coastal dialect. The listeners were also grouped according to their predominant language—either English or Spanish. The speech stimuli were four lists of 50 bisyllabic Spanish words each recorded by a male speaker of Mexican origin. The Spanish recognition words were imparted to the listeners in three different signal-to-noise ratios, particularly, SNR +6, +3 and 0 dB. The listeners were asked to orally repeat and write the word down as they heard it. The Highland listeners scored significantly better whereas the Caribbean/Coastal listeners scored poorly in the speech tests. The listener's dialect was found to have a significant impact on the score.

More recently, investigators have also examined the effects of unfamiliar dialects or accents on speech perception processing (Adank & McQueen, 2007). They found that listeners showed a delay in recognizing words when the speech stimuli were not in their own dialect. Thirty listeners, who were all native Dutch speakers aged between 18 and 49 years, and lived in the middle of the Netherlands, took part in the study. The listeners were divided into two groups of 15 members each. The first group was asked to assess a familiar accent (Local Dutch). The second one was requested to recognize an unfamiliar accent (East Flemish). The speech stimuli were 200 Dutch nouns spoken in those two accents. The results showed that the listeners had a delay of approximately 100 ms in processing the words spoken in the unfamiliar accent.

The findings of Shi and Canizales (2012) and Weisleder and Hodgson (1989) were different from those of Crews (1990), Schneider (1992), and Nissen et al. (2012) regarding the effects of dialects on speech recognition. The differences among these interesting works might stem from dissimilarities in methodologies and the listeners' age. For the methodologies, the speech stimuli were presented via recorded material in Weisleder's study and via monitored live voices in Crews's study. Besides, the presentation levels of speech stimuli also differed across the studies, leading to inconsistency in findings. Even the listener's ages were different. While the listeners were seven-year-old children in the research of Schneider and Crews, they were adults in the works of Weisleder et al. and Shi et al.

This section began with a summary of the effects of dialects on speech recognition. The findings suggest that dialects strongly affect speech recognition. Based on these findings, the researchers emphasized that audiologists should use speech material suited to a client's linguistic background to make hearing assessments (Shi & Canizales, 2012; Weisleder & Hodgson, 1989; Carhart, 1965). As mentioned earlier, using inadequate speech materials means increasing hindrances and decreasing client's capacities of speech recognition, leading to misidentifications in clinical interpretations. The phonetic distance between dialects might contribute to declines in speech intelligibility for the clients who are non-native speakers of the languages being tested. The following brief report will provide phonetic-phonemic distances across dialects of the Vietnamese language.

## **2.2 Vietnamese language and dialects**

### **2.2.1 Vietnamese**

Vietnamese is a language spoken by around 94 million people in Vietnam and above 3 million expatriates inhabiting mainly in Cambodia, the United States, Canada, and European countries. Worldatlas.com (June 2016) ranked Vietnamese as the sixth common language spoken at home in

the United States, which accounted for around 1.4 million speakers. Besides, the language is also taught in some universities in the United States, Australia, Japan, and Germany.

With respect to the origin of Vietnamese, the language belongs to the Vietic branch of the Austroasiatic language family. It is the largest spoken Austroasiatic language and typologically varies from its Austroasiatic neighbours due to intensive contacts with Chinese in terms of lexicon and phonology (Brunelle, 2015). For example, the currently used Vietnamese lexicon has roughly 70% loanwords from Chinese, which are called Sino-Vietnamese. Vietnamese has also been influenced by other languages, such as French and English. With respect to French (colonial era) and English, scientific and technical terms became a part of the lexicon of Vietnamese. These terms are accepted in two different ways. They were either retained their original word form or were transcribed into Vietnamese.

Regarding phonology, Vietnamese syllables include five phonemic components: onsets, medial glides -w-, nucleus, codas, and tones. These elements can occur in eight following syllable structures: (1) V, (2) wV, (3) VC<sub>2</sub>, (4) wVC<sub>2</sub>, (5) C<sub>1</sub>V, (6) C<sub>1</sub>wV, (7) C<sub>1</sub>VC<sub>2</sub>, (8) C<sub>1</sub>wVC<sub>2</sub> (Vương & Hoàng, 1994). The syllable structures can be generalized into a formula (C<sub>1</sub>)(w)V(C<sub>2</sub>). Tones are suprasegmental phonemes that combine with components of an optional medial glide, w-, a monophthong or diphthong, or a coda to create a rhyme. Recently, some researchers have proved that Vietnamese tones do not affect onsets (the initial position), but influence only the final part of a syllable (Trần et al., 2005). In tonal languages, changes in tones or pitches of sounds help discriminate semantics and identify words. A classic example is a consonant-vowel combination *ma* in Vietnamese. The monosyllabic word *ma*, which is grounded upon the tone patterns of vowel /a/, refers to six objects, for example, *ma* “ghost” (*ngang*-level), *mà* “but” (*huyền*-low falling), *má* “mother” ( *sắc*-high rising), *mạ* “rice seedling” ( *nặng*-low glottalized), *mả* “tomb” ( *hỏi*-dipping-rising), and *mã* “horse” ( *ngã*-high-rising glottalized).

### 2.2.2 Vietnamese dialects

The Vietnamese dialects are traditionally based on geographical divisions, including the three following main dialects: Northern, Central, and Southern (Hoàng, 2009; Hwa-Froelich et al., 2002; Vương & Hoàng, 1994; Thompson, 1991). However, the linguistic reality is far more complex (Brunelle, 2015). The Northern Vietnamese (NVN) is represented by the Hanoi speech, which is used widely in media and schooling. Therefore, Kirby (2010) stated that listeners from the other dialects might be somewhat experienced in the Northern dialect. The Saigon dialect is regarded as a

standard speech for the Southern Vietnamese (SVN) spoken by people living in Saigon and other cities surrounding Saigon (Cao and Lê, 2005; Huỳnh, 1999). Due to political and economic issues, the Saigon speech is a convergence of the different dialects—the Northern and Central (Bùi, 2009). Regarding the Central Vietnamese dialect (CVN), scholars consider the Hue speech as its representative in contrast to Hoàng (2009), who proposed Nghệ An or Thanh Hóa as the representative. This is probably due to geographical reasons. Huế province is located towards the south. So, the Huế speech shares some features with the SVN. As a result, speakers of Huế and Đà Nẵng can easily communicate with each other without any misunderstanding. The differences across the three dialects will be mentioned further in the following parts.

According to Alves (2010) and Huỳnh (2014), remarkable differences exist among three Vietnamese dialects in terms of the variations in lexicons, segments, and tones. Regarding the segments, the elements (onsets, nucleus, medial glide -w-, and codas) of a syllable structure across the dialects are a central point in this report. With respect to the suprasegments, the tone system across the dialects, including ngang (A1), huyền (A2), sắc (B1), nặng (B2), hỏi (C1), and ngã (C2), will be examined. Finally, the lexical differentiation among dialects will also be depicted in this section.

### **Initial phonemes across the dialects**

The standard Vietnamese phonological system includes 21 consonant phonemes in the initial position, namely, /b, m, f, v, t, t', d, n, z, z', s, s', c, t, ɲ, l, k, χ, ŋ, ʎ, h/ (Vương & Hoàng, 1994). The initial consonant /p/ is excluded from this report because the phoneme is considered to be a foreign phoneme stemming from French (*pin* “battery”, *pê-dan* “pedal”) and does not exist in the initial consonant system in Vietnamese (Đỗ & Lê, 2005). Across dialects, the initial system also varies, not only by the number of phonemes but also by the phonological features. In the NVN, barring three retroflexes /ʂ, z', t'/, there are only 18 initial phonemes (Pham, 2009; Hoàng, 2009). But the SVN has 21 phonemes containing these retroflexes (Huỳnh, 1999). The initial phoneme system in the Southern-Central dialect (Huế, Quảng Trị) is resemble to that in the SVN. However the Northern-Central has 23 initial phonemes and includes the phoneme /p/ (Alves, 2007).

Due to a lack of the retroflexes /ʂ, z', t'/ in NVN, speakers constantly pronounce /s/, /z/, and /c/ alike, in contrast to the rest of the dialects in which speakers can distinguish among these retroflexes.

Orthography	NVN	CVN	SVN	Glossary
<i>sông</i>	/soŋ/	/ʂoŋ/	/ʂoŋ/	“river”
<i>sát</i>	/sat/	/ʂat/	/ʂak/	“very close”
<i>tre</i>	/cɛ/	/tɛ/	/tɛ/	“bamboo”
<i>trời</i>	/cɤj/	/tɤj/	/tɤj/	“god”
<i>trầu</i>	/zǎu/	/zǎu/	/zau/	“vegetable”
<i>rõ</i>	/zɔ/	/zɔ/	/zɔ/	“clear”

In the SVN, a semi-educated speaker may make mistakes when he/she pronounces the retroflex phonemes, for example, common words like *sáu*—*xáu* “six”, *trà*—*chà* “tea”, and *ruợu*—*gượu* (*gư*) “wine”. In close relationships—for example, kinship or friendship—a well-educated speaker also allows himself to pronounce these in such ways. In workplaces, the native speakers of SVN, however, are more aware of their speech and enunciate more accurately. Additionally, the onset phoneme /v/ does not occur in the SVN. It is replaced by /j/ (Thompson, 1991), which sounds similar to /j/ (yes) in English.

The onset /j/ in SVN dialect not only substitutes the phoneme /v/ but also /ʒ/ (*gi*, and *d* in orthography) in standard Vietnamese phonemes (Phạm, 2006). The phoneme /z/ varies within the CVN dialect. Huế accent (Southern-Central area) is a part of the Central dialect. However, due to language contacts with the Southern dialect, the onset /z/ is also identified as the semivowel /j/ (Vương, 1992), just like their counterpart in the SVN. On the other hand, in the Nghệ An accent (Northern-Central area), the /z/ is recognized in the same way as that in the NVN (described in graphemes “gi” and “d”). Back to the phoneme /v/, it is represented by /j/ in the Southern-Central speech in contrast to the Northern-Central speech and NVN, which remains unchanged. Below are some examples:

Orthography	Saigon	Hanoi	Vinh	Hue	Glossary
<i>dạy</i>	/jai/	/zǎi/	/zǎi/	/jai/	“to teach”
<i>gió</i>	/jɔ/	/zɔ/	/zɔ/	/jɔ/	“wind”
<i>già</i>	/ja/	/za/	/za/	/ja/	“be old”
<i>về</i>	/je/	/ze/	/ze/	/je/	come back

### Medial glide /-w-/

According to Phạm (2016), the medial glide /-w-/ shares an identical feature between the NVN and CVN. However, this phoneme is more variable in the SVN. The medial glide /-w-/ in SVN represents *o* and *u* in the orthographic form and has variations depending on the context. For example, a syllable structure (CwV) contains medial glide /-w-/. However, in daily speech, the initial is either non-existent (wV) or is preceded by the onsets /h, k, x, ɲ/ at the beginning of a syllable. There are “two patterns in this simplification process” (Pham, 2008). In the first pattern, the onset is deleted, and the /-w-/ remains. For example, *oa* “sound of cry” is *wa*, *hoài* “constantly” is *wài*, *hoa* “flower” is *wa*, and *qua* “pass” is *wa*. In the second pattern, the medial glide /w/ is completely deleted. The consonant remains but completely changes into another phoneme, for instances, *khỏe* “healthy” becomes *phẻ*, and *khoan* “drill” becomes *phang*. Another noticeable alteration is that the medial glide /-w-/ can shift to the nucleus when the medial glide stands after phoneme /a/ in several combinations. In this case, instead of vowel /a/, SVN speakers tend to eliminate the vowel /a/, and replace it with /-w-/ due to complexities of pronunciations. For example, (*cái*) *loa* “loudspeakers” becomes *lo*, *toa* (*xẻ*) “wagon” becomes *to*, and *toan* (*toan tím*) “to intend” becomes *ton*.

In addition, the medial glide /-w-/ disappears after the initial / t, t<sup>h</sup>, t̚, c, s, ʃ, l, ɲ, j/ in several SVN regional accents (Vương & Hoàng, 1994). For example, *tuệ*—*tẻ*, *thuê*—*thẻ*, *chuyên*—*chiẻn*, *truyền*—*triẻn*, *xuyên*—*xiẻn*, *lòe loẻt*—*lẻ lẻt*.

As we have seen, the medial glide /-w-/ has many variations among the Southern speakers, which reflects the diversity of this dialect. In contrast, the /w/ in the NVN or the CVN retains phonetic properties or undergoes slight changes compared with standard Vietnamese. In the Hue accent, the /-w-/ occurs when the syllables contain the nucleus /ɔ/ and end with a semi-vowel /-j/ (Vương, 1992). As a result of this combination, the nucleus /ɔ/ is altered by the nucleus /a/ in conjunction with the medial glide /-w-/. Hence, *nói* “to talk” becomes *noái*, and *đói* “hungry” becomes *đoái*.

### Nuclei across Vietnamese dialects

The nucleus system of standard Vietnamese comprises 11 monophthongs, namely /i, u, e, ə, ɔ, o, ɛ, ǎ, a, ɔ/; and three diphthongs /i\_ə, u\_ə, u\_ə/. Each diphthong represents two orthographic forms depending on the context, for instance, *mưa* /mɯ\_ə/ “to rain” and *cười* /kɯ\_əj/ “to laugh” (Pham, 2008; Vương & Hoàng, 1994). Some notable features of discriminations of the nuclei will be shown in the following part.

Firstly, the diphthongs /iə, uə, uə/ are shifted into long vowels /i:, w:, u:/ in the SVN, whereas they are distinguished separately from the long vowels in the CVN. In the NVN, these diphthongs behave similar to those in the CVN, except for the diphthong /uə/, which is produced similarly to diphthong /iə/. For example, *ru* “wine” is *riəu*, and *tr* is *ciu* “to help”. In the CVN, the Vinh and Huế listeners can distinguish clearly between two diphthongs /uə, and iə/. However, it has to be noted that the second element of each diphthong is shorter in length and narrower in openness than their counterpart in the NVN dialect. Because of this feature, Thompson (1991) mentioned that the three diphthongs in Hue behave like those in the Southern dialect, which become sequences of long vowels /i:/, /u:/, /u:/. But these vowels are somewhat different from their counterparts in the Southern speech.

Secondly, the vowels /ɛ, e/ behave like /i/ in the SVN when these are combined with codas /-m, -p/. For example, *bép*—*bíp* “kitchen”, *xép*—*xíp* “to fold”, *kém*—*kìm* “pincer”, *chém*—*chim* “to wedge”. In contrast, the NVN and CVN discriminate clearly between the vowels /e/ and /ɛ/.

Thirdly, the Southern speakers do not distinguish /o/ from /ɔ/ when the vowels are followed by the codas /-ŋ and -k/. For example, *trông*—*trong* “inside”, *không*—*khong* “no, without”, *độc*—*đoc* “poison”. In contrast, the Northern and Central speakers can differentiate /o/ from /ɔ/ quite well.

Next, in conjunction with the final semi-vowels /-j/ and /-u/, the Southern speakers use the short vowel /ă/ like a long one /a/. For example, *tay*—*tai* /taj/ “hand”, *lau*—*lao* /lau/. Speakers of the other dialects can distinguish between these. This does not mean that the short vowel /ă/ does not exist in the Southern dialect. It occurs, but in a specific distribution (Pham, 2008), for example, *căm*—*căm* /kăm/ “to resent”.

Finally, and quite interestingly, the southerners do not distinguish between a pair of short vowels /ă/ and /ɤ/. For example, *ân nhân*—*ăn nhân* “benefactor”, *cân*—*căn* “to weigh”. In contrast, the Northern and Central speakers can distinguish between those. Below are some typical examples to show the noticeable dissimilarities in terms of the nuclei across dialects.

Orthography	Southern	Northern	Central	Glossary
<i>tiếp</i>	/ti:p/	/ti.ɛp/	/ti.ɛp/	“to continue”
<i>mướt</i>	/mwp, məp/	/mu.ɛp/	/mu.ɛp/	“luffa”
<i>rượu</i>	/zu, zu, yu/	/zieu/	/zu.ɔu/	“wine”
<i>nép</i>	/nip/	/nep/	/nep/	“rice”
<i>êm</i>	/im/	/em/	/em/	“smooth”
<i>hai</i>	/haj/	/haj/	/haj/	“two”
<i>hay</i>	/haj/	/hăj/	/hăj/	“great”
<i>hao</i>	/kaw/	/kaw	/kaw/	“tall”
<i>cau</i>	/kaw/	/kăw/	/kăw/	“areca”
<i>sân</i>	/sãŋ/	/sɔ̃n/	/sɔ̃n/	“courtyard”

### Codas across Vietnamese dialects

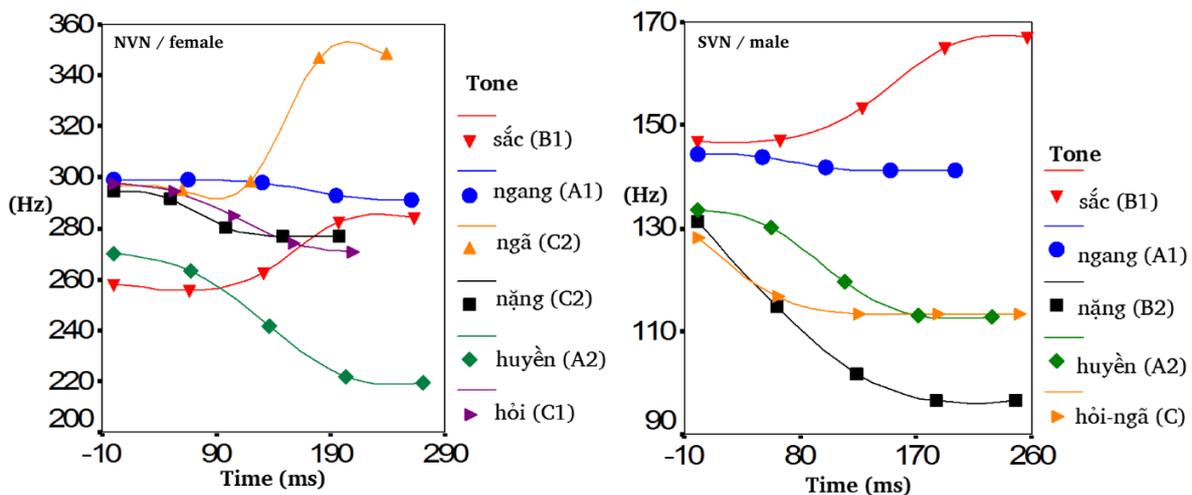
The system of codas in Vietnamese is restricted. It includes only six consonants /p, t, k, m, n, ŋ/ and two semivowels /w, j/ (Phạm, 2008; Vương & Hoàng, 1994). Examining the distinction between codas, Phạm (2016) stated that the six consonants are homogenous across the dialects. But in actual uses of the language, speakers of SVN reflect a variety of the two velar codas /-k and ŋ/ in their speech. They cannot distinguish well between the alveolar codas /-t/ and /-k/, and /-n/ and /-ŋ/. Basically, the phonemes /-t and -n/ coalesce into /-k/ and /-ŋ/ respectively in SVN. In contrast, speakers of NVN and CVN can distinguish between these codas and are aware of pronouncing these precisely. Due to language contacts, Hue speakers produce these exactly like the speakers of SVN. In a perceptual study on word-final stops in Vietnamese conducted by Tsukada et al. (2006) with speakers of two dialects, SVN speakers scored significantly worse in their discrimination of final stops /-t and -k/ than the NVN speakers did. In terms of perception tasks, the SVN speakers were less accurate in the production of these codas than the Northern ones. Although the codas /-n, -t/ are indistinguishable, it does not mean that the codas /t/ and /n/ do not exist in SVN and the Hue accent. These occur in a few certain contexts. For example, the /-t and -n/ come after the front vowels /i/ and /e/ (Pham, 2008; Vuong, 1992; Thompson, 1991), especially in the flowing rhymes: *inh, ích, ênh, êch, anh, ách*.

Orthography	SVN	NVN	CVN	Glossary
<i>in</i>	/u:n/	/in/	/in/	“print”
<i>ít</i>	/u:t/	/it/	/it/	“few”
<i>hình</i>	/hu:n/	/hiŋ/	/hiŋ/	“image”
<i>ích</i>	/it/	/ik/	/ik/	“useful”
<i>bệnh</i>	bə:n	/beŋ/	/beŋ/	“disease”
<i>ếch</i>	/ə:t/	/ek/	/ek/	“frog”

The above instances show that /n and t/ remain in SVN. However, its occurrence leads to a turning in the qualities of the vowels. Because of this, the phonemes /i, e, and a/ are produced as /ɯ, ə, and ǎ/ respectively. Apparently, there are no one-to-one correspondences between orthography and its sound when it is combined with codas /-n, -t/ in SVN.

### Tone across the dialects

SVN has six tones in orthography, including *ngang* “high level” (A1), *huyền* “low falling” (A2), *sắc* “high rising” (B1), *nặng* “low glottalized” (B2), *hỏi* “dipping-rising” (C1), and *ngã* “high-rising glottalized” (C2). Each tone is marked by a corresponding diacritic (except tone A1). According to Pham (2008), only about a third of the Vietnamese population conveys all the six tones in speech. The rest use only five or even four tones in their daily speech, especially those who originally spoke the North-Central dialects (Vuong, 1992). Previous phonetic studies have investigated the fundamental frequency of Vietnamese tones (Michaud, 2004; Kirby, 2010; Brunelle, 2009; Pham, 2003; Nguyễn & Edmondson, 1998; Alves, 1997). These studies reveal that the heights of F0 and voice qualities are major acoustic parameters that characterize tones in Vietnamese.



**Figure 1:** The average F0 contours of Northern and Southern Vietnamese tones, modified from Brunelle and Jannedy (2007). The F0 values for NVN are based on a female voice, whereas the F0 values for SVN are based on a male voice

Across dialects, the tones differ not only on perception but also on acoustics (Brunelle, 2009). Acoustically, features of pitch height and pitch contours bring distinctions into the tone systems among dialects (Figure 1). The Southern dialect has only five tones—A1, A2, B1, B2, and C1 (Figure 1, right). Tone C2 never appears in this dialect due to its coalescence with C1 (Hoang, 2009; Pham,

2008; Vương & Hoàng, 1994; Thompson, 1991; Vũ, 1982). Vũ (1981) and Kirby (2010) stated that the system of five tones is based primarily on pitch height and contours, and do not differ in terms of voice quality as compared with those in NVN. The Northern dialect has six tones (Figure 1, left), like the figure of tones in standard Vietnamese. Speakers of NVN, for example, can distinguish pitch contours between the C1 and the C2 separately, in which the C2 is a high-pitch direction and the C1 is a low one (Vương & Hoàng, 1994). In contrast to SVN, the six tones in NVN contrast each other in terms of both pitch contour and voice quality (Brunelle, 2009; Alves, 1995).

The tonal system of the Central dialect includes only five tones, of which C1 and C2 merge into a single one. Across the accents of the Central dialect, the tonal systems are more variable (Vũ, 1982). So, the tonal assimilation differs. The accent of Vinh also has five tones (Pham, 2005; Vương, 1992), of which C2 never merges into C1 but they coalesce into a single form as B2. Interestingly, some accents have only four tones, such as the ones spoken in Quảng Bình and Thanh Hóa. The Thanh Hóa accent has only tones A1, A2, B1, and B2 because the tones C1 and C2 merge with B1. The Quảng Bình accent also has four tones but these are completely different from those in Thanh Hoá. In Quảng Bình, C2 merges with B2, and C1 coalesces with B1. In Huế accent, tones C2 and C1 coalesce into B2, and B1 behaves like tone C1. Below are some extensive examples, illustrating the differences of the tonal system among dialects.

Orthography	Saigon	Hanoi	Vinh	Hue	Glossary
<i>cũ</i>	<i>cũ</i>	<i>cũ</i>	<i>cũ, ɔũ</i>	<i>cũ, ɔũ</i>	“old”
<i>đỏ</i>	<i>đỏ</i>	<i>đỏ</i>	<i>đỏ</i>	<i>đỏ</i>	“red”
<i>nó</i>	<i>nó</i>	<i>nó</i>	<i>nó</i>	<i>nỏ</i>	“he, she, it”

Compared with the two remaining dialects, tones of the CVN are different from those in NVN and SVN regarding the averages of F0 values or laryngealization (Vũ, 1981).

Due to these acoustic differences, non-native listeners misperceive the tones from unfamiliar dialects more than the tones from their own familiar dialects (Brunnel & Jannedy, 2013). Regarding the perceptual aspect, too, the way the listeners identify tones differ among the dialects. Brunelle (2009) conducted a study on two groups of speakers—the Northern and Southern. The listeners of both groups identified tones using stimuli from NVN. The results showed that the listeners who spoke NVN identified the tones correctly in their own dialect, which was not surprising. In contrast, the Southern listeners found it more challenging, as they misidentified the tones B2, C1, and C2 in the NVN. The tone B2—a low and glottalized tone—was identified as A2, which is a low-falling tone in

Southern. The C1—a low falling-rising—was misidentified as B2, a kind of low-falling tone in SVN. The C2—a high falling-rising and glottalized tone—was perceived as either the high-rising tone B1 or the high falling rising C1 in SVN. From these findings, the author concluded that the listeners of the SVN did not use voice qualities as a primary cue for tonal identifications. They depended mainly on the pitch contours for this task. On the other hand, the NVN speakers used both pitch contours and voice qualities as hints for tone identification. Interestingly, the contour does not play a similar fundamental function in the tone of NVN and SVN (Pham, 2005).

### Lexical Variations

Each dialect has a different developmental origin. The SVN, the youngest<sup>1</sup> of the dialects, was founded nearly five centuries ago (Hoàng, 2009). It was born of a synthesis of Southern Chinese dialect, and speeches of Cham and Khmer, and a local dialect spoken by the immigrants from Phú Yên to Bình Định provinces in early 18<sup>th</sup> century (Mika, 2013; Hoàng, 2009). In contrast, both NVN and CVN have a longer period of development. The CVN retains archaic features of phonology and lexicon. The NVN borrowed a great number of words from Chinese, known as Sino-Vietnamese and widely used till date. Apparently, the various origins led to differences across the dialects not only in terms of sound systems but also lexicon. Alves (2012) also stated that the grammatical vocabulary of the CVN differs from standard Vietnamese in terms of phonetic changes and archaic words. His fascinating work showed several cases of phonological correspondence patterns. For instance, the diphthong /uə/ in CVN becomes the monophthong /a/ in standard Vietnamese, e.g. *đường* vs *đàng* “street”. Then, the vowel phoneme /i/ in Central becomes the rhyme /ăj/ in standard Vietnamese, e.g. *mi* vs *mày* “you”. Then again, the rhyme /uj/ in Central becomes /oi/ in standard Vietnamese, e.g. *tui* vs *tôi* “I, me”. Some archaic grammatical words are completely different for standard Vietnamese and across the dialects, such as, *bao lăm* (CVN) vs *bao nhiêu* (standard, NVN, and SVN), “how much”; *nặng* (CVN) vs *thường, hay* (standard, SVN), “often”; *mụ* (CVN) vs *bà* (standard), an older woman. Below are some illustrations for lexical differences across the dialects.

Firstly, there are some differences in demonstrative and interrogative pronouns.

---

<sup>1</sup> The southern boundary of Vietnam was extended in the 17<sup>th</sup> century and remains unchanged. The Southern speakers originated from south-central provinces (Phu Yen, Binh Dinh) and groups of immigrants who came from China, Khmer, and Cham.

	SVN	NVN	CVN	Glossary
1	<i>này</i>	<i>này</i>	<i>ni</i>	“this”
2	<i>đâu</i>	<i>đâu</i>	<i>mô</i>	“where”
3	<i>sao</i>	<i>sao</i>	<i>răng</i>	“how”
4	<i>vậy</i>	<i>vậy</i>	<i>nữa</i>	“so, too”
5	<i>bao nhiêu</i>	<i>bao nhiêu</i>	<i>bao lăm</i>	“how much”
6	<i>thường, hay</i>	<i>thường, hay</i>	<i>nặng</i>	“often”
7	<i>Đó</i>	<i>kia</i>	<i>tê</i>	“that”
8	<i>tui</i>	<i>tôi</i>	<i>tau</i>	“I”
9	<i>mày</i>	<i>mày</i>	<i>mi</i>	“you”
10	<i>ông</i>	<i>ông ấy</i>	<i>ông nó</i>	“elder man”
11	<i>bà</i>	<i>bà ấy</i>	<i>mụ nó/ mẹ nó</i>	“elder woman”
12	<i>cô</i>	<i>chị ấy</i>	<i>o nó</i>	“she”
13	<i>anh</i>	<i>anh ấy</i>	<i>eng nó</i>	“he”

As the examples show, a remarkable feature of the Southern speech is phonetic changes to the third personal pronoun *anh ấy*, *chị ấy*, *ông ấy*, *bà ấy*, (see cases 10 to 13). In these grammatical words, the demonstrative pronoun *ấy* is eliminated. The tone C1 replaces the tones A1, B1, and A2, and finally takes a single form as *anh*, *chị*, *ông*, *bà*. The grammatical word of the SVN, compared with the NVN, is apparently similar (cases 1 to 6). However, it is different from those in the CVN. This difference is a result of the ancient origin that is still evident in the CVN.

Secondly, content words, which used to label the names of objects, apparently differ among the dialects, particularly in labelling non-living and animated objects. For instance, each dialect has a different way to indicate the relationship of a man with his child. For example, the words *bố*, *thầy* “father” are commonly used in both CVN and NVN, whereas Southern talkers prefer to call their father as *cha*, or *ba* (*tía* stems from the Khmer language, and rarely used in several provinces in the Mekong Delta). The word *nén* “candle” is used in the NVN, but a CVN speaker would prefer to use *đèn sáp*. The SVN speaker will call it *đèn cầy* instead. The word *bổ* “tiger” is spoken by the northerner, whereas the southerner uses *ợp*, and the central speaker calls it *khá*. Below is a list of words made by Hoang (2009) and synthesized through the dialect map by Kondo (2013).

NVN	SVN	CVN	Glossary	NVN	SVN	CVN	Glossary
<i>ngô</i>	<i>bắp</i>	<i>bắp</i>	“corn”	<i>lợn</i>	<i>heo</i>	<i>heo</i>	“pig”
<i>quả</i>	<i>trái</i>	<i>trái</i>	“fruit”	<i>hoa</i>	<i>bông</i>	<i>hoa</i>	“blossom”
<i>cá quả</i>	<i>cá lóc</i>	<i>cá trầu</i>	“fish”	<i>roi</i>	<i>mận</i>	<i>đào</i>	“plum”
<i>mì chính</i>	<i>bột ngọt</i>	<i>vi tinh</i>	“glutamate”	<i>chè</i>	<i>trà</i>	<i>chè</i>	“tea”
<i>xe đạp</i>	<i>xe đạp</i>	<i>xe đạp</i>	“bicycle”	<i>ô-tô</i>	<i>xe hơi</i>	<i>xe con</i>	“car”
<i>thìa</i>	<i>muỗng</i>	<i>thìa</i>	“spoon”	<i>bát</i>	<i>chén</i>	<i>đọi</i>	“bowl”
<i>thuyền</i>	<i>ghe</i>	<i>nóc</i>	“ship”	<i>nép</i>	<i>đèn cây</i>	<i>đèn sáp</i>	“candle”
<i>phong bì</i>	<i>bao thư</i>	<i>bì thư</i>	“envelope”	<i>chăn</i>	<i>mền</i>	<i>chăn</i>	“blanket”
<i>tất</i>	<i>vớ</i>	<i>tất</i>	“shocks”	<i>màn</i>	<i>mùng</i>	<i>màn</i>	“mosquito net”
<i>súp lơ</i>	<i>bông cải</i>	<i>bông cải</i>	“broccoli”	<i>bóng</i>	<i>banh</i>	<i>banh</i>	“ball”

Due to the various origins, there are many inequalities among the dialects in labelling an object. Instead of *mận* “plum” in the SVN, the Northern speakers call it *roi*, whereas the Central speakers name it *đào*. A loanword *súp lơ* (stems from the French “chou-fleur”), is transcribed correspondently *súp lơ* in the NVN. The remaining dialects prefer to use the native term *bông cải* for it. The Southern and several parts of the Central use *trái banh* for “ball”. However, it is called *quả bóng* in NVN. In daily conversation, these terms lead to some confusion for the non-native listeners of a dialect, or those unfamiliar with these terms. Fortunately, the mass media has gradually shortened the lexical or phonological distances across the dialects. Because of this, the non-native speakers of a dialect might understand better what native speakers from another dialect speak. In addition, some typical words occurring in the dialects have become common lexical items. These are used widely through the entire territory. For example, *sầu riêng* “durian”, *mãng cầu* “mangosteen”, *chôm chôm* “rambutan”, stem from the SVN. But these words have now turned into the common lexical items in Vietnamese.

In short, the three dialects differ from each other to some degree in terms of lexical and phonological aspects. Phonologically, each dialect has slight differences in terms of onset, medial glide, nucleus, coda, and tone. Lexically, each dialect also has some object names that are completely distinct from the others. It probably leads to confusion for the non-native listeners of a dialect. For the purpose of the current study, some remarkable features were synthesized among the three dialects of Vietnamese. The following section considers the tone identification of older listeners in tonal languages.

### 2.3 Tonal perceptions of older listeners

A considerable amount of literature has been published on tonal perceptions and productions in older adults for Cantonese, Mandarin, and Thai (Yang, 2015; Kasisopa, 2015; Varley & So, 1995). The studies first compared the abilities of young adults and older adults to distinguish the tone.

Secondly, they determined the lexical tones that are more confusable or troublesome to distinguish for both groups of listeners. In the Vietnamese language, previous research concentrated primarily on tone perceptions in young adults only (Brunelle, 2013; Kirby, 2009; Brunelle & Jannedy, 2007 and 2009; Vũ, 1981 and 1982). These found acoustic and perceptual differences in Vietnamese tones across the dialects. It is necessary to take a brief look at these previous researches on tone perception in older adults.

Varley and So (1995) performed an experiment studying Cantonese tones on the subject of tone production. For this experiment, two groups of listeners—young and older adults—were recruited. The participants had to distinguish among printed pictures that represented the speech stimuli conveying the tones. The result showed that the young listeners gave more correct responses than the older adults. The listeners above 50 and 69 years showed a deterioration of tone perception. They accounted for 61% (50–59 years) and 53% (60–69 years) of the errors. Among the six Cantonese tones, the authors revealed that tone T4 (a low-falling tone) was more confusable than those with high pitch levels and rising contours in the older adults' tone perception. This finding led the researchers to conclude that difficulties of tonal identifications were relevant to difficulties in auditory perception due to the listeners' age.

Similarly, Yang and his colleagues (2015) conducted a research on Mandarin tones and compared the vowel perception of young and older adults. The participants were asked to differentiate among the tones and vowels. The results showed that the older listeners scored significantly lower in tone and vowel identification than the young listeners. Comparing the error rates between tones and vowels within the group of older listeners, the authors indicated that the listeners had mostly misidentified the tones rather than the vowels. In addition, among the four Mandarin tones, tone T3 (a low-dipping one) was the one the older listeners confused the most. From these findings, the authors suggested that aging strongly affects the perception of tones in Mandarin, whereas no evidence reflects the impact of this factor on vowel identification.

Likewise, Kasisopa and his colleagues (2015) also found that among the five Thai tones, the falling tone was difficult for older listeners to recognize. The research was on the distinction of tonal contrasts in adults and older adults in noisy and quiet conditions. The participants were asked to determine whether two tones in each tonal pair were identical or different within a time limit. The results showed that the stimuli that included high-rising tones were significantly more distinguishable than the stimuli comprising falling tones. The younger listeners did significantly better in this task.

Based on these findings, the authors concluded that aging deteriorates tonal recognition in Thai, especially the falling tones.

In short, the results suggest that the tones with a falling contour negatively influence tonal identification in older adults. However, the tones with a high-rising contour do not affect tonal discrimination among older listeners. Furthermore, the younger listeners recognized tones significantly better than the older listeners. With regard to the misidentification of falling or low-dipping tones, there are several simple explanations. Firstly, some tones are acoustically similar on pitch heights or contours. So, the older listeners could not distinguish well between two tones that share similar features (Yang et al., 2015; Varley & So, 1995). Secondly, difficulties in tonal recognition may reduce psychoacoustic processing or degradation of cognitive ability (Yang et al., 2015; Kasisopa et al., 2015).

With regard to the Vietnamese language, previous research perceptually examined the tones across its dialects, but primarily focused on younger listeners. Thus, there is no information on tone recognition by older listeners, in particular, by the native listeners of SVN. Although languages and methodologies were tested differently across the studies in Vietnamese, Thai, Mandarin, and Cantonese, the findings shared similar features relevant to tonal perceptions. That is, the tones with low pitch and falling contours are more challenging to the listeners.

Vũ (1981) examined tonal perceptions in three groups of dialects, including NVN, CVN, and SVN. In this experiment, the listeners of each dialect were asked to determine tones in speech stimuli presented in their own dialect as well as other dialects. The results suggested that the listener's dialect plays a significant role in determining tones. In addition, the listeners scored significantly better on the tone when the speech stimuli denoted the context than when the speech stimuli contained only individual syllables. The Southern listeners could recognize well two tones in their own dialect, namely *ngang* (A1) and *sắc* (B1). However, they were more confused when it came to *huyền* (A2), *ngang* (B2), and *hỏi* (C). The tones A2 and B2 share similar acoustic features. This led to a lot of errors. This finding was comparable to that of Brunelle and Jannedy (2013). They conducted an experiment on perceptions of tone across the dialects. The participants, who were native listeners of SVN, were asked to discriminate among tones in their own dialect and NVN. The listeners could identify four tones, A1, A2, B1, and B2, in both dialects quite well. However, they could not identify well the tones C1 and C2 in NVN. The listeners made lots of errors in recognizing the tones A2 and B2 in their dialect. Interestingly, they recognized these two tones significantly better in NVN. The authors

emphasized that due to the similar contour and heights of pitch, the tones A2 and B2 are less distinct. Hence, confusion perceptually exists for these two tones in SVN. In contrast, A2 acoustically differs from B2 in NVN. So, even the SVN speakers could make out these tones well. In the same vein, Kirby (2009) conducted an experiment examining tone recognition among tonal pairs. The research revealed that the Southern listeners found tone A2/C1 pair (low, falling) to be significantly more confusing than other tonal pairs (high, rising).

Overall, these results are consistent regarding the identification of tones with low pitch-falling contours in the tonal languages. With respect to tone identifications in Cantonese, Mandarin, and Thai, the researchers suggested two possible reasons for the misidentification of the falling tones among older adults: aging factor and similar acoustic features of tones. However, the findings from Vŭ, Brunelle, and Jannedy showed that the confusability of the low-falling tones for the young listeners of Vietnamese (SVN) was probably because of the similar acoustic features of the two tones (A2 and B2). Yet, it is still unclear whether this is relevant to tonal identification by the older adults in Vietnamese. Particularly, it is not known how well the older listeners of SVN identify tones in their familiar dialect, or which among the fives are more confused by them, as it was observed in the cases of Cantonese, Mandarin, and Thai.

#### 2.4 Tones in Southern Vietnamese

A five-tone system is representative of SVN. It can be classified depending on their pitch levels and pitch contours (Kirby, 2010; Brunelle, 2009; Vŭ, 1981). The acoustic features of pitch levels and pitch contours are illustrated in figure 1 (a right panel) in the section of Vietnamese dialects. In this system, tones *hỏi* and *ngã* merge into a single tone *hỏi* (C), a low-dipping rising tone in the SVN. Tone *ngang* (A1) has a high pitch level and a high contour. Tone *huyền* (A2) has a lower pitch and a falling contour, whereas the tone *sắc* (B1) has a high level and a rising contour. Finally, the tone *nặng* (B2) is produced as a low-falling one, starting at mid level and then dropping rapidly to the bottom range. Brunelle and Jannedy (2013) stated that the most important distinction between the SVN and the NVN is relevant to laryngealization (creaky voice). This laryngealization is missing altogether in the SVN, leading to the coalescence of tones C1 and C2 into one tone (C). So, as mentioned earlier, young listeners are perceptually confused between tones A2 and B2 due to very similar F0 heights and contours.

In this study, all five tones were employed to find out the effect of tonal patterns of syllables in SRTs, while the three tones of A1, A2, and B1 were employed to study the effects of F0 on phoneme recognition.

## 2.5 Speech perception in older adults

### 2.5.1 Deficits in speech perception

Speech perceptions of older listeners have received considerable attention (Stam et al., 2015; Lee, 2015; Schneider, 2011; Cervera et al., 2009; Jenstad, 2001; Studebaker et al., 1997; Gelfand & Piper, 1987; Townsen & Bess, 1980). Most of the studies support the hypothesis that older listeners have deteriorated speech perception. The deteriorations of speech reception are apparent in listeners aged 60 years or above. Gelfand and his colleagues examined the ability of consonant recognition using nonsense-syllable tests for listeners aged between 21 and 68. The results showed no significant differences in consonant scores between the young and the middle-aged listeners. However, significant difference was found in a group of listeners aged above 60 years, especially when the stimuli were presented in a noisy condition. Similarly, Studebaker et al. measured WRS on the listeners aged between 20 and 90 years by using monosyllabic words. The authors indicated that the word scores of the 70-year-old listeners was significantly deteriorated as compared with the 30-year-old listeners. More noticeably, the 80-year-old listeners scored significantly worse than the 70-year-olds. Regarding the extent of differences in speech recognition among adult listeners, Divenyi et al. (2005) found an increment of 6.32dB/decade for listeners aged between 60 and 83.7 years. Besides, Pronk et al. (2013) analyzed the results of a Longitudinal Aging Study of Amsterdam (Huisman et al., 2011), which focused on the changes in SRTs in noise among 1,298 participants aged between 57 and 93 years. The analysis showed a deterioration of roughly 0.13 to 0.27 dB per year in speech recognition thresholds across four decades. The discrepancy in the recognition thresholds across the age groups between the two studies might stem from sample sizes. Divenyi et al. worked with a small number of participants whereas Pronk et al.'s study had a significantly large sample size.

In addition to the difficulties of speech perception in noise, high-frequency hearing loss is considered to be the second primary problem of older listeners. While listening to speech materials, older adults performed poorly while recognizing high-frequency phonemes, for example, fricative sounds: *s*, *sh*, *f*, *th*, *ʒ*, *v*, as compared with other phoneme categories (Hull et al., 2012; Kuk, 2010; Kramer, 2008). Also, for pure-tone stimuli, duo-tone thresholds changed the

most at high frequencies of 3–8 kHz, and did not change or minimum change at low frequencies of 0.5 and 2 kHz for the 50–60 age group (Wiley et al., 2008).

### **2.5.2 Causes of speech perception deficits**

The researchers gave the following two remarkable reasons for the deterioration of speech recognition by older listeners: the perceptual or cognitive decline related to aging (Fortunato et al., 2016; Mukari et al., 2015; Cervera et al., 2009; Akeroyd, 2008; Gelfand & Piper, 1987). However, speech deterioration could also be due to “hearing loss or age-related changes in cognitive functioning, or both of these factors” (Schneider, 2011).

To study the effects of perceptual decline, Jerger (1972) compared the speech performances of three groups of listeners: normal-hearing, young hearing-impaired, and older hearing-impaired listeners. The normal-hearing listeners, of course, performed significantly better than the two groups of hearing-impaired listeners. However, at the same level of hearing impairment, the young hearing-impaired listeners performed better than the older hearing-impaired listeners. Lunner and Sundewall-Thorén (2007) found that the high-cognitive-score group showed better mean-speech thresholds (4.5 dB lower) than those by the low-cognitive-score group. Similarly, Gatehouse et al. (2003) found that the high-performing groups scored significantly better on speech identification of words (9% higher) than the low-performing groups. These results supported the hypothesis that the cognitive decline in older adult listeners is a primary factor for the deterioration of speech performances.

For the effects of sensitive declines, Cervera et al. (2009) measured PRSs and working memory scores in two groups of listeners: one was young listeners (19 to 25 years), and the other was younger-older adult listeners (55 to 65 years). The result indicated that the younger-older listeners performed worse on both phoneme recognition and working memory than the young listeners. This result might support the concept that sensory deterioration is a major cause of decrement in speech and working memory capacities in older listeners. Akeroyd (2008) surveyed 20 studies that measured the association between speech recognition in noise and aspects of cognition. The author once again confirmed that there was a link between speech recognition and cognitive capacity (in noise). The author also found that sensory declines, and not cognitive declines, remained the primary reason for the deterioration of speech recognition. Cognitive decline only had a secondary effect.

Regarding sensory declines, presbycusis-related hearing loss is prevalent among older adults. It is associated with aging. This kind of hearing loss accounts for more than 90% hearing impairment in

older patients (Fortunato et al., 2016). For listeners with presbycusis, the input signals are deteriorated due to sensory decline (Lorienne, 2001). As a result, it is difficult for them to encode the speech signals. Cognitive ability includes working memory, attention, speed of information processing, and intelligence. Among these, working memory has been intensively studied. Cervera et al. (2009) stated that working memory was “crucial in language processing” because the “intermediate product of comprehension has to be kept active until the listeners can understand the message”. Thus, the working memory is considered to be “an active system” for the allocation, storage and processing of the information. Thus, deficits in the working memory by older listeners contribute to declines in speech perceptions. As mentioned earlier, the declines in speech recognition among older adults is probably due to hearing loss, age-related cognitive decline, or both. Besides, some researchers provided evidence to illustrate that hearing loss can also result from impoverished signal input. For example, speech signals may be distorted or delivered to older listeners with noise. Another example relevant to realistic conversations is that older listeners often find it difficult to identify who is talking and exactly what they are talking about.

## **2.6 Speech audiometry**

The aim of an audiological evaluation is to identify the types, levels, and configurations of hearing impairment. An audiological assessment consists of otoscopy, tympanometry, auditory reflex, pure-tone audiometric testing, and speech audiometric testing. Otoscopy provides physical inspection relevant to the status of the ear canal and eardrum. Tympanometry is employed to assess the mobility of the tympanic membrane and discover fluid in the middle ear or injuries in the eardrum. Auditory reflex testing provides extra information on an involuntary muscle contraction occurring in the middle ear when sound stimuli are delivered. When these observations are combined, the audiologist can diagnose the etiology of hearing impairment.

The speech audiometry can judge how well an individual can recognize and understand speech stimuli. The speech audiometric measurement includes the SRT, the speech detection threshold (SDT), and the WRS. The SRT in a particular language is defined as the level of hearing at which a listener can correctly identify 50% of a set of spondees (Martin & Clark, 2008). The SRT is also utilized to validate pure-tone thresholds due to high association between SRTs and the average of pure-tone thresholds (0.5, 1, and 2 kHz). Thus, it is considered as a cross-check (Kramer, 2008, p. 180). The SDT is merely employed when an SRT cannot be achieved. SDT is also called a speech-awareness threshold (SAT). The objective of the assessment is to gain the lowest hearing level at

which the speech stimuli can be detected 50% of the time. The test needs clients to merely point when a stimulus is delivered. Due to its easy understandability, spondees like *airplane*, *football*, *cowboy*, and *mushroom* are used as speech stimuli in SRT and SDT. For listeners with normal hearing, the SDT is roughly 5 to 10 dB lower than the SRT (Kramer, 2008, p. 183), which requires listeners to iterate the presented words.

Along with pure-tone thresholds, SRTs are useful in illustrating the degree of hearing loss. However, SRTs are not representative of the hearing levels at which individuals listen to speech in their actual environments. Word recognition (WR) testing estimates how well a listener can identify speech stimuli at one or more hearing intensity levels. The result is described as a percent of correct words identified, which is called word recognition score (WRS). Another way to report the result of WR testing is PRS. SRT and PRS are the focuses of the current research. PRS is different from the SRT in which the entire word list is scored as a whole, rather than each speech stimulus individually, as in SRT.

### **2.6.1 Speech materials in Vietnamese**

Regarding local speech materials, three studies are relevant to the designs of speech audiometry materials in Vietnamese (Hanson, 2014; Nguyễn, 1986; Ngô, 1977). However, the speech materials have not yet been implemented in clinical assessments. Hanson and Nguyễn have developed both monosyllabic word recognition and disyllabic speech recognition materials. In Hanson's work, the monosyllabic material included 200 words divided into four sub-lists of 50 words each, and the disyllabic speech materials comprised 89 words. The materials were balanced in terms of speech intelligibility for each speech stimulus. In the monosyllabic speech material, the mean psychometric function slope at 50% was roughly 5%/dB for both female and male speakers. In the disyllabic speech material, the mean psychometric function slope at 50% was 11.3%/dB for male speakers and 10.2% /dB for female speakers. In Nguyễn's work, the disyllabic word speech test comprised 100 items, distributed into 10 lists of 10 items each, and the monosyllabic word material included 200 items, divided into 10 lists of 20 items each. Both speech materials were equalized on phoneme distributions and the frequency of the tonal patterns (high to low pitch) to ensure that the selected words in each list matched its counterpart in Vietnamese. In an early study, Ngô constructed audiometric tests including a digital stimuli test (ranging from 11 to 99) and a speech test (200 monosyllabic words). These speech tests (developed by Nguyễn and Ngô) were then conducted on individuals with hearing impairment. The authors suggested in the early 1990s that these speech

materials could be adapted to the needs of clinical assessment in Vietnamese. However, we think that due to a lack of normative values of SRTs, and dialectal variation the two speech materials have so far been inapplicable to a clinical setting. Hanson only assessed the psychometric equivalence of speech audiometry materials for research purposes. Thus, the normative values of SRT or WRS are still unavailable for speech audiometry materials in Vietnamese.

### 2.6.2 Adaptive Auditory Speech Test

The Adaptive Auditory Speech Test (AAST) is an automatic procedure to determine the SRT of young children and adults in both quiet and noisy conditions. Using a closed set of only six spondees, the test procedure is minimally dependent on an individual's lexicon.

The listeners have to click on the correct picture representing a speech stimulus after hearing it. When the responses are correct, the intensity levels of the next speech stimulus will be 5dB softer (3dB softer in noise). When the responses are incorrect or there is no response, the next stimulus will be 10dB louder (6dB louder in noise). The test ends after seven wrong answers from a listener. The maximum mean testing time is two minutes per condition.

The speech material is available in several languages: German, Dutch, Spanish, Polish, Luxemburgish, Chinese, and Ghanaian. The speech material uses spondees like *Eisbär* "polar bear", *Schneemann* "snowman", *Fussball* "football", *Flugzeug* "airplane", *Handschuhe* "gloves", and *Lenkrad* "steering wheel" as speech stimuli (Coninx, 2005). These words have a redundancy that is comparable to short everyday sentences comprising only two keywords. When the spondee words do not exist in a particular language, for example, Spanish, trisyllabic words are considered as a replacement.

The speech test uses an automatic adaptive procedure. Thus, the evaluation of an individual's SRT is reliable. Moreover, the average testing time is short and the learning effects are correspondingly fast and negligible. In addition, children aged four years show their motivation towards the test, and they can perform it quickly and comfortably.

With respect to the German AAST, Coninx (2005, 2008) has validated the reliability of AAST in testing German children aged between four and 12 years and determined its normative values. He found that children aged four years performed significantly worse on SRTs (10dB) than those aged 11 years. Children aged eight years could achieve average speech thresholds that were comparable to adults' speech thresholds. A possible explanation for the difference was a lack of concentration in

younger children. Based on these findings, the author suggested that the listener's age and their SRTs in German AAST were interdependent. The author found that at 50 percent correct responses, the slope of the psychometric curve was 14%/dB for speech-in-noise measurement. This finding was in line with that of the Oldenburger Kinder Satztest (Wagener & Kollmeier, 2005), which had slopes of 6–8%/dB and 12–14%/dB in quiet and noise respectively.

## 2.7 Summary

The purpose of this section is to review the literature on the dialectal effects on speech audiometry testing and dialectal variations across the Vietnamese dialects, tonal identification in tonal languages by older listeners, and speech perception in older adults.

With respect to dialectal effects, in almost all the studies, the native listeners of a dialect obtained low speech thresholds or high phoneme scores (the best) in contrast to the non-native listeners. The latter achieved high speech thresholds and somewhat low phoneme/word scores (Weisleder & Hodgson, 1989; Schneider, 1992; Le et al., 2007; Shi & Canizales, 2012; Nissen et al., 2013). This means that a strange dialect was the reason for a decline in speech intelligibility by listeners. However, the conclusions drawn from these findings are not in agreement regarding their clinical significance. Schneider, Crews, and Nissen et al. found effects of dialects on speech perceptions, but the effects were not significant for clinical application. In contrast, Weisleder and Hodgson, and Shi and Canizales have argued that the dialect negatively affected the clinical assessment of clients.

The section also provided an overview of different features across Vietnamese dialects relevant to phonetics and lexicons. About phonetics, the initial consonants (onsets), prevocalic, nuclei, and final consonant (codas) reflect a wide variety of dialects. In the initial consonants, the retroflexes /ʂ, ʐ, tʂ/ are enunciated as /s, z, c/ respectively in NVN. In SVN, these sounds are produced as /s, ʃ, c/, such as, *sáu—xáu* “six”, *trà—chà* “tea”, and *gwơu—gu* (*gu*). The initial /v/ behaves as /j/ in the SVN, in contrast to the NVN and CVN, in which the /v/ does not change. For the medial glide, speakers of the SVN produce the prevocalic /-w-/ in two different ways: (1) they remove the onset in a syllable and then the prevocalic becomes the onset instead, (2) they delete the prevocalic from the syllable, and another consonant replaces the initial phonemes at the onset. In contrast, the speakers of the remaining dialects retain the prevocalic /w/ in their speech. For the nuclei, diphthongs /i\_ə, u\_ə, u\_ə/ are shifted into long vowels /i:, w:, u:/ in SVN, whereas these are distinguished manifestly from the long vowels in NVN and CVN. Furthermore, the vowels /ɛ, e/ are enunciated like /i/ in the SVN when it precedes the final phoneme /-m, -p/ in syllable

structures. These vowels are discriminated well from each other in the NVN and the CVN. The final consonants /-t/ and /-k/ perceptually and acoustically differ between NVN and SVN. The SVN speaker uses the phoneme /-t/ as /-k/, and /-n/ as /-ŋ/ in some cases. However, in most cases, the speakers of NVN and CVN distinguish /-t/ from /-k/, and /-n/ from /-ŋ/. Regarding the tonal systems across dialects, the standard Vietnamese has six tones. But in the dialects, only five or four tones are used. SVN speakers do not differentiate between *ngã* (C2) and *hỏi* (C1) with respect to pitch heights. These two tones are coalesced into the single tone C in the SVN. In contrast, the NVN speaker can distinguish between these tones based on the height of tones (low pitch for C1, and high pitch for C2). The tones C2 and C1 are united into the single tone B2, and B1 behaves as C1 in some regional dialects of the CVN. Due to perceptual and acoustic distinctions, non-native listeners misperceive the tones from an unfamiliar dialect more than those from their own dialect (Brunnel & Jannedy, 2013).

Apart from the difference in tones and phonetics, the dialects also differ in terms of their lexical aspects. Non-living and animated objects are named differently across the dialects. These differences in tones, phonetics, and lexicons across the Vietnamese language might lead to confusions for native and non-native listeners of a dialect in daily conversation.

With regard to tonal identification by older adults, the results suggested that the tones with low F0 values (low-falling contour) affect a listener's speech identifications in tonal languages (Thai, Cantonese, and Mandarin). In contrast, the tones with a high-rising contour do not affect the older listeners regarding either speech recognition or tone discrimination. No reference studies have so far examined tonal or speech identifications by older listeners under the effects of lexical tones in Vietnamese. However, the findings of Brunelle and Jannedy (2013) and Vũ (1981) found that younger listeners frequently misidentified the tones that had similar contours and height of pitch, such as *huyền* (A2) and *ngã* (B2).

Regarding the speech perception of older adults, previous research has shown that speech reception deteriorated in listeners aged more than 60 years. An increment of 6.32 dB per decade was found for listeners aged between 60 and 83.7 years (Divenyi et al., 2005). The reasons for this decline of speech recognition in older adults were derived from the perceptual or cognitive declines associated with aging (Fortunato et al., 2016; Mukari et al., 2015; Cervera et al., 2009; Gelfand & Piper, 1987). Sensory declines are known to have the primary effect on the deterioration of speech recognition. Cognitive declines have only a secondary effect (Akeroyd, 2008). The difficulties of speech perception in noise

and high-frequency hearing loss are considered as the two properties of sensory impairment in older listeners.

For the speech audiometry materials in Vietnamese, Hanson (2014), Nguyễn (1986), and Ngô (1977) have designed the speech materials for both monosyllabic and disyllabic words. However, these speech materials have not yet been implemented for clinical purposes due to dialectal variation.

In the next section, the designs of speech materials—AASTs and NAMES in Vietnamese—will be described.

### 3. DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

To achieve the aims of the current research, we developed two kinds of speech materials depending upon the linguistic features of SVN. First, the AAST was designed based on the phoneme frequency in written languages. The speech stimuli were two-syllable noun phrases. The test was used to measure an individual's SRT. Second, NAMES was based on the phoneme frequency in the spoken language. The stimuli were meaningless disyllabic structures (CV-CVC). This test was used to assess a listener's PRS. These two speech materials were utilized as the stimuli to determine (1) the normative values, (2) the dialectal effects, and (3) the effects of the tonal pattern of syllables on speech perception, and the learning effects (AAST) on speech recognition thresholds.

#### 3.1 Designs of a speech material of AAST

In designs of the AAST, researchers have suggested the following criteria: the selected words should be familiar to the listeners; the phonetic elements should be different across the words; the phonetic elements of a speech material must duplicate the distribution of phonemes in that language (Carhart, 1951 and 1952; Coninx, 2006); and the speech stimuli must be more homogeneous in terms of speech intelligibility (Ramkissoon, 2001).

The purpose of auditory speech material is to measure an individual's speech threshold, not his/her cognitive capacity or intelligence (ASHA, 1988). Especially in AAST, the speech materials are designed for children. Hence, the selected words must be familiar to them, simple, and age-appropriate. Phonetic dissimilarity is the most important feature when the words are selected for AAST. Similar phonology among speech stimuli can cause confusions for listeners when the stimuli are presented at softer presentation levels. More importantly, Carhart proposed that phoneme distribution in a speech audiometry test should be similar to its counterpart in a language. Even if AAST includes only six spondees, it covers the phonetic range of SVN. Due to the limited numbers of words, some less-frequently-used phonemes might occur in the AAST. The homogeneity of speech intelligibility means that each speech stimulus within a speech test or across the speech tests of AAST has a balanced amplitude level. Thus, the speech stimuli are also adjusted in their intensity level to ensure that their audibility within the subtest and across subtests of AAST are homogeneous.

In the current study, these four criteria have been adhered to in designing the five subtests of AAST in SVN. Below are more details of the choices of AAST words, the phoneme frequencies, and the intensity balances for AAST.

### 3.1.1 Phoneme frequencies in Southern Vietnamese—written text

Unfortunately, there has been no literature on phoneme distribution in Vietnamese. Early studies only examined the frequencies of tones taken from Vietnamese dictionaries (Phạm, 2003; Hao, 1997). Due to the lack of findings on phoneme frequency, phoneme distribution in SVN had to be calculated (Nguyễn, 2014). Sample texts were collected from different sources, including books for children and young readers, along with local online newspapers. This corpus consisted of 157,337 monosyllables, corresponding to 424,175 phonemes. Overall, vowel phonemes account for 46.3% while consonant phonemes make up 53.7% of the SVN text. The detailed frequency of the occurrence of phonemes in SVN is reported in Appendix A.

It has to be noted that the phonemic system in SVN misses some pairs of phonemes due to the merger as compared with those in NVN and CVN. These mergers occur in several contexts. For example, the phonemes /v- and z-/ are replaced by /j-/ in the initial position, and the phonemes /-t/ and /-n/ are altered by /-k/ and /-ŋ/ respectively in the final position. These have already been mentioned in the literature section (§2.2.2). The frequency of the occurrence of phonemes in SVN was categorized based on this regulation.

### 3.1.2 Choices of disyllabic noun phrases

There is ambiguity in terms of the definition of “word” in Vietnamese. The term “word” can be understood as “the smallest meaningful unit” that refers to an object’s name. It can stand by itself, and combine with another to constitute a sentence (Mai et al., 1997). Based on this definition, each word in Vietnamese is considered as a linguistic unit, which has one or more morphemes to name objects. Similarly, other linguists, like Cao (1999) and Nguyễn (2013), have also proposed that the term “word” is “the smallest meaningful unit”. However, they have suggested that each word corresponds to a syllable. In other words, the term “word” in Vietnamese is considered to be a monosyllable. A word unit including two syllables is considered to be a disyllabic phrase and not a compound word as mentioned in other languages.

In the current research, the authors followed the word definition proposed by Cao and Nguyễn: *a word is the smallest meaningful unit, and each word corresponds to a syllable*. Therefore, a combination of two syllables (noun) is referred to as *a disyllabic noun phrase*. For the sake of convenience, however, the authors preferred to use the term *word*, and in some necessary cases, the term *disyllabic noun phrase* to indicate the speech stimulus in AAST.

Selected words were generated from the most-frequently-used words in Vietnamese. This was done to make sure that younger children could easily recognize the words. The words were derived primarily from the SVN dictionary (Huynh, 2007). Fifty-four words were collected, with each disyllabic noun phrase comprising six to seven phonemes. Eventually, 26 disyllabic noun phrases were considered. They met the following criteria (Coninx, 2005; Offei, 2013):

The phoneme frequency in AAST duplicates its counterpart in SVN:

- The words phonetically differ from each other
- The words have the same prosodic patterns
- The participants, especially the younger children, know well the meaning of words.

The five subtests of AAST have been designed with a tonal contrast in pitch heights and pitch contours. In the SVN, speakers can acoustically and perceptually distinguish only five tones—*ngang*: A1, *huyền*: A2, *sắc*: B1, *ngã*: B2, *hỏi-ngã*: C (tone C1 and C2 merge into a single form). Below are explicit mentions of tonal patterns of words in each subtest of AAST.

- AAST-a1: high levels, flat contours, including only tone A1
- AAST-a2: high pitch levels with rising contours, including tones A1 and B1
- AAST-a3: low pitch levels with falling contours, consisting of tones A2, C, and B2.
- AAST-a4: high and low levels with rising and falling contours, including four tones A1, A2, B1, and B2.
- AAST-aTP: six words are extracted from the four aforementioned subtests. Similar to a4, high and low pitch levels with rising and falling contours are implemented. It consists of three tones, A1, B1, and A2.

The 26 two-syllable noun phrases are distributed into five subtests. These are presented in Table 1, along with the IPA transcription, prosodic and tonal patterns, and glossaries. Each subtest of AAST is designed to be phonetically balanced as compared with the references of phonemic frequency in the SVN. Subsection 3.1.3 gives the comparisons of the phoneme frequencies in each subtest of AAST and the SVN.

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

**Table 1:** The 26 selected words for AAST in Vietnamese, grouped into five subtests with six words each

	Word	IPA	PP	TP	Glossary
<b>a1</b>	<i>thanh long</i>	/t <sup>h</sup> ɛŋ <sup>1</sup> lɔŋ <sup>1</sup> /	S-S	HL-HL	“dragon (fruit)”
	<i>con trai</i>	/kɔn <sup>1</sup> tʃaj <sup>1</sup> /	S-S	HL-HL	“boy”
	<i>chim sâu</i>	/cim <sup>1</sup> ʃəw <sup>1</sup> /	S-S	HL-HL	“bird”
	<i>ban công</i>	/ban <sup>1</sup> kɔŋ <sup>1</sup> /	S-S	HL-HL	“balcony”
	<i>cam tươi</i>	/kam <sup>1</sup> tuw <sup>1</sup> əj <sup>1</sup> /	S-S	HL-HL	“fresh (orange)”
	<i>dây đeo</i>	/zəj <sup>1</sup> dɛw <sup>1</sup> /	S-S	HL-HL	“wearing (chain)”
<b>a2</b>	<i>sóng lớn</i>	/sɔŋ <sup>5</sup> lɔn <sup>5</sup> /	S-S	HR-HR	“big (waves)”
	<i>trái đất</i>	/tʃaj <sup>5</sup> dət <sup>5</sup> /	S-S	HR-HR	“earth”
	<i>pháo bông</i>	/faw <sup>5</sup> boŋ <sup>1</sup> /	S-S	HR-HL	“fireworks”
	<i>viên thuốc</i>	/ji <sup>1</sup> ən <sup>1</sup> t <sup>h</sup> u <sup>1</sup> ək <sup>5</sup> /	S-S	HL-HR	“pill”
	<i>túi xách</i>	/tuj <sup>5</sup> sɛk <sup>5</sup> /	S-S	HR-HR	“hand(bag)”
	<i>mắt kính</i>	/măt <sup>5</sup> kiŋ <sup>5</sup> /	S-S	HR-HR	“glasses”
<b>a3</b>	<i>mặt trời</i>	/măt <sup>6</sup> tʃəj <sup>2</sup> /	S-S	LBr-LF	“sun”
	<i>lồng bàn</i>	/lɔŋ <sup>2</sup> ban <sup>2</sup> /	S-S	LF-LF	“dish (cover)”
	<i>hình tròn</i>	/hiŋ <sup>2</sup> tɔn <sup>2</sup> /	S-S	LF-LF	“circle”
	<i>chậu cảnh</i>	/cəw <sup>6</sup> kɛŋ <sup>4</sup> /	S-S	LBr-LR	“flower(pot)”
	<i>điện thoại</i>	/di <sup>1</sup> ən <sup>6</sup> t <sup>h</sup> waj <sup>6</sup> /	S-S	LBr-LBr	“tele(phone)”
	<i>giỏ quà</i>	/jɔ <sup>4</sup> kwa <sup>2</sup> /	S-S	LR-LF	“gift(basket)”
<b>a4</b>	<i>hương thực</i>	luw <sup>1</sup> əŋ <sup>1</sup> t <sup>h</sup> uik <sup>6</sup>	S-S	HL-LBr	“food”
	<i>bóng đèn</i>	/bɔŋ <sup>5</sup> dɔn <sup>2</sup> /	S-S	HR-LF	“bulb”
	<i>học sinh</i>	/hɔk <sup>6</sup> ʃiŋ <sup>1</sup> /	S-S	LBr-HL	“pupil”
	<i>nốt nhạc</i>	/not <sup>1</sup> ŋak <sup>6</sup> /	S-S	HR-LBr	“musical (note)”
	<i>măng cụt</i>	/măŋ <sup>1</sup> kut <sup>6</sup> /	S-S	HL-LBr	“mangosteen”
	<i>quạt giấy</i>	/kwat <sup>6</sup> jəj <sup>5</sup> /	S-S	LBr-HR	“paper (fan)”
<b>a-TP</b>	<i>con trai</i>	/kɔn <sup>1</sup> tʃaj <sup>1</sup> /	S-S	HL-HL	“boy”
	<i>túi xách</i>	/tuj <sup>5</sup> sɛk <sup>5</sup> /	S-S	HR-HR	“bag”
	<i>lồng bàn</i>	/lɔŋ <sup>2</sup> ban <sup>2</sup> /	S-S	LF-LF	“dish (cover)”
	<i>bóng đèn</i>	/bɔŋ <sup>5</sup> dɔn <sup>2</sup> /	S-S	HR-LF	“bulb”
	<i>cầu thang</i>	/kəw <sup>2</sup> t <sup>h</sup> aŋ <sup>1</sup> /	S-S	LF-HL	“stair”
	<i>đôi mắt</i>	/doj <sup>1</sup> măt <sup>5</sup> /	S-S	HL-HR	“eyes”

The words are transcribed on IPA with the superscript above each syllable showing its tone (1: A1, 2: A2, 4: C, 5: B1, 6: B2). Abbreviates (PP: prosodic patterns, TP: tonal patterns, S: strong, HL: high level, HR: high rising, LF: low falling, LBr: low broken, LR: low rising)

3.1.3 Phoneme frequencies

The consonant phonemes in the speech test are grouped in terms of the manners of articulation: aspirated (Asp), plosive voiceless (PVs), plosive voiced (PVed), nasal, fricative voiceless (FVs), fricative voiced (FVed), and lateral (Lat). The vowel phonemes are exhibited individually and coupled with tones. Table 2 and Figures 2 and 3 display the phoneme distribution of each speech material compared with the reference distribution.

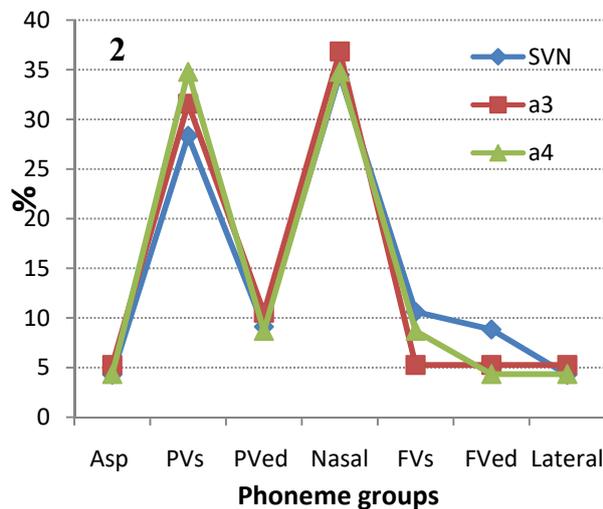
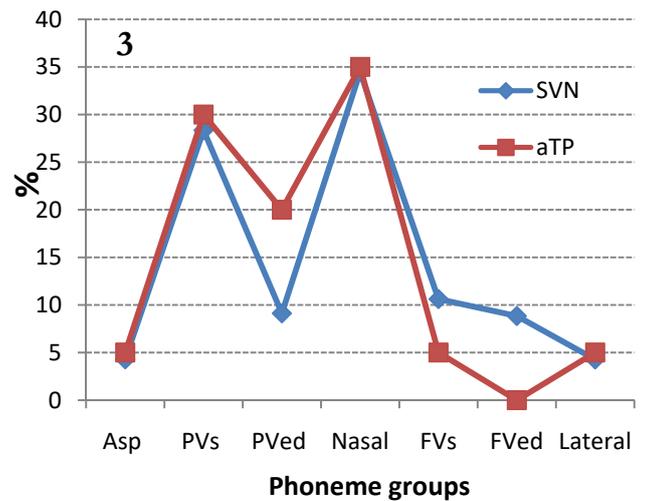
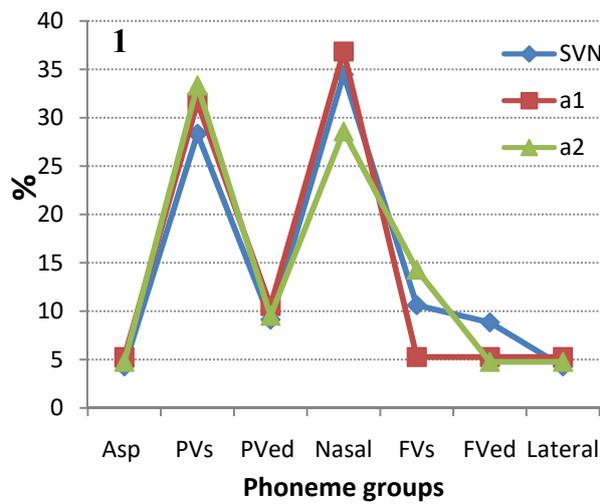
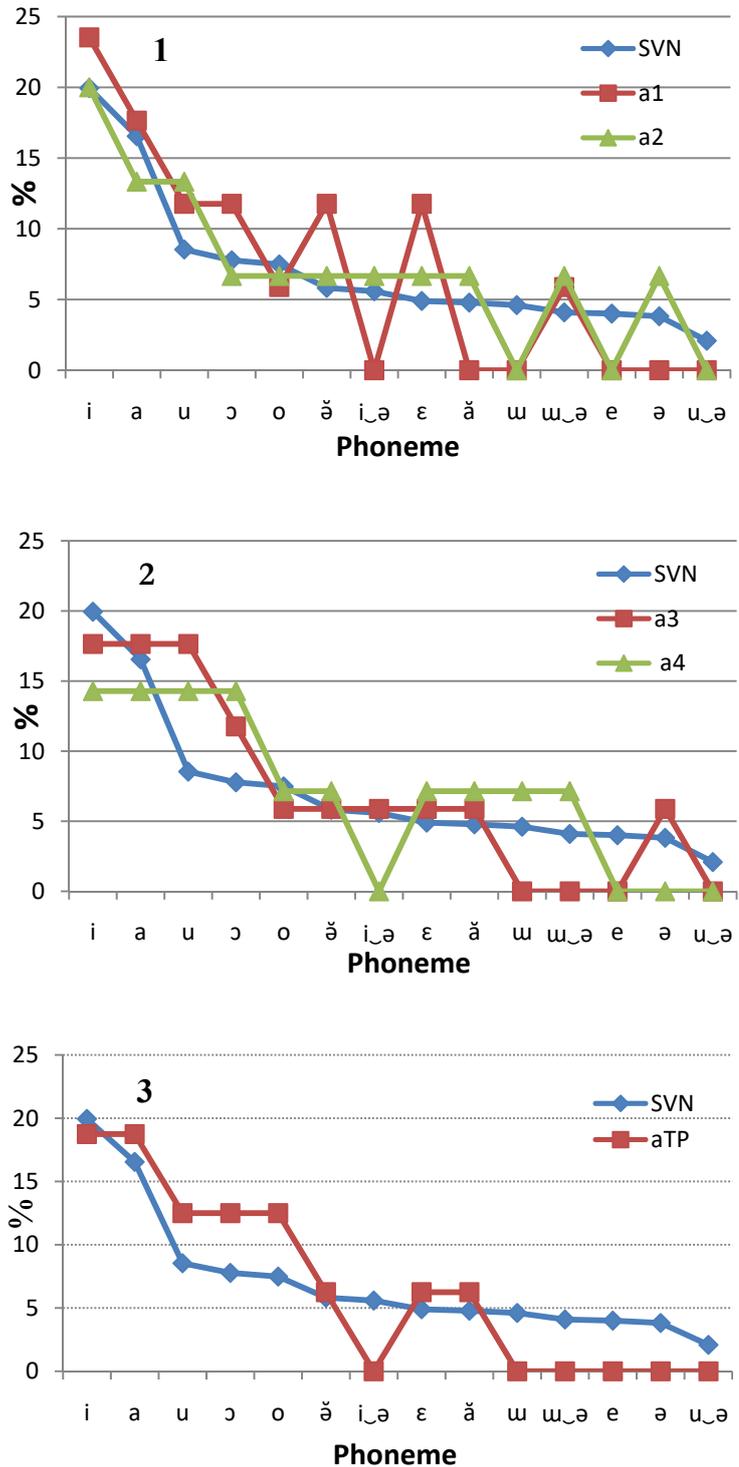


Table 2: The frequency of consonants (percent) in the subtests of AAST compared with those in SVN

Manner	SVN	a1	a2	a3	a4	aTP
Asp	4.3	5.3	4.8	5.3	4.4	5
PVs	28.4	31.6	33.3	31.6	34.8	30
PVed	9.1	10.5	9.5	10.5	8.7	20
Nasal	34.5	36.8	28.6	36.9	34.8	35
FVs	10.6	5.3	14.3	5.3	8.7	5
FVed	8.8	5.3	4.8	5.3	4.4	0

Figure 2: The frequency of consonants in each subtest of six AAST words compared with those in the written language

The curves illustrate the phoneme distribution of AASTs and the language. They are relatively homogenous. The distribution of the vowel phonemes is not very close as compared with the consonant phonemes. First, since the subtests included only six words each, the distribution of phonemes in each AAST may not be 100% same as those in the written language. Second, in the vowel phonemes (Figure 3), the speech tests are designed on the basis of frequently used vowels. So, some phonemes with a low frequency of occurrence, for example, diphthongs, might appear in the selected words of the speech materials. However, the proportions of phoneme distributions in each subtest are comparable to those in the SVN.



**Figure 3:** The frequency of vowel phonemes in each subtest of AAST in comparison with their counterpart in the written language

3.1.4 Picture drawings

Twenty-six simple pictures (Figure 4 and 5) corresponding to the indigenous Vietnamese culture were drawn. All the pictures were in the following format: JPG, 201x174 pixels.



Figure 4: Test screens of the subtests of AAST, a1 to a4

Both speech materials a1 and a2 were conducted to measure hearing in children. To find out whether the images were identifiable, 20 participants aged between four and six were asked to identify all the pictures from the speech materials of a1 and a2. Over 90 percent of the selected children could label the pictures. Five students were asked to identify the remaining pictures in the other subtests. They could easily label the pictures. So, it can be said that these pictures are suitable for the screening tests.

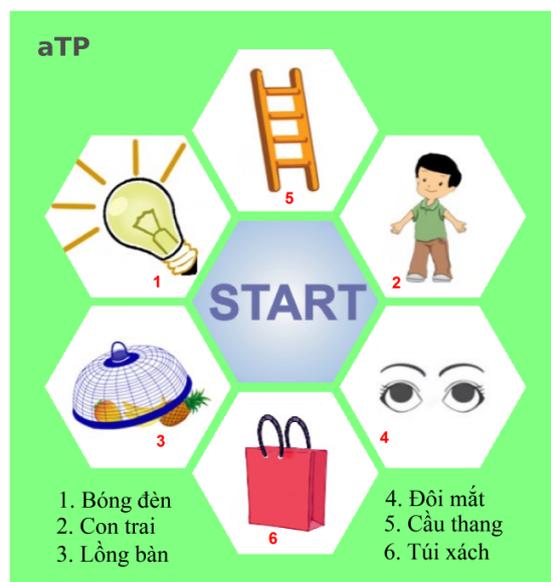


Figure 5: Test screen of the subtest aTP

### 3.1.5 Sound recordings

The speech stimuli were recorded in the speech stream by a native female speaker of SVN in a natural and clear voice. The speaker was asked to maintain constant speech levels and distance from the microphone. The recordings of the AAST speech materials took place in a sound-treated room in VOH (Voice of Ho Chi Minh City's people), setting sampling rates of 44.1 kHz and a resolution of 32 bits. The files were stored as mono sound (one channel, one microphone). Each word was recorded twice: first in the order of 1-2-3-4-5-6, and second, in the reverse order, 6-5-4-3-2-1.

A native SVN speaker evaluated the recorded stimuli to ensure the best quality in terms of the speech rate, loudness, intonations, and clarity. The best version of each speech stimuli was selected. The duration of silence between the first and second syllables in each word was trimmed by the acoustic software Cool Edit Pro 2.1. The trimmings of the speech stimuli ensured that these speech stimuli were perceptually natural and suitable for the native listeners of SVN. For example, a rapid cut-off at the end of each syllable was avoided. Finally, each syllable in each speech stimuli was digitally edited to have the similar root mean square (RMS) level, which could make amends for potential intensity differences during the recordings. Thirty stimuli were selected for five subtests of AAST. Table 3 shows the balances of the RMS levels between the syllables within a word and across words.

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

**Table 3:** The average and tonal RMS powers for the stimuli of AAST

AAST	Stimuli	Average power (dBFS)	Total power (dBFS)	AAST	Stimuli	Average power (dBFS)	Total power (dBFS)
<b>a1</b>	<i>thanh long</i>	-18.8	-17.3	<b>a3</b>	<i>mặt trời</i>	-21.3	-19.4
	<i>con trai</i>	-18.4	-17.1		<i>lồng bàn</i>	-21.4	-20.2
	<i>chim sâu</i>	-15.2	-13.5		<i>hình tròn</i>	-20.6	-19.6
	<i>ban công</i>	-20.2	-18.5		<i>chậu cảnh</i>	-21.2	-19.5
	<i>cam tươi</i>	-15.7	-14.2		<i>điện thoại</i>	-20.1	-18.7
	<i>dây đeo</i>	-15.8	-14.4		<i>giỏ quà</i>	-21.0	-20.0
<b>a2</b>	<i>sóng lớn</i>	-19.2	-18.1	<b>a4</b>	<i>hương thực</i>	-17.1	-15.3
	<i>trái đất</i>	-16.5	-15.8		<i>bóng đèn</i>	-17.6	-16.6
	<i>pháo bông</i>	-19.0	-17.4		<i>học sinh</i>	-20.1	-18.1
	<i>viên thuốc</i>	-15.5	-14.7		<i>nốt nhạc</i>	-17.4	-15.6
	<i>túi xách</i>	-16.3	-15.1		<i>măng cụt</i>	-18.0	-16.6
	<i>mắt kính</i>	-17.7	-15.8		<i>quạt giấy</i>	-19.2	-17.6
<b>aTP</b>	<i>con trai</i>	-13.3	-12.9	<b>aTP</b>	<i>bóng đèn</i>	-16.1	-15.6
	<i>túi xách</i>	-15.8	-15.2		<i>đôi mắt</i>	-16.8	-15.7
	<i>lồng bàn</i>	-16.8	-16.8		<i>cầu thang</i>	-17.5	-16.6

(All values are defined in dBFS (dB full scale))

SRTs in speech audiometric tests of AAST were measured not only in quiet but also in noise. A speech material in noise is considered the best approach to replicate background noise in daily life situation (Neumann et al., 2012). So, a noise sound file was recorded by the same speaker. The speaker was asked to read aloud a piece of daily news from her broadcast program. The recording lasted approximately two minutes.

### 3.1.6 Preparations of pilot tests for AAST

The first version of AAST in Vietnamese was prepared in April 2014 at IfAP in Solingen, Germany. The purpose of the pilot test was to get the initial data to equalize speech intelligibility across individual stimuli within the speech tests and among the speech tests as far as possible. The pilot test measurements were performed in the University of Đồng Tháp (Vietnam) with 40 normal-hearing participants aged between 20 and 30 years (mean age=24 years). The normal-hearing subjects were corroborated by duo-tone audiometry,  $\leq 30$  dB HL for octave frequencies of 0.5 and 4 kHz. The 40 listeners were divided into two groups of 20 subjects each. The first group was tested by using the speech tests a1 and a2 and the second one with a3 and a4. The participants were screened

individually for both listening conditions of quiet and noise. The speech tests were delivered monaurally to the listener in an interleaved test order, for example, either a1 and a2, or a2 and a1. The purpose of this method was to avoid possible learning effects in the individual's SRT. The equipment was calibrated before the commencement of the hearing measurement. In the test, the subject listened to the speech stimuli and pointed out the picture matching the meaning of the stimulus they had just heard. Eventually, these initial data were analyzed to determine whether the speech stimuli within the speech test and across the speech tests were perceived to be equal in terms of audibility. The speech intelligibility of each stimulus was presented separately in the psychometric curve. Based on these psychometric curves, the intensity levels were accommodated and equated for each stimulus within the speech test to bring the psychometric curves as close as possible to the optimal psychometric curve. Adjustments were also made across the subtests of AAST.

### **Psychometric curves**

Psychometric functions illustrate the associations between hearing ability and the acoustic levels of stimuli, which move up with an increase in the stimulus level. Based on the psychometric curve, how well each stimulus is recognized can be visualized. The intensity levels of stimuli can be equated with other stimuli to make sure that all of them are homogeneous in terms of speech intelligibility. The psychophysical theory states that when a stimulus increases in intensity levels, sensitivity towards this stimulus increases monotonically (Macmillan & Creelman, 2005). The psychometric function indicates the perceptual sensitivity measured twice—before and after intensity adjustments. The purpose of calculating the psychometric curve is that it can bring an internal balance for each stimulus within the subtest and across the subtests of AAST. Figures 6 to 9 show the psychometric curve before and after balancing the intensity level in noisy and quiet conditions.

### **Psychometric curves before intensity adjustments**

As it can be seen in figure 6, some stimuli seemed to be easy but the others were quite difficult for the participants. For example, in a1, the words “orange”, “chain”, and “bird” were too difficult. These were modified by increasing the intensity level of the test files from 2 to 3 dB. The rest of the test was quite easy. Thus, the words were adjusted by reducing the intensity level by 1 dB. In a2, the words “pill”, “bag”, and “earth” were somewhat difficult, and needed an increase of about 2 dB in the intensity level. The other words, “fireworks” and “wave” were too easy. So, the intensity of the files was reduced by 1 dB. In a3, the word “pot” was somewhat difficult. It was modified by

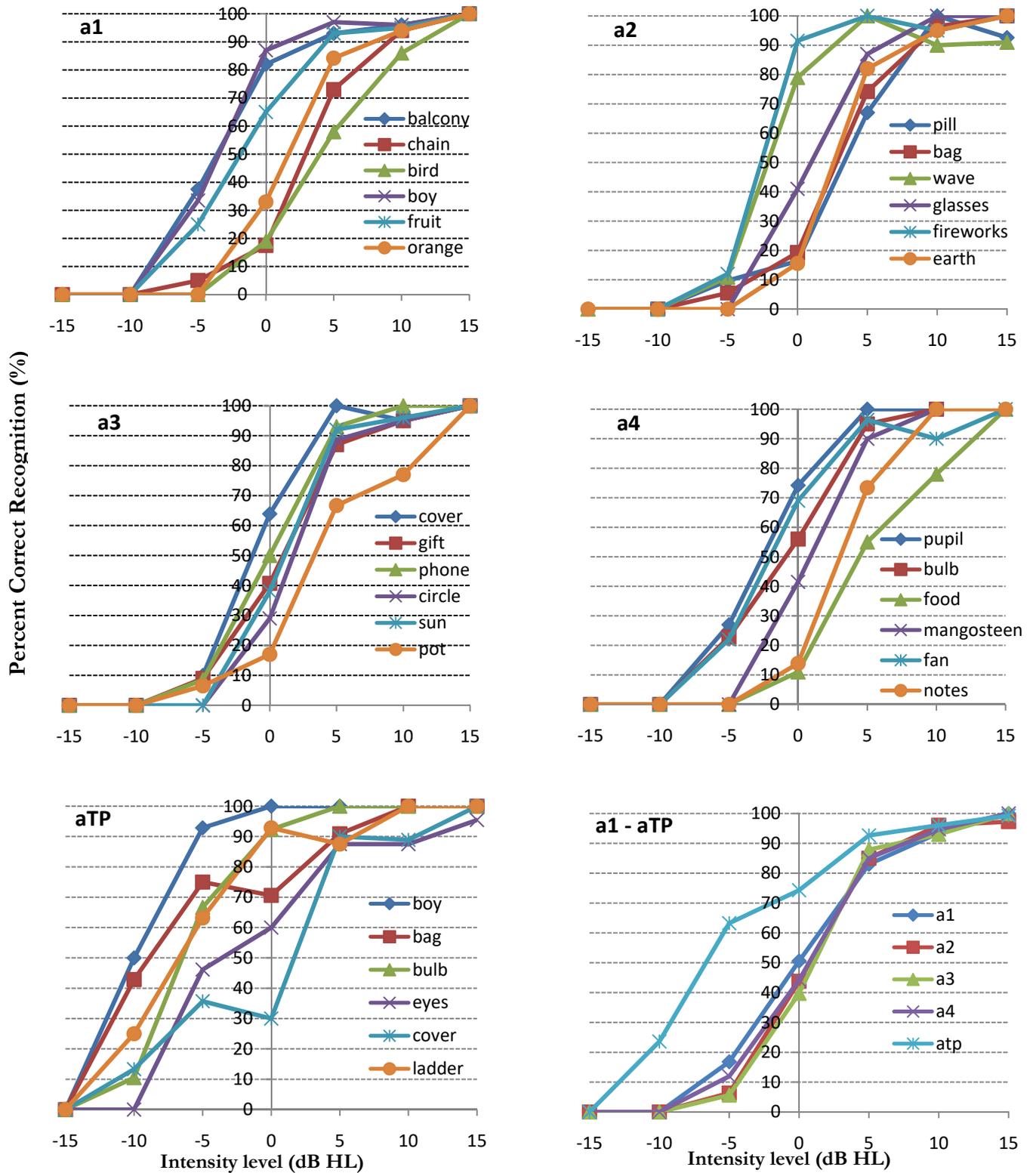
increasing the intensity of the test files by 1–2 dB. In a4, the words “pupil” and “fan” were too easy. So, these were changed by decreasing the intensity of test files by 1–2 dB. In contrast, the words “food” and “note” were too difficult. These were corrected by boosting the intensity by 2–3 dB. In the aTP, all stimuli seemed to be easy. Thus, these were also modified by decreasing the intensity level of the test files by 2 dB. In quiet, the average of the slope steepness for the 30 stimuli at the 50% threshold resulted in almost 7.5%/dB (Figure 6).

We have just mentioned the psychometric features in quiet with regard to psychometric features in noise (Figure 7). There are similarities in the psychometric curves in noise and in quiet. The speech stimuli that seemed to be difficult for the listener to recognize in quiet also appeared to be difficult for the listeners to recognize in a noisy condition. In relation to psychometric function slopes, the average slope values in noise before internal balancing is 9.7%/dB for the subtest of AAST, which was higher than those in quiet.

Based on the analysis of psychometric functions before intensity adjustments, an intensity-level correction was made to the speech stimuli in AAST. This was to make sure that the subtests of AAST received after the internal balances were equal in terms of speech intelligibility. In all, 20 normal-hearing listeners aged between 18 and 22 years were tested by the same procedure as used for the previous measurement (before the intensity adjustments). The analyses showed new psychometric curves (Figure 8 for quiet, Figure 9 for noise) that are more homogeneous with respect to speech intelligibility between the stimuli within an AAST and across AASTs.

The psychometric function slope values of AASTs were relatively equal between quiet and noisy conditions. In quiet, the steepness value was 8.2%/dB whereas the steepness value was 8.4%/dB for noise. Compared with other works, the slopes of the psychometric curves in the current research are somewhat gentler than those in AAST Ghanaian (Offei, 2013), which had a slope value of 10.2%/dB. They are also gentler than those in OlKiSa, which had a slope value of 13%/dB (Wagener et al., 2005). The differences in slope values between the present study and others can be ascribed to linguistic factors that make speech less intelligible in the speech stimuli (Wagener & Brand, 2005; Zokoll et al., 2013). Indeed, as compared with tonal languages, the psychometric slope functions of the Vietnamese AAST are close to the slope values reported for speech material developed in Thai and Cantonese. Hart (2008) found slopes between 8.6%/dB and 9%/dB for a list of 28 disyllabic words in Thai. Similarly, Nissen et al. (2011) found a slope of 7.6%/dB for a list of 28 disyllabic words in Cantonese.

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE



**Figure 6:** Psychometric curves of the subtests of AAST in quiet before the intensity adjustments; slope 7.5%/dB

DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

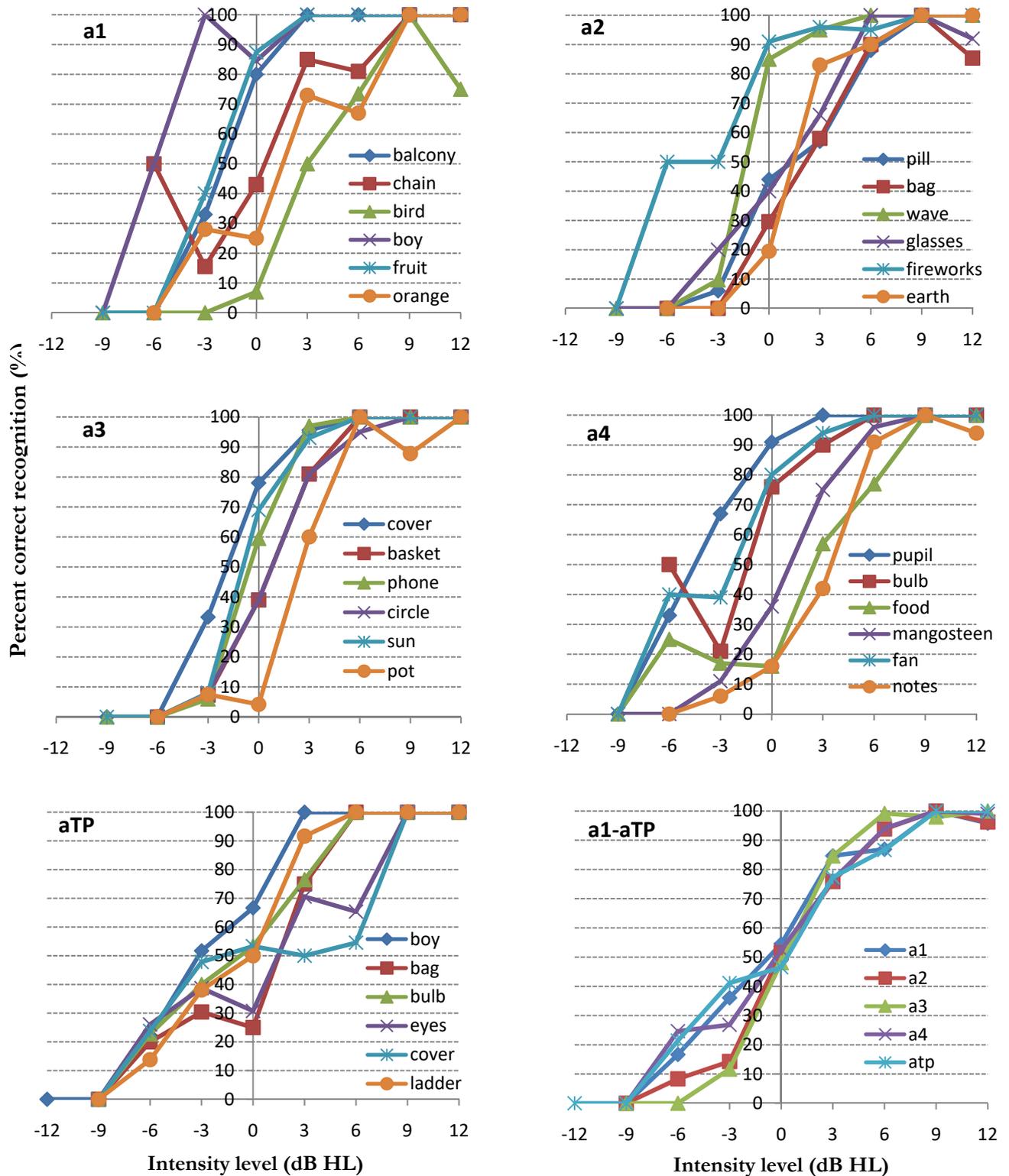


Figure 7: Psychometric curves of subtests of AAST in noise before intensity balances; slope at 50%=9.8%/dB

DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

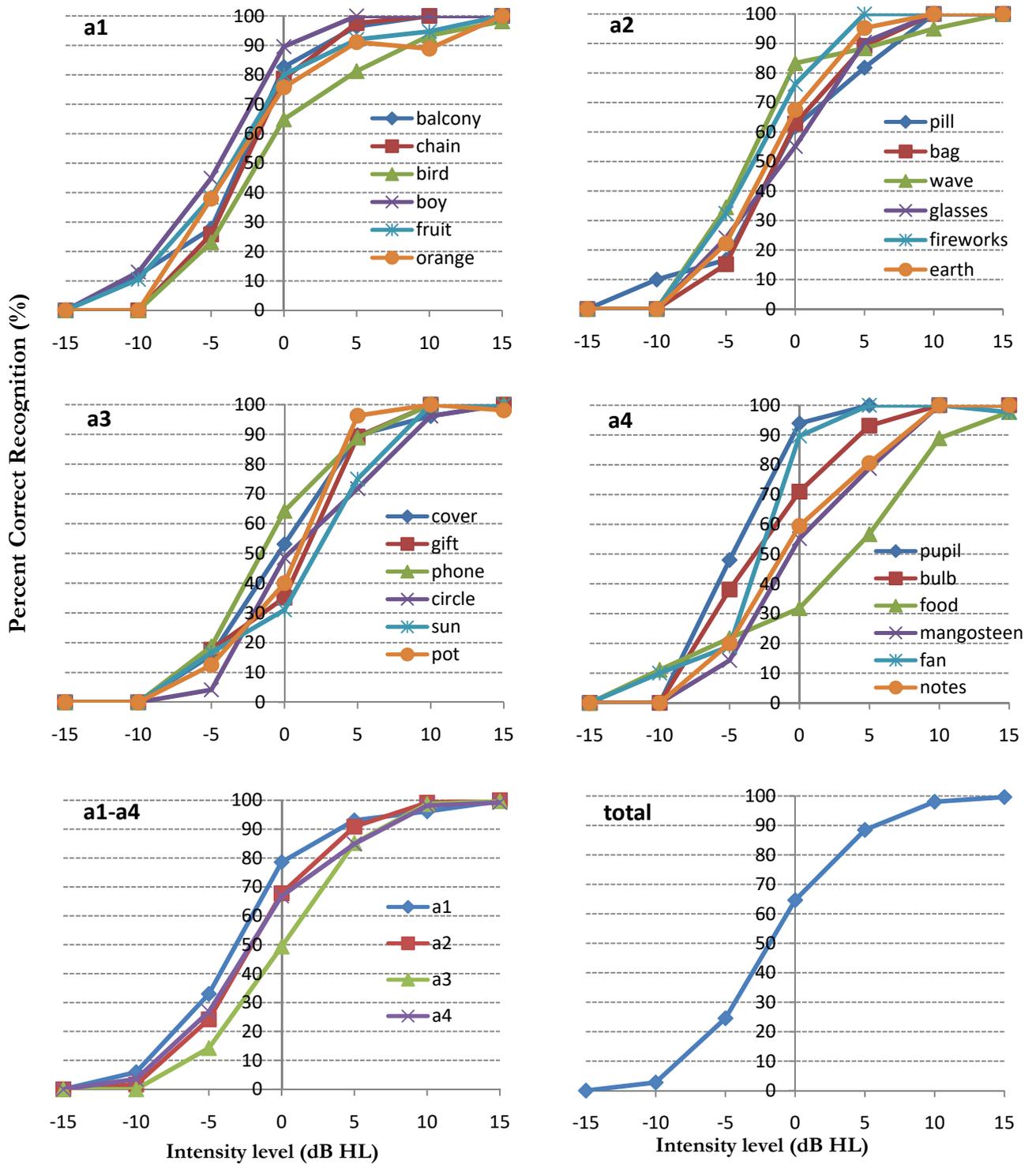
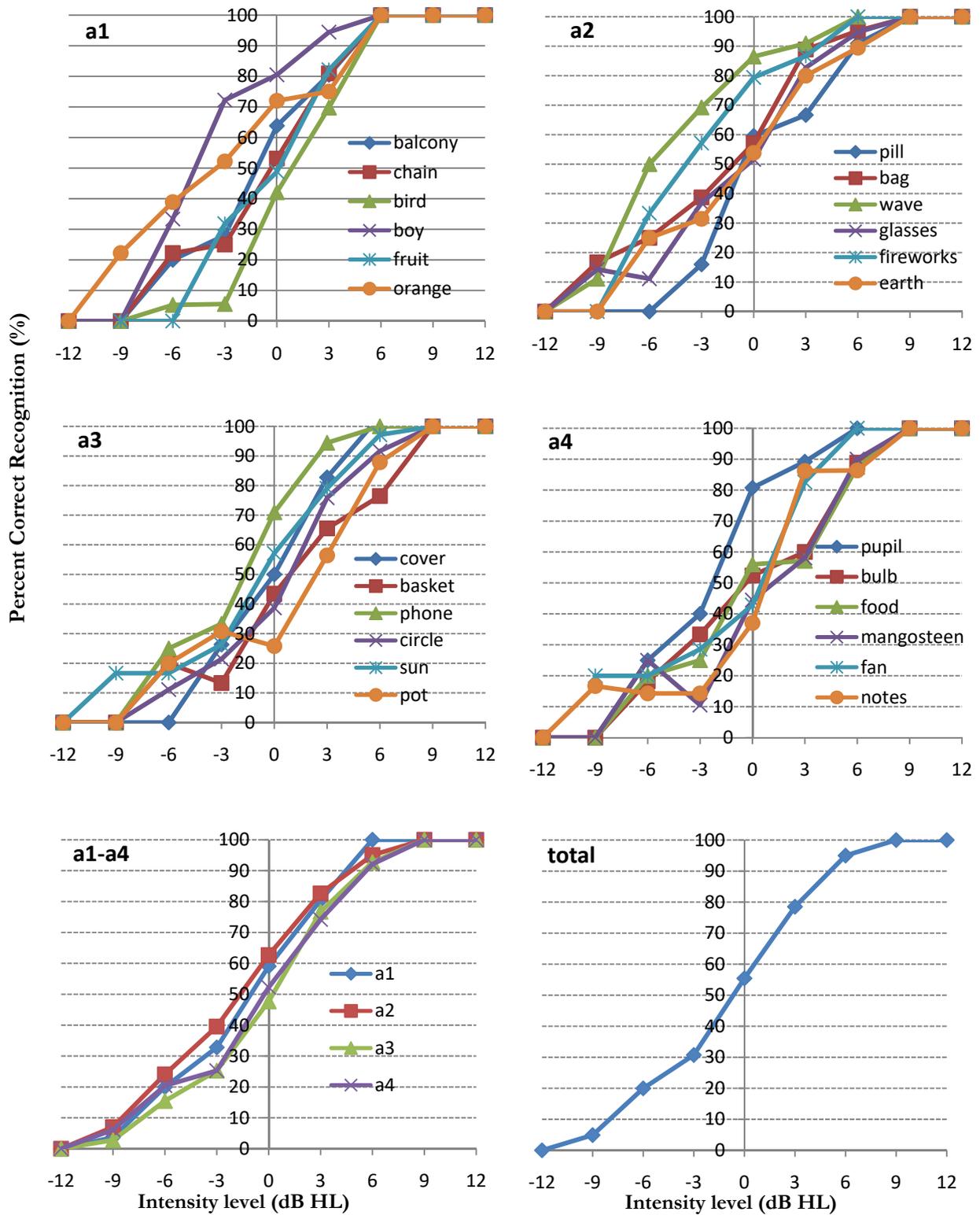


Figure 8: Psychometric curves of the subtests of AAST in quiet after intensity adjustments; slope=8.2%/dB

DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE



**Figure 9:** The psychometric curves of the subtests of AAST after intensity corrections in noise, the slope=8.4%/dB

**Homogeneities of mean SRTs among AAST subtests**

The data collected after intensity adjustments were used to assess homogeneities in speech thresholds across the subtests of AAST (a1 to aTP). Table 4 illustrates the threshold for the individual speech test. The mean SRTs across the subtests of AAST were more than 5 dB away from the expected threshold values, which are expected roughly at  $25 \pm 5$  dB (SPL) and  $-16 \pm 3$  dB (SNR).

**Table 4: Mean SRTs for all five subtests of AAST**

Speech materials	[dB SPL] $\pm$ SD	[dB SNR] $\pm$ SD
AAST-a1	$31.0 \pm 2.3$	$-14.0 \pm 1.1$
AAST-a2	$33.4 \pm 2.0$	$-12.5 \pm 1.0$
AAST-a3	$36.0 \pm 2.0$	$-9.0 \pm 1.1$
AAST-a4	$36.0 \pm 2.2$	$-10.1 \pm 1.0$
AAST-aTP	$31.0 \pm 2.5$	$-14.0 \pm 3.0$

Based on the expected threshold values, the speech stimuli in the speech test had to be modified in terms of the intensity levels to make sure that the mean SRTs were closer to the expected SRTs. The adjustments made to the stimulus intensities in each speech test were different from each other. The maximum adjustment was 5.8 dB, and the minimum adjustment was 1.8 dB. However, the amount of intensity level correction per speech test needed to be kept within bounds to preserve the natural speech signals. With the level adjustments, the average SRTs of  $25 \pm 5$  dB SPL in quiet and  $-16 \pm 3$  dB SNR in noise could be predicted for the speech material.

**Confusion analyses in the five subtests of AAST**

To examine whether a speech stimulus within a subtest was well recognized, we made a word confusion analysis. Aside from the two bars “?” and “t.o”, the six bars of the listener’s answers for each stimulus showed the number of times the listeners chose a response. Figures 10 and 11 show the word confusion of the four tests of AAST. In a1 (top panel), listeners often confused between “balcony” and “fruit”, and “orange” and “fruit”. However, the number of confusions was trivial. In a2 (bottom panel), most of the listeners responded to the word “bird” with “?” (did not hear or very unsure about the answer). In a2, the proportion of confused words was less than those in a1. Most listeners responded to the word “pill” with “?”.

In a3 (top panel), the most confused words were “circle” and “cover”. These were confused seven times. However, this number is negligible. In a3, the word “pot” was responded with “?”. In a4 (middle panel), a lot of confusion occurred between the words “bulb” and “pupil” (17 times), and a little lesser between “food” and “mangosteen” (12 times). In addition, the word “food” was most frequently responded with “?” compared with the rest of words in a4. Lastly, for aTP (bottom panel), more confusion happened between the words “bag” and “eyes”. This was probably due to the similar tonal patterns. Similarly, the word “cover” was misidentified as “ladder”. The word “cover” also drew the most “?”s from the majority. It differed a lot from the word “cover” in a3.

However, there was no word that the listeners confused very often (>30–40%) with another. This indicated that the words in the four AASTs in Vietnamese were suitable for the actual hearing measurement. Of course, after modifying the intensity level of the test files, which we have stated earlier, the AAST words might be more balanced and better in terms of speech intelligibility.

In short, the Vietnamese AAST was the first screening test for this language. This speech test was designed on the basis of the procedures used in the existing AASTs in German, Polish, and Ghanaian. The pilot results showed that the subtests of AAST in Vietnamese matched the other AASTs well regarding speech intelligibility.

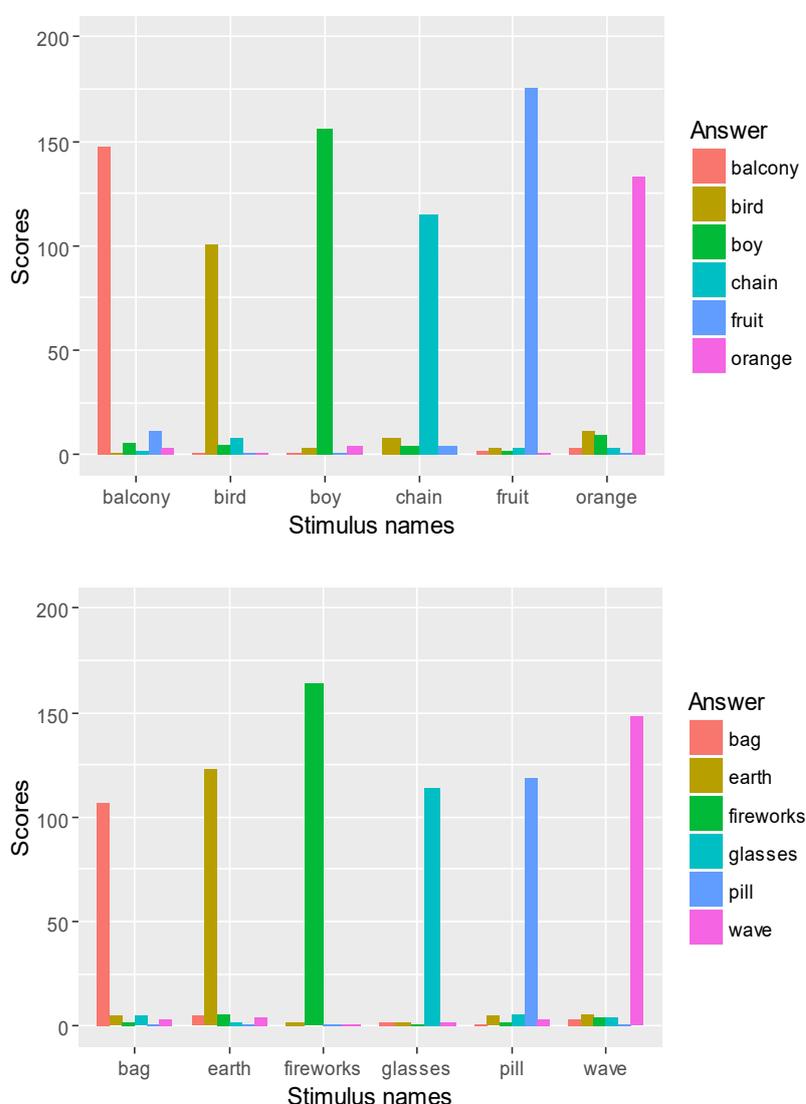


Figure 10: The confusion analyses of the subtests a1 and a2

DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

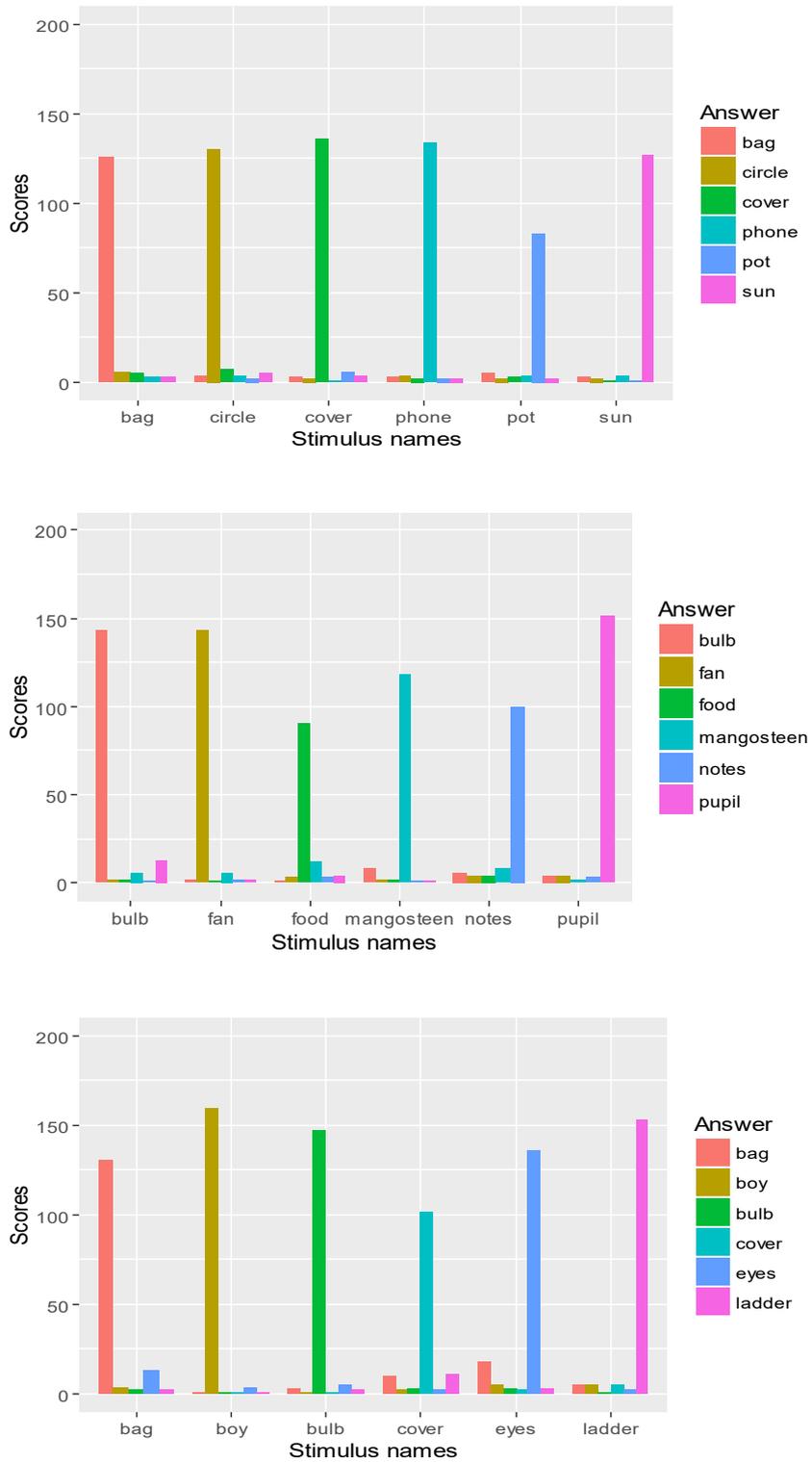


Figure 11: The confusion analyses of the subtests of a3, a4, and aTP

### 3.2 Designs of the speech material of NAMES

#### 3.2.1 Phoneme frequencies in Southern Vietnamese—in spoken text

The construction of the NAMES test was based on phoneme distribution in SVN. Instead of using the phoneme frequency in the written language, this speech material was based on the phoneme allocation in the spoken language. As Carhart (1951) stated, the distribution of phonemes in the speech test must match that in the spoken language.

To establish phoneme frequency occurrences in the SVN spoken text, spoken samples, including interview texts, were gathered from an in-depth interview of a PhD project<sup>2</sup>. The interviewees, aged between 28 and 60, were mostly born and brought up in southern Vietnam. They were different kinds of people: farmers, officers, journalists, and lawyers. Each recording session took around 30 minutes. All sound files were recorded and stored separately. Ultimately, all were converted into word files. The numbers of syllables were 136,129 units, consisting of 362,527 phonemes. The phoneme frequency in the spoken language is shown in Appendix A. In this section, the frequency for the selected phonemes that were used to establish the NAMES test is illustrated.

#### 3.2.2 Phoneme selections and nonsense disyllabic CV-CVC structures

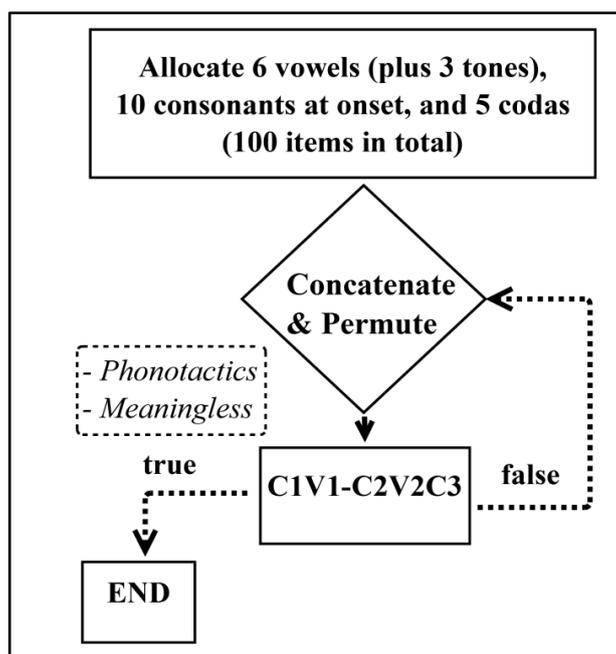
The NAMES test, which comprises nonsense disyllables, is constructed depending on the syllable structures of Consonant-Vowel-Consonant-Vowel-Consonant type ( $C_1V_1C_2V_2C_3$ ) along with their tonal patterns. To achieve the lists of nonsense disyllables, which are phonetically different, balanced, and representative of everyday languages, the most frequently used phonemes in the spoken text were chosen for the onsets, nuclei, and codas. Frequently used tones were selected. The initial phonemes included /k, d, t, t', l, v, m, h, b, j/. The vowel phonemes consisted of /a, o, ɔ, ɔ̃, ɛ̃, i/, and the final phonemes comprised /-k, -t, -m, -ŋ, -j/ in conjunction with three tones *ngang* (A1), *huyền* (A2) and *sắc* (B1). The consonant phonemes included 60 items for three positions ( $C_1$ ,  $C_2$ , and  $C_3$ ), with 20 items each. The nuclei contained 40 phonemes, which were equally distributed among the positions of  $V_1$  and  $V_2$ . First, these phonemes were computed into an Excel file based on their position: the initial ( $C_1$ ), medial ( $C_2$ ), and the final one ( $C_3$ ). Second, an initial consonant phoneme was coupled with one in six vowel phonemes in conjunction with two tones for a two-syllable combination. Next, the phonemes were permuted in a random order to create four lists of disyllables, with 20 items each. Every speech material is made up of the same onsets, vowels, codas, and tones to make sure that the phonetic contents are similar across the lists of syllables.

<sup>2</sup> The project is being conducted by PhD student Tran Tu Van Anh of the University of Bonn (forthcoming 2017)

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

Finally, a native listener of Vietnamese (a linguist) inspected whether these disyllables follow phonotactic rules in Vietnamese or whether these are meaningful syllables in the language. For example, the tone B2 and A1 never come with final stops of /-k and -t/. The short vowel /ə/ never occurs in a syllable without a final sound. To solve such cases, the unqualified syllables were left out and substituted by a corresponding syllable extracted from the same row in the Excel file. The flowchart is shown in Figure 12. It shows the steps followed to develop the words. The speech test included 80 nonsense disyllables, divided into four sublists of 20 each (see Appendix B). In the SVN, the onset of labiodentals “v” is spoken like phoneme /j/, as a representative of “v” in transcription.

As mentioned earlier, to develop NAMES, we selected phonemes that frequently occur in the spoken text. It was assumed that phonetic correspondences between the speech material and the particular language would increase the validity of the speech material. The phoneme proportion of the NAMES words in each sublist has been shown in tables 3.5 a–b in comparison with those in the spoken text.



**Figure 12:** The flowchart illustrating the steps to create meaningless  $C_1V_1-C_2V_2C_3$  syllables

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

**Table 5:** Proportion of consonant phonemes in NAMES test

Phoneme	Occurrence	Percentage	Expected	Rounded	C <sub>1</sub> -C <sub>2</sub>	C <sub>3</sub>
/k/	17589	11.5	9.1	9	4	5
/d/	13054	8.5	6.8	7	7	
/t/	11394	7.4	5.9	6	3	3
/t'/	9776	6.4	5.1	5	5	
/l/	9756	6.4	5.1	5	5	
/v/	9184	6.0	4.8	5	5	
/m/	8818	5.7	4.6	5	3	2
/h/	8793	5.7	4.6	4	4	
/b/	7711	5.0	4.0	4	4	
/-ŋ/	23829	27.0	5.4	5		5
/-j/	20473	23.0	4.6	5		5

Onset: C<sub>1</sub>, C<sub>2</sub>; coda: C<sub>3</sub>

**Table 6:** Proportion of vowel phonemes in NAMES

Phoneme	Occurrence	Percentage	Expected	Rounded
/ɔ/-B1	9862	14.2	5.7	5
/a/-A2	8306	12.0	4.8	5
/o/-A1	6856	9.9	4.0	4
/i/-A2	6374	9.2	3.7	4
/a/-A1	6104	8.8	3.5	3
/a/-B1	5555	8.0	3.2	3
/ɔ/-A1	3772	5.4	2.2	2
/ə/-A1	3769	5.4	2.2	2
/ə/-B1	2800	4.0	1.6	2
/i/-A1	2655	3.8	1.5	2
/ă/-A1	2509	3.6	1.4	1
/o/-A2	2504	3.6	1.4	1
/ă/-A2	1805	2.6	1.0	1
/i/-B1	1778	2.6	1.0	1
/o/-B1	1234	1.8	0.7	1
/ə/-A2	1152	1.7	0.7	1
/ă/-B1	1150	1.7	0.7	1
/ɔ/-A2	1110	1.6	0.6	1

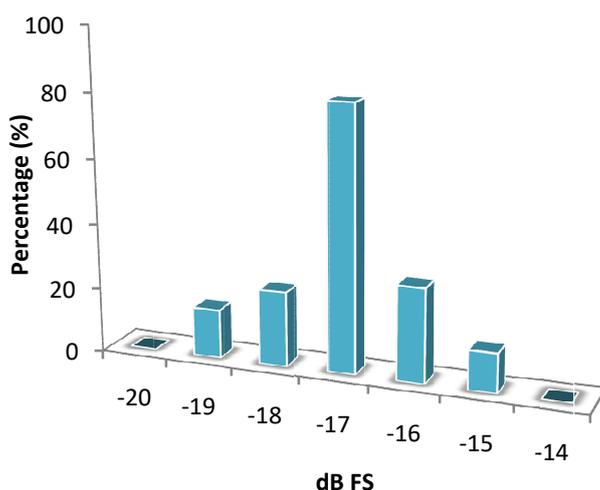
Each vowel phoneme (V) combines with one of the three tones: A1 (High level tone), A2 (low-falling tone), and B1 (high rising tone).

### 3.2.3 Stimuli recordings

All meaningless disyllables were read out by a 40-year-old woman who spoke standard SVN. The speaker was asked to pronounce four lists of C<sub>1</sub>V<sub>1</sub>-C<sub>2</sub>V<sub>2</sub>C<sub>3</sub> syllables with the following instructions: to maintain constant intonation during the recordings, to avoiding an asking intonation, and to maintain a reading

speed within a sublist and among the sublists in a natural pronunciation. The acoustic stimuli were recorded as a mono with sounds digitized at 44.1 kHz and 24 bits and saved separately in PCM wave file format. Avoiding background noises, the recordings were arranged in a sound-treated room of VOH with the ambient noise around 25 dBA.

Then, each stimulus was modified with a 50 ms silent duration between the first and second syllables (Le et al., 2011). The aim was to ensure that all two-syllables were uttered as they are done in daily conversations. The syllable durations were also adjusted. Next, to make sure that all 80 stimuli were balanced in terms of energy, the RMS values of each stimulus (first and second syllable) were measured with Cool Edit Pro 2.1. Each syllable in each disyllable combination was equalized at a similar total RMS level. Figure 13 shows the distribution of RMS for 80 stimuli of NAMES. Mostly, the stimuli were close to -17dB.



**Figure 13:** Distribution of energy (total power) in the stimuli of NAMES

### 3.2.4 Phoneme categories in the speech material of NAMES

The aim of this study was to examine phoneme and tone identification scores by older listeners. We assumed that older listeners would be less accurate in performances as stimuli carry a high-rising tone, as compared with a low-falling or high-level tone. Furthermore, fricative sounds presumably would be more challenging to recognize for older listeners. Based on these assumptions, the phoneme categories of NAMES were grouped on the basis of manners of articulation. The vowels were combined with three tones, A1, A2, and B1, in combinations of vowel-plus-tone.

- (1). Vowel phonemes /a, o, ɔ, ə, ă, i/ plus the high level tone (A1)
- (2). Vowel phonemes /a, o, ɔ, ə, ă, i/ plus the high-rising tone (B1)
- (3). Vowel phonemes /a, o, ɔ, ə, ă, i/ plus the low-falling tone (A2)

The consonant phonemes were classified into five subcategories based on the manners of articulation.

- (4). Voiced stops /b, d/
- (5). Unvoiced stops /t, k, t<sup>h</sup>/
- (6). Laterals and nasals /m, ŋ, l/
- (7). Fricatives /j-, h/
- (8). Semi-vowels /-j/

### 3.2.5 Preparations of the pilot test

The 200 speech stimuli of NAMES were imparted to five normal-hearing subjects, who were students at the University of Bonn (Germany). The pilot test aimed firstly, to examine the validity of NAMES test, and secondly, to observe the listener's behavior to the nonsense speech stimuli.

The speech stimuli were presented binaurally in the listener's ears through the headphones HAD 280 with the presentation level at 80 dB (SPL). Then the listeners were asked to iterate the speech stimuli they had just heard. The data were extracted from the Excel file CSV to compute the PRS. It is the number (percentage) of phonemes correctly identified out of 100 phonemes per list. As expected, the subjects were interested in our pilot test due to the novelty of the speech stimuli.



Figure 14: A screenshot of the NAMES test

The normal-hearing listeners obtained a grade of “highly recognizable” and an average PRS of 98%. This PRS was in line with those in Kuk et al. (2010), which had an average score of 98% for male speakers and 97% for female speakers. The number of correct words in the current study was 19 out of 20 tokens. The observed results revealed that all NAMES stimuli are initially suitable for the actual phoneme recognition measurements.

Figure 14 illustrates an example of a test screenshot during administration. After the listeners’ responses, the tester or helper would type the listener’s responses into the box(es) corresponding to each phoneme. The tester clicks on either the individual virtual keyboard or the box “All correct” if the listeners had 100 % correct responses—or “All wrong” if the listener had no response at all. The listeners were encouraged to ask for a reiteration of the speech stimulus if they could not hear it obviously. Once the response was recorded clearly, the next speech stimulus would come to the listener’s ears.

In each list of 20 words, the software automatically tabulated the total number of words, phonemes, and vowel and consonant categories, and ensured that they had been correctly repeated. Furthermore, the speech stimuli and the corresponding responses were saved on the computer’s hard drive for each list of the test. The listeners’ responses were also recorded for further assessment.

The NAMES test is a supra-threshold speech-screening test. Its scores are calculated as the percentage or number of phonemes correctly recognized. The test is built based on the nonsense CVCVC structures, which are independent of the individual’s literacy and education (Cooke et al., 2010). The speech test is also suitable for non-native listeners who have a little experience of the language being tested (Paglialonga et al., 2014). The nonsense speech test is independent of the listener’s cognition, too (Akeroyd, 2008). These include short-term memory or speech processing, which is considered to be the causes of deterioration in speech recognition, particularly in older listeners (Gordon-Salant, 2005). The authors of this research found the above-mentioned factors interesting.

### **3.3 Summary**

In summary, this section has described the designs of the two speech materials: AASTs and NAMES. AAST consists of five subtests with six two-syllable noun phrases each. It is well established and has been carefully assessed through pilot studies. The NAMES test was initially developed for Vietnamese. Eighty meaningless disyllabic structures are divided into four lists of 20

## DESIGNS OF THE SPEECH MATERIALS IN VIETNAMESE LANGUAGE

items each. Ideally, this kind of speech test is used to determine supra-threshold PRSs for listeners who are older children, adults, and older adults. However, it might be unsuitable for younger children. The initial assessments of the two speech tests indicated that these speech materials could serve as speech stimuli for further measurements. The next section will move on to the details of the methods used in the present study.

## 4. OBJECTIVES AND RESEARCH QUESTIONS

### 4.1 Objectives

The preceding introductory section has given some indication of the uses of the study. The objectives of the current research are proposed in detail:

- To standardize and validate two speech audiometry materials of AAST (in quiet and noise) and NAMES by groups of native listeners of SVN.
- To examine the validity of using the two speech audiometry materials to assess SRTs and PRSs by non-native listeners of SVN who speak NVN and CVN.
- To investigate SRTs and phoneme identifications by older native listeners under the effects of tonal patterns (F0) of syllables.
- To determine the correlations between SRTs, PRSs, and duo-tone thresholds in quiet and noise for normal-hearing listeners of SVN.

### 4.2 Research questions and hypotheses

Specifically, the following research questions (RQ) and hypotheses (H) are raised to achieve the aforementioned objectives.

**RQ1.** What are the age-related norm values for Vietnamese AAST and are these values similar to those in other languages?

**H1a.** *The norm values depend on the listeners' age. Children and older adults perform significantly poorer on speech-threshold values than adults do.*

**H1b.** *The age-related norm values in Vietnamese are analogous to those in other languages.*

**RQ2.** Is AAST a “simple test” because learning effects are (almost) non-existent?

**H2.** *Learning effects are non-existent in the Vietnamese AAST due to the similar speech-threshold results among three test trials.*

**RQ3.** Do tonal patterns (F0) in terms of lexical tones in Vietnamese have any effect on the speech recognition of older native listeners with high-frequency hearing loss?

**H3.** *Tonal patterns (F0), in terms of lexical tones in Vietnamese, have an effect on speech recognition of older native listeners with high-frequency hearing loss.*

- *AAST: Speech threshold values are different across the subtests of AAST under the effects of tonal patterns in quiet and noise.*

- *NAMES: Scores of tonal (phoneme) identification differ among the three tones.*

**RQ4.** What are the age-related norm values for Vietnamese NAMES and are these values similar to those in other languages?

**H4b.** *The norm values depend on the listener's age, in which older adults perform significantly poorer regarding phoneme scores than younger adults do.*

**H4b.** *The age-related norm values for NAMES in Vietnamese are similar to those in other languages.*

**RQ5.** Which are relevant for dialectal aspects in AAST and NAMES?

**H5a.** *There are significant differences in speech threshold values in AAST between the native and non-native listeners of the dialect.*

*(The native listeners achieve better SRT values than non-native listeners)*

**H5b.** *There are significant differences in the response time in AAST between the native and non-native listeners of the dialect.*

*(The native listeners need a shorter response time than non-native listeners do)*

**H5c.** *There are significant differences in phoneme recognition scores in NAMES between the native and non-native listeners of the dialect.*

*(The native listeners obtain better phoneme scores than the non-native listeners)*

**RQ6.** What are the interdependencies of AAST and PTA, NAMES and PTA, and AAST and NAMES?

**H6a.** *There are strong associations between duo-tone and AAST thresholds in quiet and noise.*

*(The better the speech threshold value, the lower the duo-tone threshold value)*

**H6b.** *There is an association between the duo-tone threshold and the NAMES score.*

*(The higher the phoneme score, the better the duo-tone threshold)*

**H6c.** *There are strong relationships between the NAMES scores and the AAST thresholds.*

*(The better the speech threshold value, the higher the phoneme score)*

To attain the proposed objectives and address the research questions, the speech materials of AAST and NAMES were designed to assess speech perception by the native and non-native listeners of SVN. The primary focus of AAST is to “assess the skills at the detection level as well as at the discrimination level” (Offei, 2013, p. 65). The focus of the NAMES test is to measure the facility of phonemic identification and differentiation above an individual's threshold.

In the designs of new speech materials, norm values are expected to provide a “benchmark” for further assessments as a reference value. Thus, the normative data were collected from the normal-hearing native listeners of SVN using the two speech materials of AAST and NAMES. Learning effects reflect the reliability of the speech materials so that minimal learning effects are expected. The previous studies conducted by Mohammed (2010), and Offei (2013) have provided unclear findings. No learning effects were found in Mohammed’s work in contrast to Offei’s works, which sometimes found the existence of learning effects, and sometimes not. This is of special interest in the current study.

In speech audiometric testing, auditory perception of speech differs from client to client due to their linguistic backgrounds, such as language competence, and dialect variation, which contribute to the intelligibility of the clients being diagnosed. Hence, dialect differences can, for example, lead to misidentification of the types and levels of hearing loss. No information is currently available on this issue in Vietnamese audiology. Therefore, the aspect of dialectal effects on auditory perception of speech is of special interest in this thesis. Data on dialectal effects were gathered from two groups of non-native listeners of NVN and CVN.

Vietnamese is a tonal language. The identity of an individual syllable or word is based on its tonal pattern and the phonetic structure. The identity of an individual tone depends on the changes of pitch patterns that a syllable carries (Blicher et al., 1990). Thus, the tonal pattern (F0) is the primary cue for tonal identification. Two speech materials were used to explore how older native listeners of SVN perceive disyllabic noun phrases and nonsense disyllables under the effects of tonal patterns of syllables. For audiometric testing, the effect of tonal patterns on hearing performances have not yet been considered in the Vietnamese language, especially for older listeners with high-frequency hearing loss. We did not know how well the older native listeners perceptually identify speech stimuli at different pitch levels and pitch contours. Also, we did not know the extent to which the tonal pattern of syllables manipulates the hearing performances of older listeners. AAST and NAMES were administered to groups of older native listeners aged between 65 and 85 years. In the research on tonal effects, the term “tonal pattern” is defined only in terms of syllable tones, not word tones.

Aside from the mentioned themes, the dissertation also assessed the relationship between speech audiometry materials and duo-tone audiometry.

## 5. METHODS

### 5.1 Speech test of AAST

The AAST was designed to determine normative values in SRT by Southern listeners. The speech material was also used to determine whether the listeners' speech recognitions were affected by dialectal variations and lexical tonal patterns (F0) of syllables. With respect to the learning effects on the speech performance, AAST was also used to examine whether the learning effect still existed in the Vietnamese AAST. Below are the detailed descriptions of participants, speech materials, and test procedures in the current research.

#### 5.1.1 Speech stimuli

This AAST comprised the following five subtests: a1, a2, a3, a4, and aTP. The speech test a2 was primarily used to determine the normative values, dialectal effects, and tonal pattern effects. Along with a2, the remaining speech tests (a1, a3, a4, and aTP) were also used to determine the effect of tonal patterns on the speech thresholds. The subtest a1 was mainly conducted to investigate the learning effects, along with a2.

An acoustic analysis of the tonal patterns of speech materials was necessary to measure the fundamental frequency values of each speech stimulus. The analysis illustrated a detailed look into the differences in pitch heights and pitch contours across speech materials. Values of pitch heights and formants for every stimulus (Figures 15 and 16) were extracted from the phonetic software Praat (Boersma & Weenink, 2013). For a1, pitch values of all six words ranged between 160 and 290 Hz with an average of 230 Hz, and the pitch contours remained equal across stimuli. For a2, the pitch values were measured in a range of 194 to 450 Hz, averaged at roughly 270 Hz, with the pitch contours rising at the ending pitch (in the case of tone B1). For a3—low pitch level and falling contour tones—the pitch value was between 140 and 380 Hz, and averaged around 235 Hz. For a4 and aTP, the pitch levels were somewhat equal with wider ranges of 150 to 500 Hz (a4), and 150 to 450 Hz (aTP). The F0 values were equal—roughly 240 Hz—and the pitch contours were between falling and rising.

The remarkable distinction in pitch level values could be easily captured between the speech tests a1, a2 (with a high pitch level and high-rising contour), and a3 (with a low and low-falling contour). The remaining speech tests of a4 and aTP carried both low and high tones. Hence, the mean F0 values for both were equal. In general, the F0 values were below 500 Hz across subtests of AAST. The

measured F0 values in the speech materials were comparable to those of the findings by Pham (2003), and Brunelle (2003).

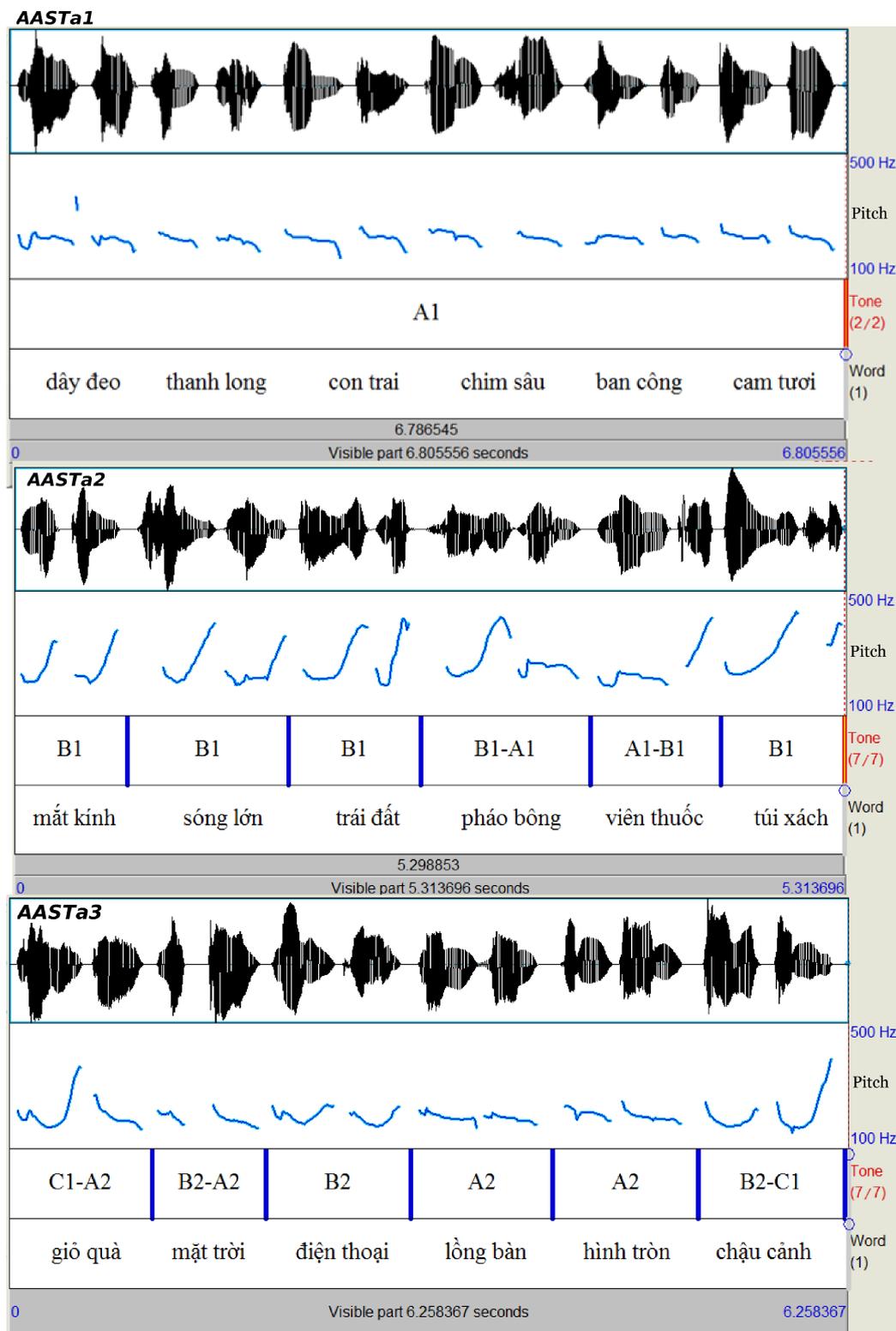
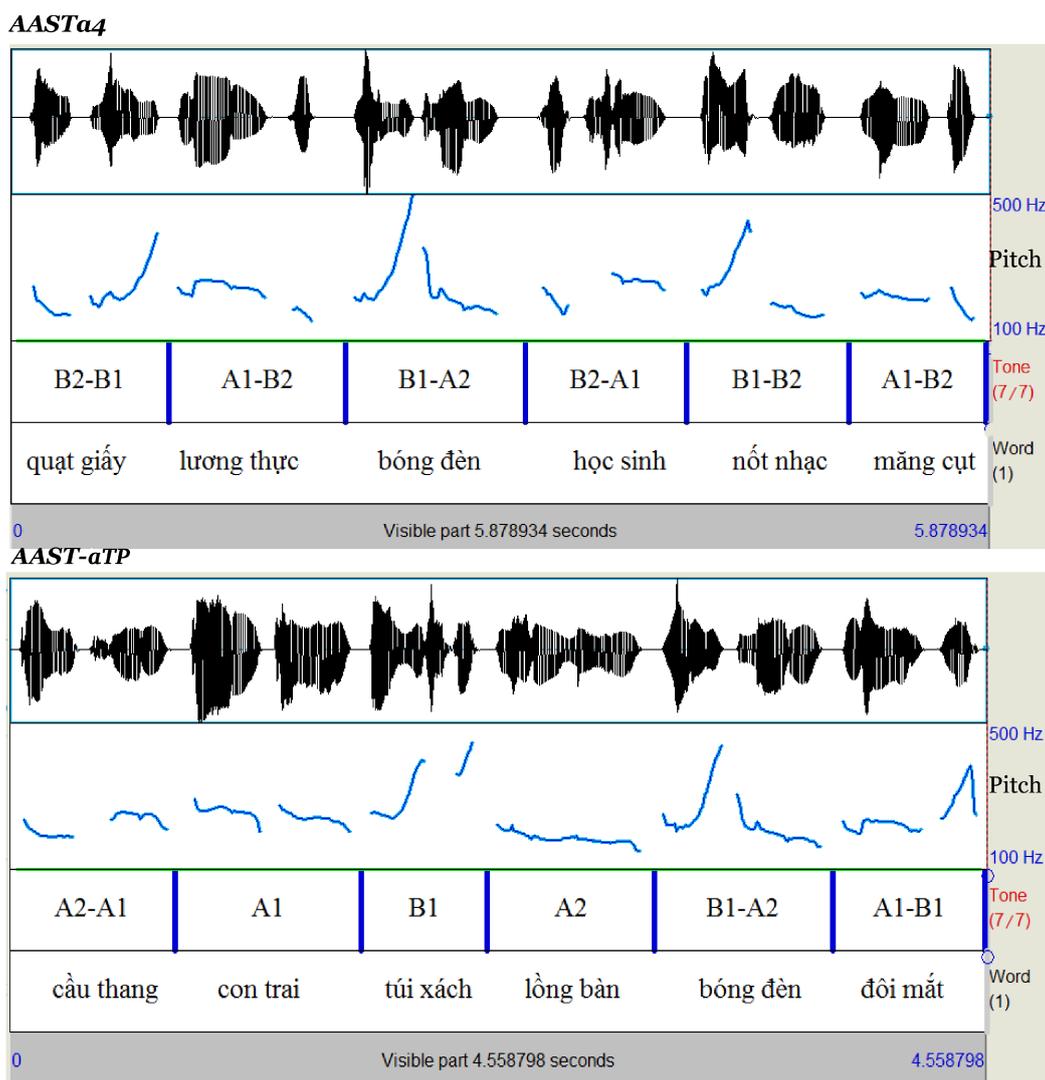


Figure 15: Amplitude waveforms and F0 of speech stimuli, a1 to a3



**Figure 16:** Amplitude waveforms and F0 of the speech stimuli, a4 and aTP

### 5.1.2 Listeners

The AAST was administered to 435 normal-hearing listeners aged between four and 85. Otoscopy and duo-tone audiometry (octave frequencies 0.5, 4 kHz) were carried out for each listener. Only listeners who have average duo-tone thresholds lower than 30 dB HL (children, adults) and lower than 45 dB HL (older adults) were recruited in this study. A large number of listeners were excluded from the measurement due to hearing impairment, cognitive decline (among older listeners), incomplete hearing measurement due to shyness, lack of attention, or misunderstood task (younger children). The hearing measurements were based on the Declaration of Helsinki (6th).

### Native listeners

In all, 200 listeners aged between four and 85 took part in the study on normative values. They were divided into the following three subgroups depending on the listeners' age: children (4–8 years), youth-adults (15–40 years), and older adults (55–85 years). All participants were born and brought up in the south of Vietnam. They mostly lived in Đồng Tháp Province at the time of the study. Some lived in the provinces of An Giang and Tiền Giang and Hồ Chí Minh City.

**Children:** The sample consisted of 74 mono Vietnamese children, divide into three age groups: four-year-olds ( $n=24$ , girl=10, mean age=4.4 years), six-year-olds ( $n=29$ , girl=15, mean age=6.4 years), and eight-year-olds ( $n=21$ , girl=9, mean age=8.4 years). The four-year-old children were preschoolers at the kindergarten of Thái Hòa. The two remaining groups comprised elementary students at the Primary School of Thực hành Sư Phạm. Besides, 12 six-year-old children (mean age=6.5) were recruited for the study of the learning effect. The 12 children were tested only on their dominant ear, based on their duo-tone thresholds.

**Youths and adults:** Sixty-six listeners aged between 15 and 40 years participated voluntarily in the measurement. They were divided into three age groups: 15 to 20 years ( $n=22$ , female=12, mean age=18.3 years), 21 to 30 years ( $n=24$ , female=11, mean age=25.3 years), and 31 to 40 years (female=11, mean age=36.7 years). Among the younger listeners (15 to 20 years), some were high school students. The remaining were undergraduate students at the University of Đồng Tháp. Of the listeners aged between 21 and 40 years, some were students and some were university employees. These participants studied English as an obligatory subject in their school or university. A fourth of them could speak at least basic English.

**Older adults:** The sample included 66 older listeners aged between 55 and 85 years, including three age groups: 55 to 65 years (female=10, mean age=59 years), 66 to 75 years (female=9, mean age=71.4 years), and 76 to 85 (female=9, mean age=80 years).

To study the tonal effect, 108 normal-hearing listeners aged between 55 and 85 were recruited. The test subjects were divided into five groups: Group 1 ( $n=21$ , female=10, mean age=67,  $SD=8.3$ ), Group 2 ( $n=19$ , female=11, mean age=65,  $SD=6.8$ ), Group 3 ( $n=23$ , female=12, mean age=67,  $SD=9$ ), Group 4 ( $n=23$ , female=14, mean age=66.5,  $SD=8$ ), and Group 5 ( $n=22$ , female=12, mean age=67,  $SD=8.7$ ). Each group performed one of the speech materials of a1, a2, a3, a4, and aTP, respectively.

Due to a lack of samples comprising older listeners, especially those above 75, listeners from other areas of the south were recruited. Some came from An Giang province, and some lived in Saigon. The listeners differed from each other in terms of educational and occupational qualifications. Some were Catholic nuns or retired personnel, and had achieved a higher education level. Some were retired handyworkers with limited education. A few of them spoke advanced English or French while the majority (roughly 85%) spoke only Vietnamese. It has to be noted that roughly half of the listeners had retained their daily habit of reading newspapers or playing chess. Due to visional impairment, some others preferred to listen to the radio or watch TV for news instead. A few did not read the news or watch TV at all. In our observations, the older listeners who had retained one of the above-mentioned habits in their daily life or had a good literacy level obtained better SRTs. Our observation was similar to that of Murphy et al. (2016), who stated that the educational level of an individual negatively affects their cognitive ability, indirectly causing a deterioration of speech.

#### **Non-native listeners**

In all, 115 non-native listeners, who originated from North and Central Vietnam, participated in the study on dialectal effect. Apart from the children who might not have been familiar with SVN, the university students or older listeners had more or fewer experiences with the dialect due to either the media or their occupation, which might have needed them to exchange information with native speakers of SVN.

**Northern listener groups:** The group consisted of 59 subjects, including the following subgroups: six-year-olds ( $n=20$ , girl=9, mean age=6.5 years), 20–30-year-olds ( $n=21$ , female=11, mean age=24.6 years), and 65–75-year-olds ( $n=18$ , female=11, mean age=69.2 years). Many of them were born and brought up in Hanoi. A few of the university students and older adults were temporarily living in Hanoi due to their job or education at the time of the study.

**Central listener groups:** Fifty-eight listeners originally from central Vietnam took part in the study. Most of them were born and brought up in the provinces of Nghệ An and Hà Tĩnh. The listeners included three groups: 19 six-year-olds (girl=9, mean age=6.4 years), 21 adults aged between 20 and 30 (female=9, mean age=24 years), and 18 older adults aged between 65 and 75 (female=10,  $M=68.2$  years).

Before the screening test, the listeners' hearing threshold was measured using duo-tone with two-octave frequencies, 0.5 and 4 kHz, for both ears. The audiograms were variable, ranging from 5 to 45

dB HL for these two frequencies (see Appendix J). The younger listeners (except for the four-year-olds) presented a perfect duo-tone threshold in both octave frequencies. The older listeners showed a wider range of duo-tone thresholds from normal to mild hearing impairments, especially on the duo-tone 4 kHz. Due to poor concentration or abstract features of the duo-tone stimuli (Kramer, 2008, p. 204), the four-year-olds performed significantly poorer duo-tone thresholds.

In the current research, the cognitive data of listeners (especially older adults) were not recorded. However, all older listeners were considered to have good cognitive ability according to their performances of duo-tone and AAST.

### **5.1.3 Test procedure**

The author of this thesis and two students, who had been guided for several days, carried out the examination.

The hearing measurements took place in a quiet room where background noise ranged from 38 dB to 48 dBA, measured with a digital sound level meter (GM1351). Furthermore, the background noises were also recorded with a recorder device (Tascam DR-05). It is difficult to find a sound booth to measure the hearing. So, the measurement took place in kindergartens, primary schools, universities (music room, libraries), churches, or even in private rooms (for older listeners). During the hearing tests, all windows and doors were shut to minimize the interferences of environmental odours.

The listeners were measured on the SRTs individually. Each measurement lasted roughly 10 to 15 minutes per person: 10 minutes for adults and 15 minutes for children and the older listeners. As it has been mentioned previously, the duo-tones 0.5, 4 kHz were used for an initial assessment of an individual's hearing capacity to determine whether he/she could carry on with the measurements. For those hard of hearing, an otoscopic exam was suggested to give them a brief idea about the current state of their ears. They were advised to visit audiologists for further assessment.

Before the hearing measurements, the listeners were familiarized with the speech stimuli of AAST in combination with the six pictures that were displayed on the laptop screen. A printed picture with six items was delivered to help the younger children familiarize themselves with the speech stimuli. After that, the listeners received this brief instruction:

“You will hear words that will become softer and difficult to hear. After hearing the word, you must point out the picture. When you do not know the word, press the question mark (?).”

The speech stimuli were presented monaurally through Sennheiser HD 280 headphones. The tester ensured that the headphones adequately covered the listener’s ears. During the test, the tester observed the listeners’ feelings or reactions to offer timely support if they encountered fatigue or confusion in the performances.

The speech material of a2 was used to determine the normative values and assess dialectal effects, while a1, a2, a3, a4, and aTP were used to evaluate the tonal effects on the older listener’s speech recognition. With regard to the learning effects, the children repeated a1 thrice and a2 twice. The dominant ear was tested. The test was performed in the following pattern:

[a2–a1–(five-minute break)–a1–a1–a2]

As mentioned in section 2, the speech threshold of young normal-hearing adults was expected to be close to  $25 \pm 5$  dB (SPL), and  $-16 \pm 3$  dB (SNR) with a small inter-subject variance.

#### 5.1.4 Data analyses

All data were extracted separately from Bells to Excel files and stored in CSV formats. The data were analyzed with a statistical software (R version 3.2.1). The analyses showed descriptive statistics across studies. To compare SRTs across the groups (age, dialect) or speech materials, a one-way analysis of variance (ANOVA) was conducted to measure the main effect. To follow up the interactive effect, a post-hoc test Tukey was conducted. The statistical significance was set at p-values of 0.05.

The descriptive statistics of SRTs across all hearing conditions were computed and illustrated through box plots, which showed values of mean, minimum, maximum, and median, the 25th percentile, and the 75th percentile for each data set. The dependent variable were the SRTs values in noise and quiet. The independent variables were the dialect groups (Northern, Southern, and Central), test trials, speech stimuli (a1, a2, a3, a4, aTP) and age groups. Correlations between aging and the SRTs across age groups were also calculated. The confusion matrix among speech stimuli and speech material was also analyzed.

#### 5.1.5 Exclusion of outliers

Some data points on AAST were considered to be outliers due to measurement errors, or misunderstood tasks. For example, the listeners were sometimes confused when the first speech

stimuli were presented. This happened particularly with the younger children or the older listeners. A few listeners responded to acoustic cues in their strategy. For example, they guessed the speech stimuli as the acoustic cues declined, leading to the lowest (best) speech threshold. Additionally, interferences of background noise (traffic noise, conversation) caused more mistakes by the listeners. Hence, data points were distant from the central variable of the data set.

Data analyses were based on Tukey's method (Tukey, 1977) to identify whether a data point was an outlier. An outlier was defined by its allocation, which was outside the fences (whiskers) of box plots. The following tables showed that some outliers (ears/audiograms) were eliminated from the analyses.

**Table 7:** Number of outliers (ears) in the normative data of AAST-a2

Groups		N (listeners, ears)	Outliers	
			quiet	noise
Children	4:00 to 4:11	(24, 48)	- 5	- 5
	6:00 to 6:11	(29, 58)	- 9	- 3
	8:00 to 8:11	(21, 42)	- 6	- 3
Youth and adults	15y to 20y	(22, 44)	- 2	- 1
	21y to 30y	(24, 48)	- 6	- 2
	31y to 40y	(20, 40)	- 2	- 4
Older adults	55y to 65y	(20, 36)	- 10	- 1
	66y to 75y	(20, 37)	- 2	- 0
	76y to 85y	(20, 31)	- 3	- 2

**Table 8:** Number of outliers in dialect datasets

Groups		N (listeners, ears)	Outliers	
			quiet	noise
Northern	Children	(18, 36)	- 4	- 4
	Adult	(21, 42)	- 2	- 3
	Older adult	(18, 34)	- 6	- 5
Central	Children	(19, 38)	- 2	- 1
	Adults	(21, 42)	- 7	- 3
	Older adults	(18, 35)	- 6	- 6
Southern	Children	(27, 54)	- 3	- 2
	Adults	(24, 48)	- 6	- 1
	Older Adults	(17, 34)	- 4	- 5

**Table 9:** Numbers of outliers across the five subtests

Stimuli	N (listeners, ears)	Outliers	
		quiet	noise
a1	(21, 41)	- 4	- 5
a2	(19, 38)	- 5	- 3
a3	(23, 46)	- 7	- 16
a4	(23, 46)	- 3	- 5
aTP	(22, 44)	- 6	- 8

In all, 435 listeners participated in the studies with 838 audiograms per hearing condition (quiet and noise). Out of the 838, the data from 110 ears/audiograms in quiet and 85 ears/audiograms had to be eliminated from the analyses, leaving 720 audiograms in quiet and 753 audiograms in noise for further analysis.

We have provided a detailed description of the methods used to collect data for the AAST. The next part of this section will give an overview of the methods used in the NAMES test.

## 5.2 Speech test of NAMES

The NAMES test was designed to find out and validate normative values in PRSs by native listeners of SVN. Besides, the speech test was also used to determine whether the dialect significantly affected the PRS of the non-native listeners of SVN, and whether the pitch height of tone (F0) affected the older listener's speech recognition. The following detailed description is about the participants, the speech stimuli, and the test procedures across the studies of the NAMES test.

### 5.2.1 Speech stimuli

Forty meaningless disyllables, divided into the two sublists A11 and A22, comprising 20 items each (see Appendix B), were used in the studies. As mentioned earlier, the phonemic distributions in A11 and A22 were homogeneous. The phoneme distributions of the two sublists were analogous to their counterpart in the spoken language of SVN. This made sure that the speech material was a duplicate of their daily spoken language. The syllables in  $C_1V_1-C_2V_2C_3$  structures carried two tones out of the three (A1, B1, and A2). Phonetically, homonymy was non-existent across the speech stimuli within a sublist.

With nonsense syllables, linguistic properties (semantics, or syntax, or word frequency) can probably be removed from speech materials (without effects from top-down processing). However, listeners

would respond to those by meaningful syllables instead (Bosman & Smoorenburg, 1995), according to their lexical size. In contrast, by the use of meaningful syllables, they might have guessed speech stimuli correctly even if they did not pay full attention to those. In addition, the meaningless speech materials are known to be independent of an individual's literacy or education (Cooke et al., 2010). Furthermore, this speech material also works for non-native listeners who have less experience of the language being tested (Paglialonga et al., 2014). Another advantage of meaningless speech materials is the independence of the cognitive ability of listeners (Akeroyd, 2008), which is considered to be the cause of deterioration in speech recognition.

The NAMES test functions through the Bells software installed in laptops. Laptops connect to the following devices: dr.dac nano-soundcard, Sennheiser HDA 280 headphones, and microphone Samson UB1.

### 5.2.2 Listeners

In all, 173 normal-hearing listeners participated in the NAMES tests. Five of the 173 participants were excluded from the study due to poor performance (due to hearing loss). They were among the older listeners above 75.

**Native listeners:** In all, 127 participants aged between 15 and 85 years took part in the assessment of norm values. Information on the listeners has been given earlier. However, the number of listeners in certain age groups changed for the NAMES task. These were: 31 to 40 years (n=19, female=10), 55 to 65 years (n=22, female=12), 66 to 75 years (n=24, female=12), and 76 to 85 years (n=19, female=8). The number of listeners in the remaining groups remained unchanged.

Regarding the study of the response modes, 18 adult listeners aged between 25 and 40 years (female=10, mean age=32.5 years), who studied at the universities of Bonn and Cologne, were recruited for the study. They spoke Vietnamese in their daily life. Half of them spoke NVN, and the other half spoke SVN.

**Non-native listeners:** The listeners were aged between 20 and 30 years. Of them, 19 spoke the Northern dialect (female=10, mean age=25 years), 19 spoke the Central dialect (female=10, mean age=25.6), and 21 spoke the Southern dialect (female=12, mean age=24 years). None of the listeners exhibited any problem of articulation. Due to the difficulty of recruitment of older listeners, the study of the dialectal effect in NAMES was conducted only on the group of adult listeners.

### 5.2.3 Test procedures

The listeners were tested individually in an alternative order of A11 and A22. The test time lasted roughly five minutes for each listener. Before the measurement started, the listeners did not familiarize themselves with the speech stimuli or the response tasks. All of them were asked to orally iterate the 40 nonsense disyllabic structures in two lists of 20 each. Before the measurement, the listeners received an instruction on the response task:

“You are going to hear 20 words that present individuals’ names. Please listen to these carefully and repeat them orally at once. If you do not hear the word properly, please ask the examiner for a reiteration.”

The speech stimuli were presented binaurally via Sennheiser HDA 280 headphones at fixed intensity level of 80 dB SPL in quiet. This is regarded to be the most comfortable intensity level for those with normal hearing.

Only the authors of this thesis carried out the examinations. The examiner gave the PRSs by assessing the correct and incorrect responses from the listeners through a graphical user interface. The tester observed the listener’s lip movements during the test. To recheck the test performances, the listeners’ voices were picked up separately by a microphone (Samson UB1) and automatically stored in a laptop’s hard disk. Furthermore, a digital device (TASCAM DR-05) recorded the background noise and the listeners’ prompts during the test performances. This recording device was placed roughly 30cm in front of the listener’s mouth.

To study the response method, the listeners iterated the speech stimuli by the following approaches: oral and written. Avoiding the learning effect might influence the phoneme scores, these two methods were interleaved during the test performances.

### 5.2.4 Data analysis

Like in the case of AAST, the data of NAMES were extracted from Bells to Excel and saved in the CSV format. All data were then analyzed using R version 3.2.1.

The overall PRS across studies were averaged from individual scores based on age groups, dialect groups, tone, and response modes. To measure the main effect, the ANOVA was carried out, and the statistical significance was set at  $p=0.05$ . Additionally, a post-hoc analysis with Tukey’s HSD was also done to calculate the p-value to compare the pairs. The proportions of phoneme

misidentification were also calculated in the percentage of errors based on individual phonemes, or phoneme classes.

To estimate the correlations between SRTs (AAST-a2) and duo-tone thresholds, and between duo-tone thresholds and PRS scores (NAMES), a Pearson's correlation coefficient was used.

This section has described the methods used in the present studies to find out the normative values for the two kinds of speech materials (AAST and NAMES), the effects of dialects and tones on speech audiometry testing. The next section will present the findings of the present work.

## 6. RESULTS

### 6.1 Results of the speech test AAST

#### 6.1.1 Normative values

The normative values were analyzed from the normative data of 200 native listeners of SVN. To determine the normative values of SRT, the listeners' SRTs were averaged for an individual age group. Simultaneously, the mean SRTs among the age groups were compared to observe the differences in the SRTs. The general results are illustrated in Figures 17 (thresholds in quiet) and 18 (thresholds in noise). A detailed description of the statistical analyses can be seen in Appendix C.

#### The SRTs in quiet

Figure 17 gives an overview of the SRTs among the different age groups for the native listeners in a quiet condition. Some generalizations can be seen. First, the maximum differences in speech recognition values appear in the four-year-olds and those above 55 years. Large inter-individual variances of SRTs are observed for the two groups as compared with the remaining groups. Second, both the youth and the adults have stable average SRT values. Third, eight-year-olds obtained nearly the same average speech threshold as those obtained by the youth or adults. The detailed results are depicted below.

#### - The groups of children

The results show a significant difference among the groups of children (A, B, C) with the main effect,  $F(2,129) = 37.7, p < 0.001$ . In particular, the mean speech threshold achieved by the four-year-olds (37.2 dB SPL, SD=5.1) were significantly poorer than those obtained by the six-year-olds (31.8 dB SPL, SD=2.5) and the eight-year-olds (31 dB SPL, SD=2.54). Additionally, a Tukey's HSD post-hoc test revealed a significant difference ( $p < 0.001$ ) in the SRTs between the four-year-olds and the six-year-olds, and between the four-year-olds and the eight-year-olds. However, there was a near statistical significance ( $p = 0.057$ ) in SRT for the six-year-olds and eight-year-olds. A slope of about -1.6 dB SPL/year was computed as the correlation between threshold and age.

#### - The groups of youth and adults

As mentioned earlier, the mean SRTs of the younger adults (D) and the two adult groups (E, F) were stable with a slight difference ranging between 0.5 dB SPL and 1 dB SPL. Specifically, the mean SRT of the 21–30-year-old group (29.4 dB SPL, 3.4) was slightly better than that of the 15–20-year-old group (30 dB SPL, 2.5), and the 31–40-year-old group (30.5 dB SPL, 3.0). A one-way ANOVA revealed no statistical significance among these groups, and the main effect,  $F(2, 122) = 3.32, p = 0.33$ .

A post-hoc test for further analyses indicated that the significant difference appeared only between the groups of 21–30 years and 31–40 years, with  $p=0.03$ . No statistical significance was found between the groups of 15–20 years and 21–30 years. In addition, the correlation between the SRTs and the subject's age in these three groups were estimated by a slope value of 0.07 dB per year.

#### **- The groups of older listeners**

Eighty listeners aged between 55 and 85 years participated in the study. Due to moderate hearing loss or cognitive deficits, data from 20 older listeners out of 80 were excluded from the analysis. Hence, only data from 60 older listeners could be used. Nearly half of the listeners had a problem with high-frequency hearing loss based on their duo-tone threshold of 4 kHz.

Figure 17 presents the mean SRT obtained by the group of 55–65 years (35.8 dB SPL, 3.3), which was significantly better than that obtained by the group of 66–75 years (40.4 dB SPL, 3.2), and the group of 76–85 years (40.6 dB SPL, 3.8). Apparently, there is a roughly 4.5 dB SPL difference between the two oldest groups and the younger one. The oldest listeners (76–85 years) seemed to have a wider range of speech threshold values as compared with the younger-older adult groups. The ANOVA revealed that the threshold differences among these groups were statistically significant, with main effect  $F(2, 86) = 17.3$ ,  $p < 0.001$ . Interestingly, the mean SRT for the 66–75-year group was exactly equal to that achieved by the 76–85-year group. Furthermore, the correlation of audibility threshold and the participant's age was expressed by a slope value of 0.24 dB per year.

#### **The SRTs in noise**

The SRTs in noise across the eight age groups were plotted. The features of SRTs in noise were relatively similar to those in quiet, with an explicit trend of increment in SRTs from the four-year-olds to the younger adults, and deterioration in SRTs across the three groups of older adults. The younger children (four years) and the older listeners (above 65 years) had poor mean speech threshold values between -9.5 and -6.5 dB SNR. The average SRTs of the children and adults were relatively equal—roughly -14.5 dB SNR. The mean SRTs of the eight-year-olds were somewhat equal to those achieved by the adults and the younger adults.

#### **- Group of children**

The results (figure 18) showed that the mean SRTs obtained by the younger children was significantly poorer (roughly a difference of 3 dB SNR) than those of the six- and eight-year-olds. A one-way ANOVA indicated that the thresholds of these groups were significantly different, with

main effect  $F(2, 134)=30.17, p<0.001$ . In addition, a close correlation existed between the speech recognition threshold and the listener's age with a slope value of roughly -1 dB per year.

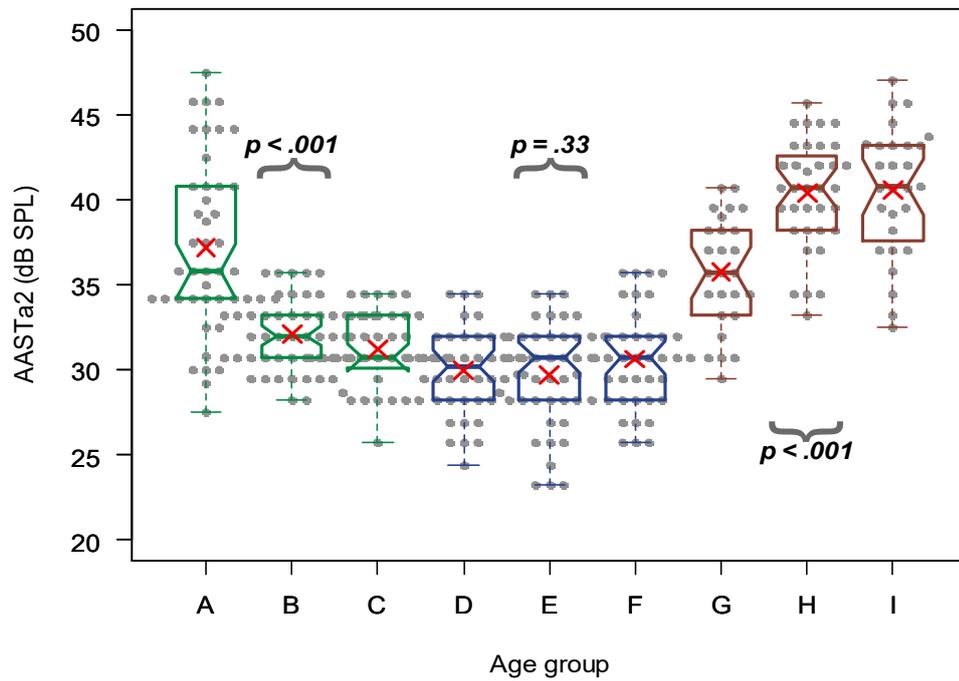
As stated earlier, the six- and eight-year-olds obtained a high SRT, which was comparable to that of the youths and adults. During the measurement, these two groups showed their interests and could run the test themselves without the support. However, the younger children encountered a lot of troubles, such as a lack of concentration. Even when they showed an interest in the task, their concentration remained for a short time.

#### **- Groups of youth and adults**

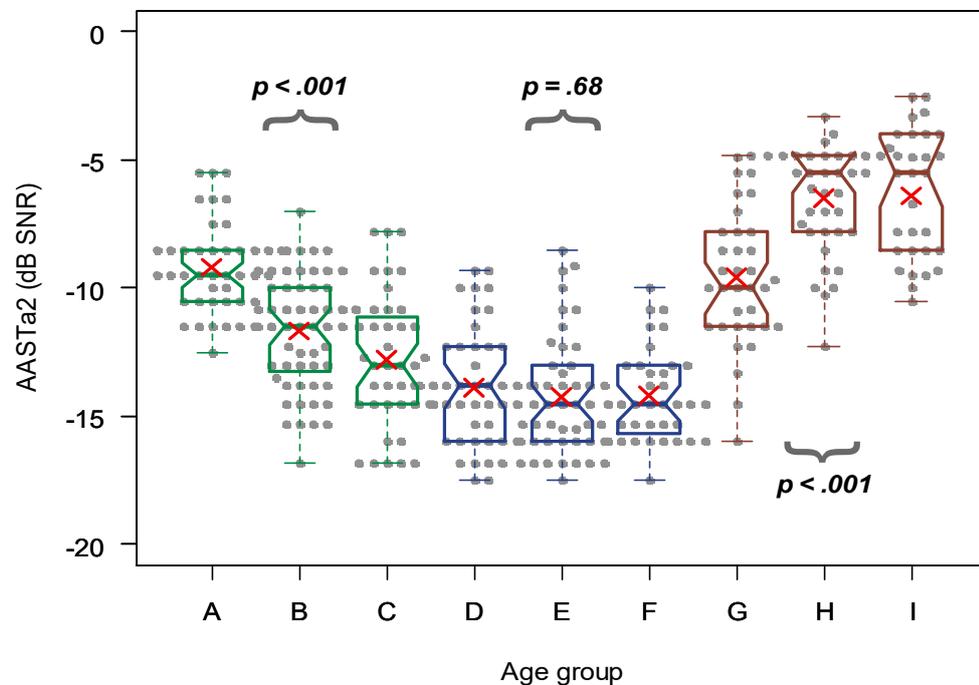
The speech threshold values of the young adults and adults were relatively equal with the average SRTs around -14 dB (SNR). To investigate the significance of the effect of age, a one-way ANOVA was carried out on the speech threshold. The analysis showed no significant difference among these groups,  $F(2, 127)=0.64, p=0.53$ . Since the speech threshold values of these were exactly equal, there was a weak correlation between the listeners' age and their speech thresholds.

#### **- Groups of older adults**

The mean SRTs of the 55–65-year group was around 3 dB (SNR) better than those obtained by the remaining older adult groups. A one-way ANOVA revealed that the difference in mean threshold among groups was statistically significant, with the main effect of  $F(2, 98)=19.3, p<0.001$ . Similarly, in quiet, the mean SRT obtained by the 66–75-year-olds and 76–85-year-olds were exactly equal. However, there was a larger range of threshold values for the 76–85-year-olds compared with those for the 66–75-year-olds. A post-hoc test revealed no significant difference ( $p=98$ ). The correlation between age and threshold for the groups of older adults was computed at 0.2 dB per year.



**Figure 17:** Speech recognition threshold values in quiet condition across the different age groups of the southern listeners (N=200), including the following groups: four years (A), six years (B), eight years (C), 15–20 years (D), 21–30 years (E), 31–40 years (F), 55–65 years (G), 66–75 years (H), and 76–85 years (I). The mean SRTs are in mark x



**Figure 18:** Speech recognition threshold values in a noisy condition across the different age groups of the Southern listeners, (N=200), including the following groups: four years (A), six years (B), eight years (C), 15–20 years (D), 21–30 years (E), 31–40 years (F), 55–65 years (G), 66–75 years (H), and 76–85 years (I). The mean SRTs are in mark x

In short, some conclusions can be drawn on speech recognition by SVN listeners. (1) The mean SRT values for younger children (four years) are 8 dB SPL (in quiet), -5 dB SNR (in noise) higher (worse) than those for the youths and the adults. (2) The two groups of older listeners (above 66 years) showed deteriorations in speech recognition (11 dB SPL and -8 dB SNR higher) as compared with those achieved by the adults. (3) The mean SRT of eight-year-olds was comparable to that of the adults. These results suggest a strong association between the listeners' age and the speech threshold values in quiet and noisy conditions.

### **6.1.2 Effects of dialects on speech audiometry testing**

This section will present the effects of dialect on speech recognition by determining the SRT values for each listener group: NVN, CVN, and SVN. Simultaneously, the average speech thresholds for each group will be compared to assess how divergent the speech threshold values are for the native and the non-native listeners. In addition, the reaction time (measured in ms) for each listener group will be calculated to compare the native and non-native listeners.

The general results (Figures 19-20 and Appendix D) show that the non-native listeners scored mean SRTs values significantly poorer than those of the native listeners in both quiet and noisy conditions. A much bigger range of speech threshold values was observed for the non-native listener groups compared with the native listener groups. Furthermore, the groups of non-native listeners took a longer reaction time in their performance in contrast to the native listeners who spent shorter response time (Figure 21).

#### ***Speech recognition thresholds in quiet***

##### **Group of children**

Children who were native listeners of SVN (S1) performed better than the two other groups of non-native listeners (N1, C1). The descriptive statistics show a noticeable difference in SRTs among the children's groups, in which the mean SRT of N1 was higher (2 dB SPL) than those of S1. The mean SRT of C1 was also higher (6 dB SPL) than that of S1. The one-way ANOVA revealed that these differences were statistically significant,  $F(2, 119) = 37.9, p < 0.001$ . In addition, standard error values (see the table in Appendix D1) for the groups of the non-native listeners were more variable than those of the native listeners.

### **Group of adults**

Likewise, the mean SRTs (see figure 19) of the native listeners (S2) were significantly better than those of the two non-native listeners (N2, C2). A roughly 3.5 dB SPL difference in audibility threshold showed a statistical significance between the non-native group and the native groups,  $F(2, 117) = 14.5$ ,  $p < 0.001$ . Moreover, the standard error values of the non-native speakers seemed to be different from those of the native listeners.

### **Group of older adults**

A few non-native listeners sometimes encountered challenges at the commencement of their performance, when the initial speech stimulus was presented. This was probably due to a novel accent or a lack of attention. Consequently, the test performances of several non-native listeners (N3, C3) seemed to be slightly affected. For example, the test took longer for each of them. In terms of the speech threshold, both N3 and C3 scored poorer SRTs than the native listeners S3. The difference ranged between 2.0 and 3.5 dB SPL. A one-way ANOVA revealed a statistically significant difference among these groups, with  $F(2, 83) = 4.8$ ,  $p = 0.01$ . The SRTs of non-native listeners were somewhat equal with no statistical significance ( $p = 0.21$ ).

### **Speech recognition thresholds in noise**

The results are displayed in figure 20 and Appendix D2. Similar to the SRTs in quiet, the threshold values in noise obtained by the non-native listeners were significantly worse compared with those obtained by the native listeners. The results under the effect of dialect will be mentioned in greater detail below.

### **Group of children**

Children who were native listeners (S1) significantly outperformed the non-native listener children (N1, C1). The degradation in speech thresholds by the non-native listeners ranged from 4 to 6 dB (SNR) as compared with the native listeners. There was a significant difference between the three groups under the effect.

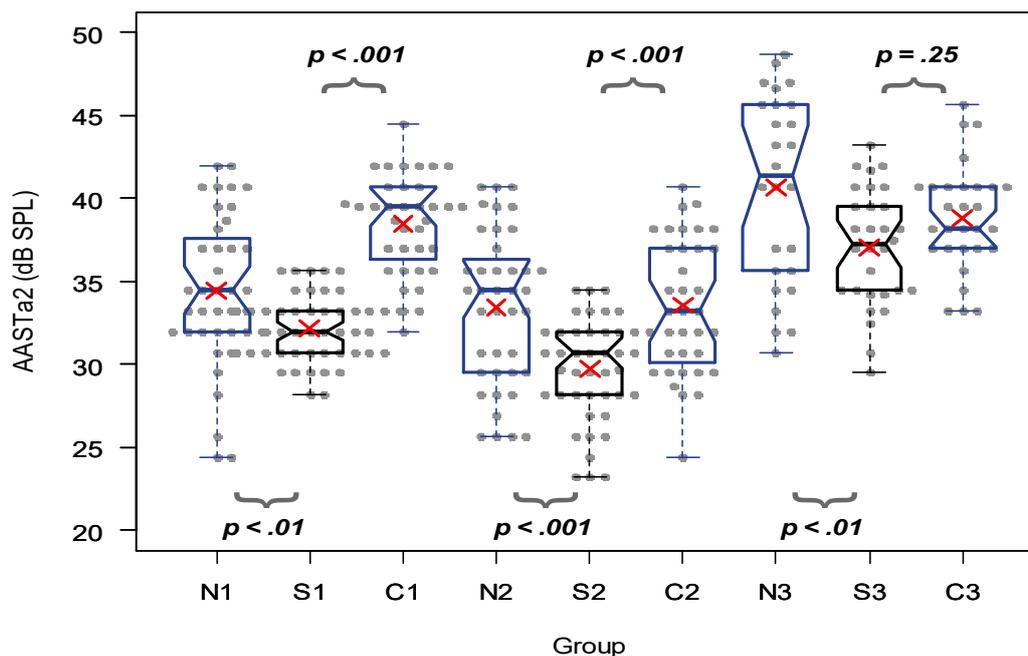
### **Group of adults**

A third of the non-native listeners (N2, C2) had minor experiences of the Southern speech due to the media, or their educational or work experiences. Hence, the dialectal variation was expected to slightly affect the speech thresholds of the non-native listeners. Unexpectedly, the results showed a huge difference in SRTs between the non-native and native listeners of SVN.

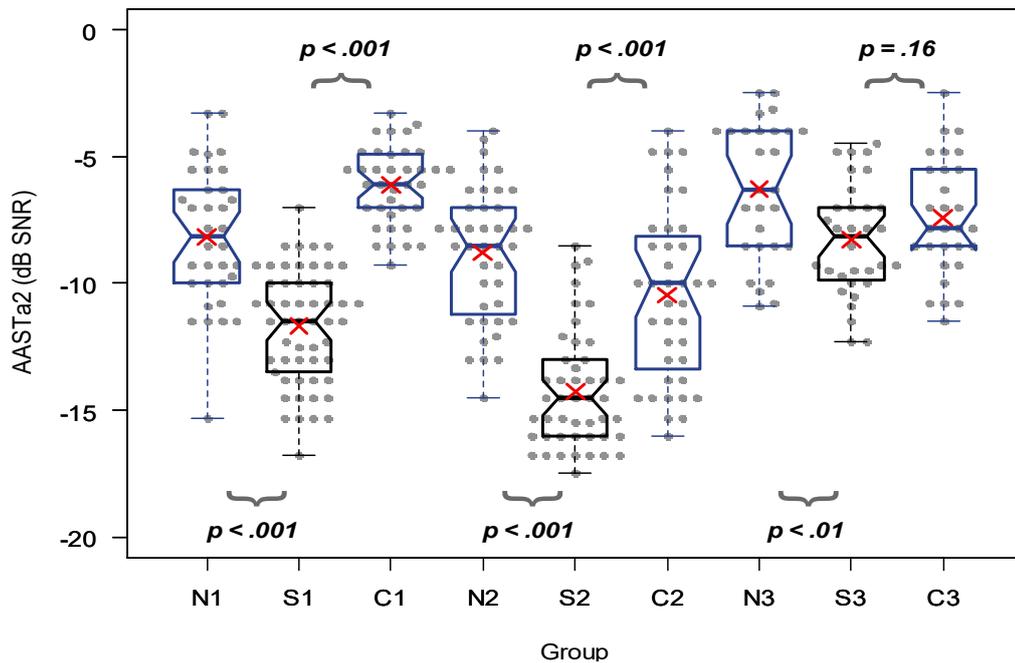
The mean SRTs of non-native listeners were 3.5–5 dB higher than those of the native listeners. This difference was found to be statistically significant, with a mean effect for dialectal examination,  $F(2, 123)=40.8, p<0.001$ .

### Group of older adults

Similar to the adult groups, the non-native listeners (N3, C3) could exchange information without any hindrance with people who came from the south. Hence, it was hoped that the SRTs would be relatively similar among the three groups. However, the results showed a significant difference among them in terms of speech-threshold values. The Southern participants' threshold values (S3) were better than the Northern (2 dB SNR) and Central listeners' (1dB SNR). Although this difference was minor, further analysis showed high significance with  $F(2, 86)=5.1, p=0.01$ . The mean SRT of the C3 was close to those of the S3 (1 dB difference). Post-hoc analysis revealed no significant difference between S3 and C3,  $p=0.16$ .



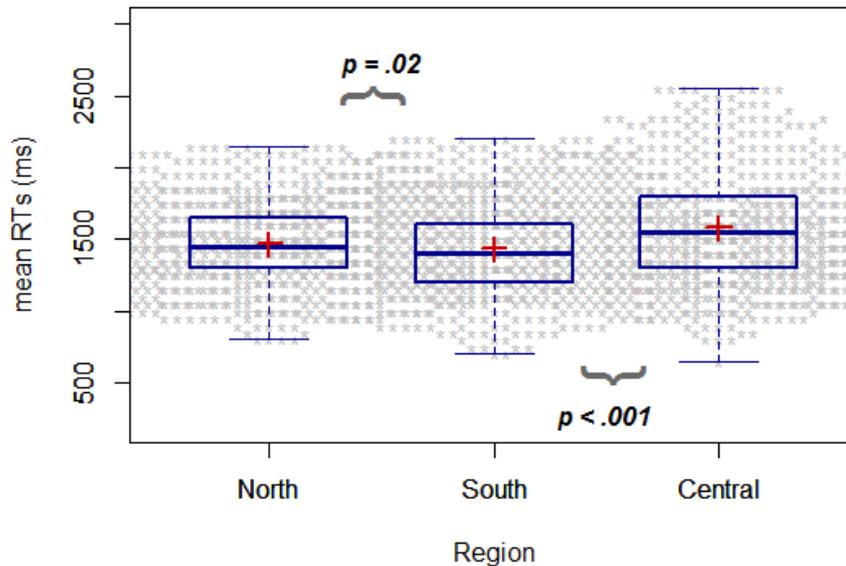
**Figure 19:** The speech recognition threshold values presented for the non-native listeners and the native listeners in quiet, including groups of Northern (N1, N2, N3), Central (C1, C2, C3) and Southern listeners (S1, S2, S3). The numbers coming after letters (N, C, S) indicate age groups (1: children, 2: adults, 3: older listeners)



**Figure 20:** The speech recognition threshold values in noise illustrated for the non-native and native listeners, including the three groups of NVN (N1, N2, N3), the three groups of CVN (C1, C2, C3), and the three groups of SVN (S1, S2, S3). The numbers following the letters N, C, S indicate age groups (1: children, 2: adult, 3: older adults)

### Reaction time (RT)

The reaction time is widely used in the area of human speech processing as a quantification of processing difficulty. In this study, the reaction times were computed when a stimulus began and lasted until the listener made a choice by clicking on one of the pictures displayed on the laptop screen. Figure 21 shows the reaction time for each dialect group (RTs of children were excluded from this study). As expected, the duration of the reaction time of the native listeners were significantly higher (between 35 ms and 146 ms) than those of the two groups of non-native listeners. The Tukey post-hoc test was carried out on values of RTs. The result shows a significant difference in terms of the RT between the Northern and the Southern listeners ( $p=0.02$ ), and between the Central and the Southern ones ( $p<0.001$ ).



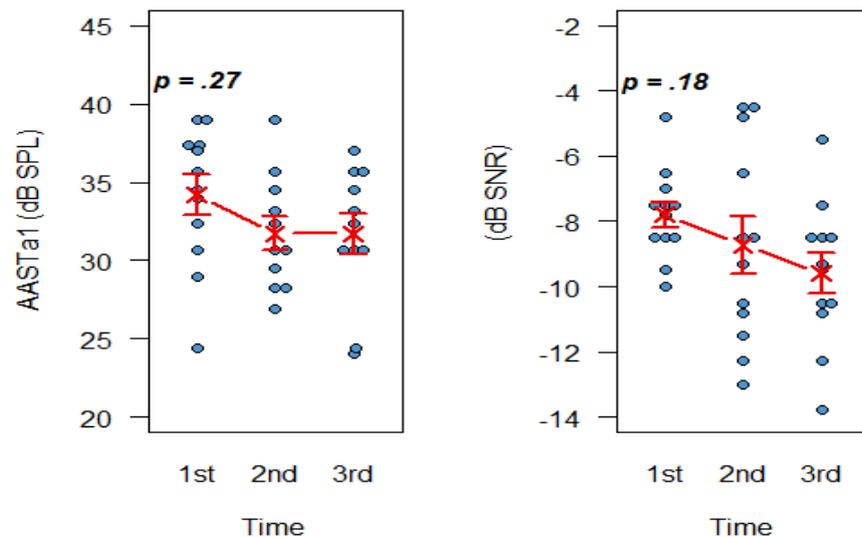
**Figure 21:** The reaction time of the three dialectal groups in the correct responses;  $n=968$ ,  $1239$ , and  $916$ , which are the numbers of correct responses for the Northern, the Southern, and the Central speakers respectively

In short, the mean SRTs and the RT significantly differ between the non-native and the native listeners of SVN. The above results show that the participants performed worse when the speech stimuli were not in their own dialect. Differences in the speech thresholds ranged from 2 dB to 6 dB between the non-native and the native listeners in quiet, and 1 to 5 dB in noisy condition. The results also showed a delay in the processing of words when the listeners heard speech stimuli presented in an unfamiliar dialect. The research findings suggested that the degradation in audibility threshold and the postponement in RTs of the non-native listeners are most likely because of the dialectal variation.

### 6.1.3 Learning effects on speech recognition thresholds

To examine the influences of learning effects on AAST, the speech threshold was computed each time the hearing test was done, and then compared. Figure 22 shows the SRTs under the learning effect for a group of 12 children (six-year-old). It shows the means, standard errors, and ranges under quiet and noise. The result shows that better performances were achieved in the second or third trial. In a quiet condition (left panel), the mean SRT for the first trial was significantly poorer (roughly 2.5 dB SPL) than the mean SRTs for the second and third trials. The mean SRTs remained unchanged in the second and third trials. There was a 2.5 dB (SPL) difference between the test and the retests in a quiet condition. However, the one-way ANOVA revealed no significant difference.

This was probably because a small number of listeners participated in this research. Likewise, the mean SRT in noise (right panel) improved gradually to a better speech threshold in the final trial. The increment in SRTs ranged from 1.0 to 1.8 dB SNR from the first to the last trials. This difference was negligible. The statistics found no statistical significance under learning effect ( $p=0.27$  for quiet,  $p=0.18$  for noise).



**Figure 22:** The speech threshold values in quiet (left) and noise (right) under learning effects on six-year-old children

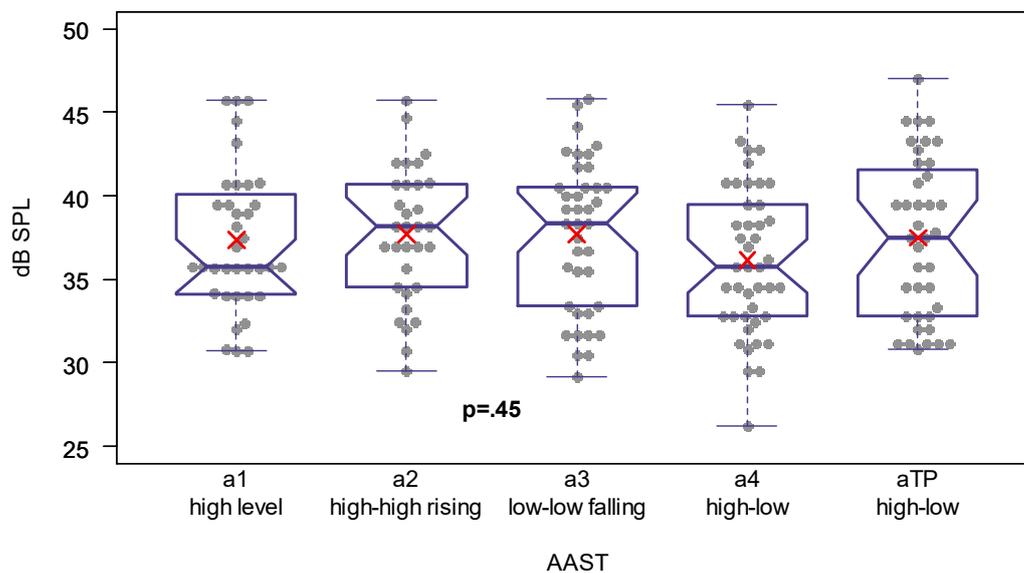
Although no statistically significant differences were found, the result might suggest that the improvements in SRTs (ranges 1 to 2.5 dB) would significantly affect the clinical findings to a certain degree.

#### 6.1.4 Effects of F0 on the speech recognition thresholds

To assess the effects of tonal patterns (F0) of syllables on an older listener's SRT, the mean SRTs were separately calculated for each subtest of AAST and then compared across the subtests (AAST-a1, a2, a3, a4, and aTP). Figures 23–24 and Appendix E show an overview of the SRT values for older adults. The values are plotted across two conditions, SRTs in quiet (figure 23) and SRTs in noise (figure 24). In general, the figures illustrate two noticeable results. First, aside from the mean SRT of a3, the mean SRTs of the remaining speech materials were relatively equal in both testing conditions. Second, the SRT in a4 seemed to be somewhat better than those in the other subtests under both quiet and noisy conditions.

### The SRTs in quiet condition

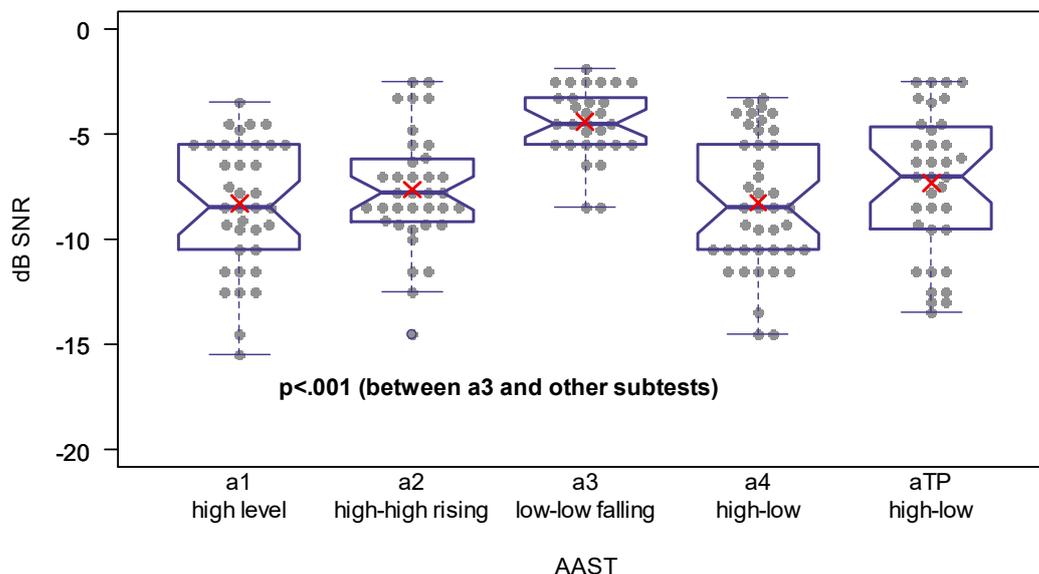
As mentioned earlier, the mean SRTs for the three subtests, aTP, a1, a2, and a3, were quite stable with a negligible difference of less than 0.5 dB SPL. For the subtest A4, the mean SRT of a4 appears somewhat better (1.5 dB SPL). However, no significant difference was found in the one-way ANOVA,  $F(4, 184) = 0.93, p = 0.45$ . Again, posthoc analysis (Tukey's HSD) was conducted to determine whether there was a difference between the pairs of speech test. The result also indicated no significance across the subtests of AAST, with the  $p$ -value ranging from 0.47 to 0.99.



**Figure 23:** Speech recognition threshold values in quiet condition across the speech materials of AAST for older listeners ( $N=86$ ), a1 and a2 bearing high-level and high-rising tones, respectively; a3 bearing low and low-falling tones; a4 and aTP bearing a high and a low tone respectively

### The SRTs in noise condition

In contrast to the SRTs obtained in quiet, those obtained in noise were more variable. In particular, while the mean SRTs in a1, a2, a4, and aTP varied minimally from one another (less than 1.0 dB SNR), the mean SRT in a3 was significantly higher (4 dB SNR). There was a main effect of the speech material  $F(4, 175) = 9.63, p < 0.001$ . To determine whether there were any significant differences between pairs of speech tests, post-hoc analyses (Tukey's HSD) were conducted. The analyses revealed a significant difference between a3 and each of the other subtests ( $p < 0.001$ ). There was no statistically significant difference between the pairs of a1, a2, a4, and aTP ( $p > 0.5$ ).



**Figure 24:** Speech recognition threshold values in noise condition across the speech materials of AAST, a1 and a2 bearing high-level and high-rising tones respectively; a 3 bearing low and low-falling tones; a4 and aTP bearing a high and a low tone respectively

Based on the results of SRTs, it can be summarized that there are no noticeable effects of tonal patterns on older adult's speech perception in a quiet condition. The average SRTs of the speech materials that bear only high pitch levels (a1, a2) are more or less equal to the average SRTs of those which carry a low pitch (a3), and a low-high pitch (a4 and aTP). Interestingly, in a noisy condition, the values for speech materials carrying low pitch levels and falling contours (for a3) differed significantly from the values of other subtests or speech materials (a-TP, a1, a2, and a3). These results suggest that the fundamental frequency of the tones (F0) do not influence speech recognition in a quiet condition. However, it may affect the SRTs if noise intervenes, especially when the two-syllable noun phrases used as stimuli in the speech tests have low and falling tones.

### 6.1.5 Response matrix and error analyses on AAST-a2

To examine whether each speech stimulus in AAST-a2 is balanced in word recognition, the proportion of word misidentification was analyzed. Table 10 and Figure 25 give an error matrix for six speech stimuli. In table 10, rows represent participants' responses and the columns represent the six speech stimuli. As it can be seen from the table and the figure, the stimuli *viên thuốc* and *sóng lớn* were confused more than the other speech stimuli, corresponding to 7.5% and 7.7% wrong choice respectively. The word *sóng lớn* was misperceived as *pháo bông* 24/977 times, and as *mắt kính* 23/977

times, while *viên thuốc* was misperceived as *túi xách* 22/1024 times. In addition, the two words, *sóng lớn* and *viên thuốc*, were responded with a question mark “?” by the majority. The participants were either unsure about the correct choice or did not hear the stimulus at all.

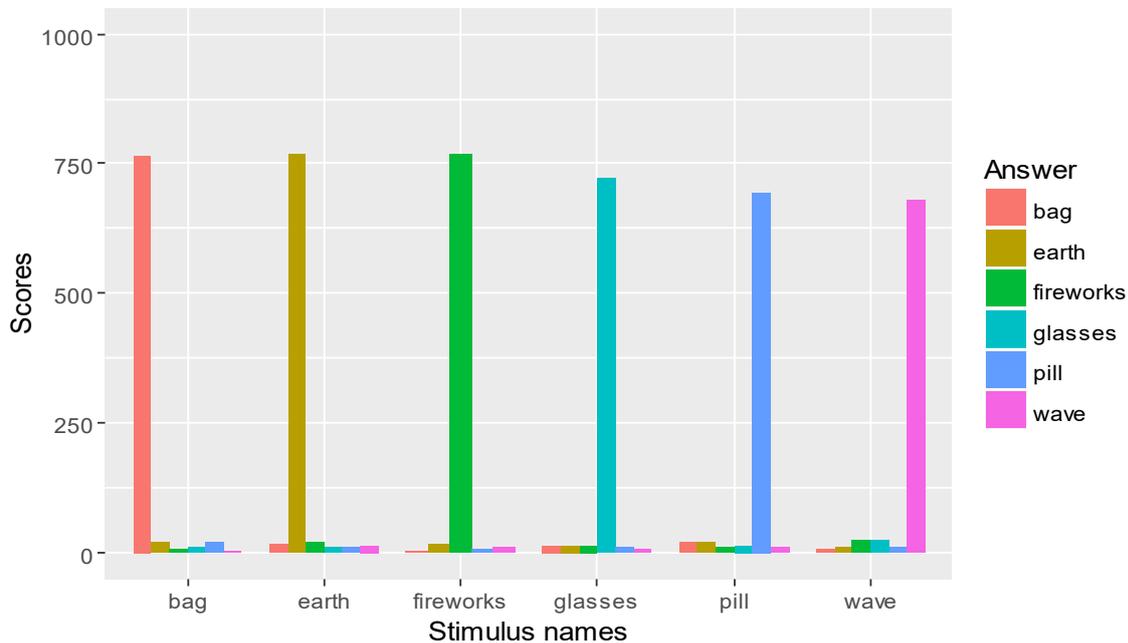
The most confusing disyllabic noun phrase for the youngest of children (four-year-olds) seemed to be *sóng lớn* (see Appendix F1). Along with *viên thuốc*, these two words also sounded confusing to the older listeners. These accounted for roughly 10 percent errors (see Appendix F3). In addition, most of the older listeners chose the “?” when these two speech stimuli were presented (roughly 40% responses). This suggested that the older listeners faced challenges in identifying fricative sounds, in particular /ʃ/ and /v/ in these words.

Taken together, the proportion of errors was minimal and slightly different among the six speech stimuli. This reflected a balance between the responses and the actual stimuli. Overall, all words were well identified.

**Table 10: Word confusion matrix for the groups of native listeners, N=6000 stimuli**

Answer Stimulus	A	B	C	D	E	F	?, t.o	?, to (%)	CoR (%)	WroC (%)
A	766	20	7	10	22	3	178	17.7	76.1	6.2
B	16	767	22	11	12	13	194	18.7	74.1	7.1
C	5	18	769	0	7	9	168	17.2	78.8	4.0
D	13	13	13	720	10	6	207	21.1	73.3	5.6
E	22	19	12	14	694	10	253	24.7	67.8	7.5
F	6	12	24	23	10	681	221	22.6	69.7	7.7

The six speech stimuli included *túi xách*, “bag” (A), *trái đất*, “earth” (B), *pháo bông*, “fireworks” (C), *mắt kính*, “glasses” (D), *viên thuốc*, “pill” (E), and *sóng lớn* “big wave” (F). Abbreviations: CoR (Correct response), WroC (Wrong choice)



**Figure 25:** The proportion of errors among the speech stimuli for the native listeners

Together, these results provide important insights. Firstly, the normative values in AAST-a2 show interdependence between the listener's age and the SRT. For instance, the young children performed significantly worse (roughly 8 dB SPL, 5 dB SNR) than the adults with respect to speech threshold. Similarly, the older listeners (over 65 years) obtained significantly worse speech thresholds (roughly 11 dB SPL and 8 dB SNR) than the adults. Secondly, a 1-2dB difference among the test trials revealed that the learning effect still existed but was negligible on the AAST in Vietnamese. Thirdly, as expected, the non-native listeners showed significant deficiencies in SRTs and a considerable delay in word processing when the speech stimuli were presented in an unfamiliar dialect. The SRTs of the non-native listeners were mostly larger than 2.5 dB (quiet and noise), and showed a delay of 40 to 150 ms in auditory processing compared with the native listener groups. Finally, it was assumed that the subtests of AAST carrying high-rising contours would have poor mean SRTs than those bearing low-falling and high-level tones. However, the findings showed no apparent evidence to prove that the tones with high pitch levels had an impact on the speech recognition of older listeners.

This section has presented the findings using the speech material AAST. The next part of this section will show the results using the speech material NAMES.

## 6.2 Results of the speech test NAMES

### 6.2.1 Normative values

#### Overall recognition phoneme scores

In all, 134 participants took part in this experiment. They were divided into six age groups, as mentioned before. The children were not included in the study. To determine the normative values for every age group, the PRS was averaged and compared across the age groups.

Table 11 shows an overview of the results of the statistical analyses for these age groups. In general, the mean PRS was quite stable and similar across the groups, except for a group of 76–85-year-olds. First, the 15- to 20-year-olds achieved slightly poorer PRS (1%) than the 21- to 30-year-olds and the 31- to 40-year-olds. ANOVA revealed a nearly-significant difference among these groups, with  $F(2, 137) = 2.64$ , and  $p = 0.07$ . Second, in the age groups between 55 and 85 years, descriptive statistics showed a descending order in the PRSs associated with the listener's age. The 55–65-year-olds scored better than the two others: around 1 percent correct phonemes compared with the 66–75-year-olds, and roughly 3 percent compared with the 76–85-year-olds. The ANOVA showed statistically significant differences among these groups,  $F(2, 124) = 8.76$ ,  $p < 0.001$ . For further analysis, a post-hoc test was also conducted to determine which pair was more significant than the other. The result revealed a significant difference between the 66–75 and the 76–85 groups,  $p < 0.001$ .

As mentioned earlier, the mean PRSs were stable and balanced across groups, barring the 76–85-year group that scored 94.5 percent correct phonemes, three percent lower than the 21–30-year group. So, why was there a slight dissimilarity in the average PRS among the age groups, barring the 76–85-year group? This question is easy to interpret. All the speech stimuli of NAMES were presented in a comfortable intensity level of 80 dB SPL under a quiet condition. Most of the older listeners—even those with mild hearing loss at 40 dB HL (duo-tone)—could identify nearly 98 percent phonemes correctly. Hence, the result suggests that the age factor does not affect NAMES's PRSs among listeners aged between 55 and 75 years. The aging factor appeared to affect more the PRSs of listeners aged over 75 years.

**Table 11:** Descriptive statistical values of the Southern listeners, N=268

Age group	N	Mean	SD	Median	Min	SE
15 to 20	44	96.7	1.98	97	93	0.30
21 to 30	50	97.5	1.69	98	93	0.24
31 to 40	46	97.4	2.15	98	92	0.32
55 to 65	48	96.8	1.95	97	91	0.28
66 to 75	46	96.0	2.21	96	89	0.37
76 to 85	34	94.5	4.0	95	87	0.69

In connection with the word perception scores, the participants aged below 40 years scored better—ranging between 13 and 20 words, with an average of 17.5. The participants aged more than 55 years scored slightly poorer, with 16 correct words for the 66–75-year-olds, and 15 correct words for the 76–85-year-olds, ranging between six and 20 correct words.

After looking at the results of statistical analysis, a brief look may be taken at the phoneme perception errors in the two age groups. The tables in Appendix H give an overview of the error rates for tonal and phoneme identifications. It suggests an absolute difference in the perception errors between adults and older adults. First, with regard to phoneme recognition, the fricative sound /h/ is the one most misidentified by the older listeners (35.7%), and to a lesser extent, the phoneme /m/ (8.9%). The most misidentified stimulus by adults was the vowel /ǎ/ (11.5%), and to a lesser extent, the phoneme /o/ (9.3%). Second, with regard to tonal identification, the older adults made fewer mistakes than the adults. The proportion of errors by the older adults was around 2.5% for tones A2, while the error rate for adults was roughly 3.7%. An additional analysis of error rate will be done in detail in a later part of error analysis.

#### **Average consonant scores based on phoneme categories**

The score values can be averaged and compared based on the phoneme categories. The list of 11 consonant phonemes employed in the current research was grouped into five groups: voiced stops /b, d/, unvoiced stops /t, k, tʰ/, lateral-nasal /m, ŋ, l/, fricative /j-, h/, and semi-vowel /-j/. The five included 11 voiced-stops, 20 unvoiced-stops, 15 lateral-nasals, 9 fricatives, and 5 semi-vowels in each list of 20 speech stimuli. As a function of phoneme classes, the results showed a significant difference in the listeners' responses for both groups: the young and adult listeners  $F(4, 695)=8.91$ ,  $p<0.001$ , and the older listeners  $F(4, 630)=86.7$ ,  $p<0.0001$ . In particular, the average scores measured for the young and the adult listeners were relatively high, and close to a small range of

97.8% and 100% correct responses across phoneme classes. In contrast, the scores of the older listeners had a wider range—between 83.5% and 99% correct responses. Of these, the scores of the fricative sound were significantly lower compared with the other phoneme classes, corresponding to 83.5%. The remaining phoneme classes had exactly equal scores and were almost perfectly distinguished. The deterioration in the fricative's scores compared with the others implies that the decline in phoneme scores was associated with high-frequency hearing loss by the older listeners (Gelfand et al., 1986; Maniwa et al., 2008).

### **Confusion pattern of phoneme recognition**

To find out the phonemes that are more confusable, data from 135 native listeners were studied in detail. The following tables give an overview of the response matrix and the proportion of errors that rely on each phoneme position in the syllable structure ( $C_1V_1-C_2V_2C_3$ ), including the elements of onsets, nuclei, and codas, along with the three lexical tones of high level (A1), low falling (A2), and high rising (B1).

The results show that the spread in inaccuracy is considerable for the consonant phonemes with a wider range of proportion of errors between 0.4% and 18%. The error for vowel phonemes is small with a range of error between 0.8% and 7.9%. This means that the listeners identified the vowel more accurately than the consonant. This was in accordance with the previous studies (Meyer et al., 2010 and 2013; Cooke et al., 2010), which found that the accuracy of the identification of vowel phonemes was significantly higher than that of consonant phonemes.

In relation to onset identification, table 12 displays the response matrix and error rates for the initial phonemes, which were added up from positions  $C_1$  and  $C_2$ : 190/1080 tokens were misidentified (18%). The phonemes /m/ (37/810 times) and /k/ (45/1080 times), corresponding to 4.6% and 4.2% respectively, were misidentified to a lesser extent. The fricative onset /h/ was mistaken the most frequently. This could be attributed to hearing loss on fricative sounds among the older listeners. Instead of responding to the fricative /h/, the older adults tended to replace this with another fricative /χ/, or a plosive /t'/. In contrast, the youth and the adults did not show such misidentification pattern. An extra analysis of phoneme errors can be seen in Appendix H.

**Table 12:** The response matrix and proportion of errors for the onset by the native listeners

Response	Wrong	/k/	/d/	/t/	/l/	/v/	/m/	/h/	/b/	/t'/	Number	% error
/k/	45	1035	0	0	0	0	0	0	0	0	45/1080	4.2
/d/	52	0	1823	2	0	0	0	0	<b>12</b>	1	67/1890	3.5
/t/	11	3	0	795	0	1	0	0	0	0	15/810	1.9
/l/	6	0	0	0	1344	0	0	0	0	0	6/1350	0.4
/v/	21	0	0	0	0	1329	0	0	0	0	21/1350	1.6
/m/	30	1	0	0	0	0	773	0	6	0	37/810	4.6
/h/	<b>179</b>	1	0	0	0	0	0	886	0	14	194/1080	18.0
/b/	10	0	2	0	0	0	2	0	1066	0	14/1080	1.3
/t'/	16	0	0	0	0	0	0	1	0	1333	17/1350	1.3

With respect to vowel identification, table 13 shows the response matrix and the proportion of errors made by the native listeners. The response matrix was aggregated from the vowel phonemes V<sub>1</sub> and V<sub>2</sub>. The raw result shows that two nuclei /o/ and /ǔ/ are more confusable than the others. The phoneme /ǔ/ has the largest proportion of errors of 107/1350 times (7.9 %), and to a lesser extent, the phoneme /o/, with 105/1620 times (6.5%). These kinds of misidentifications resulted from the acoustic similarity between two rounded vowels /o/ and /ɔ/, and between short vowels /ǔ/ and /ǎ/, or /ǔ/ and /u/. In the same vein, the native listeners perceptually confused the front vowel /i/ for the front vowel /e/. Due to the resemblance of acoustic properties, the listeners could not distinguish well between these vowels. This was especially noticed in the participants aged between 15 and 40. The older listeners distinguished between these relatively better.

**Table 13:** The response matrix and proportion of errors for vowel phonemes

Response	Wrong	/a/	/o/	/ɔ/	/i/	/ǔ/	/ǎ/	Number	error
/a/	24	2946	0	0	0	0	0	24/2970	0.8
/o/	10	0	1515	88	0	7	0	105/1620	6.5
/ɔ/	29	1	5	2115	0	9	1	45/2160	2.1
/i/	37	0	0	0	1853	0	0	37/1890	2.0
/ǔ/	44	1	0	0	0	1243	62	107/1350	7.9
/ǎ/	11	0	0	0	0	3	796	14/810	1.7

The result of tonal identification is presented in Table 14. The tone A2 (low pitch level and falling contour) was most commonly mistaken among the three lexical tones: 106/3500 tokens of A2 were misidentified as B2, corresponding to 3.1% of the incorrect responses (B2 was not included in this

experiment). This finding was in line with the findings of Vü (1981), and Brunelle and Jannedy (2013). The confusion between the tones B2 and A2 stemmed from the similar acoustic features—both are low-falling tones in the SVN. Based on the observation in this study, in the case of open syllables, for example, *là*, *bì*, *thì*, and *đì*, the listeners produced either the tone A2 or B1. A lesser extent of the error rates happened in tone B1, in which 72/3433 tokens were misidentified, corresponding to 2.2% of the error rates. A surprising difference in the perception of tone A2 was found between the adults and the older listeners. The latter made fewer mistakes than the younger adults (see details in Appendix B). A possible explanation for this was that the older listeners paid more attention to speech stimuli with low-falling tones.

**Table 14:** Error rates and response matrix for tonal identifications

Response	Wrong	B1	A2	A1	Number	% error
<b>B1</b>	72	3433	0	5	77/3510	2.2
<b>A2</b>	106	2	3401	1	109/3500	3.1
<b>A1</b>	30	0	2	3748	32/3780	0.8

For the coda identification (table 15), the nasal /-m/ was the most common error among the five codas, in which 35/540 tokens of /-m/ were replaced by another nasal /-n/. However, the coda /-n/ was not included in this study. So, the exact number of responses in which /-m/ was misidentified as /-n/ could not be calculated. To a lesser extent, 36/810 of the final stop /-t/ were identified by either /-k/ or /-t/. This was not surprising when the final stops /-t/ and /-k/ were not acoustically distinguished in the SVN. Therefore, the Southern listeners were perceptually more confused when they heard these sounds. Similarly, there was no distinction between the final nasal /-ŋ/ and /-n/. Thus, the phoneme /-ŋ/ also had a high proportion of errors: 41/1350 tokens of /-ŋ/ were misrecognized, corresponding to 3.0% of incorrect responses.

**Table 15:** Error rates and response matrix for coda recognitions

Response	Wrong	/-k/	/-ŋ/	/-j/	/-t/	/-m/	Number	% error
<b>/-k/</b>	7	1332	0	0	<b>11</b>	0	18/1350	1.3
<b>/-ŋ/</b>	41	0	1309	0	0	0	41/1350	3.0
<b>/-j/</b>	7	0	0	1343	0	0	7/1350	0.5
<b>/-t/</b>	5	<b>31</b>	0	0	774	0	36/810	4.4
<b>/-m/</b>	34	0	1	0	0	505	35/540	6.5

In short, the fricative /h/ (onset) and the nasal /-m/ (coda) were commonly misidentified among the consonants while the nuclei phonemes /o/ and /ə/ were frequently mistaken across the vowels. For tonal identification, tone A2 was the most misidentified of the three tones. The older listeners made different error patterns compared with the youths and the adults in terms of both phoneme and tone identification. The following part of this section will show the comparison between adult and older adults with respect to consonant scores by phoneme categories.

### 6.2.2 Effects of response modes

As it has been seen earlier, the PRSs in the verbal responses were relatively high—around 97 percent—for adults. To check whether there were any discrepancies between the two response modes being used in the NAMES performances, 18 listeners were recruited for the study. The results showed that the listeners who responded verbally to the NAMES test had a higher phoneme score (96%) than those who gave written answers (90%),  $t(72) = 7.16, p < 0.0001$ . Table 16 shows all descriptive statistical values for this study.

**Table 16:** Descriptive statistical values of PRS for effects of test responses

Manner	N	Mean	SD	Median	Min	SE
Oral	36	95.7	2.3	96	89	0.45
Written	36	90.3	3.7	90	83	0.61

The data included 10 native listeners and eight non-native listeners of SVN. Because of this, the average phoneme scores were not in line (1 percent lower) with normative values obtained from youths and adults. An interesting result of the word score must be reported. The number of correct words scored by the verbal response (articulation) was 16, and by the written response (phoneme-grapheme correspondences), only 12. These results revealed that the written response was more risky than the verbal response in the case of NAMES.

### 6.2.3 Effects of dialects on speech audiometry testing

#### Overall phoneme recognition scores

To investigate the significance of dialectal effects on phoneme scores of NAMES, two groups of students, who were non-native listeners, took part in this research. The percentage of correct phonemes of the non-native listener groups were averaged and compared with those achieved by the

native listeners. Table 17 shows the statistical analyses for the two groups along with the results of the native listeners.

Overall, the non-native listeners scored poorer (roughly 1.5%) than the native listeners on the PRS. The two non-native listener groups scored the same—about 96 percent correct phonemes. The ANOVA showed a statistically significant difference relevant to speech threshold among the dialect groups,  $F(2, 143)=12.53, p<0.001$  though the difference of 1.5% was negligible. The result suggested a weak effect of dialect on the listeners' phoneme scores.

**Table 17:** Descriptive statistical values for the three listener groups

Group	n	Mean	SD	Median	Min	SE
North	50	95.9	2.27	96	89	0.32
Central	42	95.8	2.28	96	90	0.35
South	54	97.6	1.73	98	93	0.24

With respect to the word scores, the number of correct words across the dialect groups was also calculated. Similar to the phoneme level, the native listeners scored significantly better—roughly 18 correct words. The two groups of non-native listeners scored slightly lower—around 16 correct words.

### **Confusion patterns of phoneme recognition**

A comparison of the phoneme errors between the non-native and native listeners revealed some interesting points. As expected, the non-native listeners scored greater error rates in identifying phonemes. The confusion patterns differed across the three dialect groups as well (see Appendix I). The following is a description of phoneme confusion patterns relevant to the dialectal effect.

With respect to onsets, the non-native listeners made more mistakes in identifying the palatal onset /j-/, which was roughly 10% incorrect responses. The native listeners made fewer mistakes in this onset—around 1.6%. When the non-native listeners heard speech stimuli with the phoneme /j-/ in SVN, they seemed to confused it with /v-/ or /z/ in their own dialect. They produced the /v-/, /z-/ and /j-/ interchangeably. In contrast, most of the native SVN listeners produced either /v-/ or /z-/ in standard Vietnamese for the palatal onset /j-/ (Vương & Hoàng, 1994). So, the native listeners did not make such errors. Instead, they made more errors in the fricative voiceless /h/ for the initial position.

With respect to nuclei, the native listeners had higher error rates in identifying the short vowel /ǎ/ (roughly 8.0%). The non-native listeners did not make such mistakes, but had larger error rates in identifying the phoneme /i/ instead. The Northern listeners made 11% errors while the Central listeners made 7% errors. The errors in identifying the phoneme /ǎ/ can be explained simply by the linguistic behaviors of the native SVN speakers. They produce /ǎ/ and /ǎ̃/ somewhat similarly in terms of acoustics when /ǎ/ is followed by the nasal sounds /n and ŋ/. As a result, some (62/1350) misidentified /ǎ/ as /ǎ̃/ in terms of perception. In the case of the front vowel /i/, the non-native speakers frequently perceived /i/ as /e/. This was probably due to the acoustic resemblance of the vowel /i/ borne by the falling tone A2 without an ending sound. In this context, this vowel was more confusable with /e/. This error was also found in the native speaker's response, but the proportion of error was minor.

With respect to codas, the non-native listeners were usually more confused between the two nasal sounds /-ŋ/ and /-n/ in their performance as compared with the native listeners: 19% incorrect responses for the Northern listeners, 27.6% for the Central listeners, and about 3.0% for the Southern listeners. As mentioned in the section on literature, the SVN speech does not distinguish well between the coda alveolar /-n/ and the coda velar /-ŋ/, which the NVN and the CVN do. Although an ending sound /-n/ was not included in the speech test, the non-native listeners would remind themselves that the ending sound /-ŋ/ in the SVN might be either /-n/ or /-ŋ/ in their response. This caused more confusion for them. The non-native listeners made a large number of error on the phoneme /ŋ/, while the native listeners made a mistake on the alveolar /-t/ and the bilabial /-m/. This kind of misidentification has been elucidated in the preceding part.

With respect to tone identifications, all three groups of listeners seemed to find it difficult to identify the low-falling tone A2, whereas they easily recognized A1 and B1. The non-native listeners made equal mistakes in recognizing A2 (roughly 6%) while the native listeners gave 3% incorrect responses. The reasons for tonal confusion will be explained in the coming part.

Lastly but surprisingly, both the native and non-native listeners showed the same phonemic confusion patterns, as they were less accurate in the identification of the nasal /m/ and replaced it with the nasal /n/ as a corresponding response.

### 6.2.4 Effects of F0 on supra-threshold phoneme recognition scores

Disyllabic structures were used, as the speech stimuli carried the three tones of *ngang* (A1), *sǎc* (B1), and *huyẻn* (A2). As interpreted earlier, the tones A1 and B1 represented high pitch level in contrast to A2, which represented low pitch level with a falling contour. In our experimental design, each tone was associated with a certain vowel (vowel-plus-tone). Because of this, the PRSs were aggregated from the identification scores of both vowels and tones. The PRSs have been counted, compared and illustrated in Table 18 for the groups of older listeners.

**Table 18:** Descriptive statistical values of PRS in tonal identification task

Stimuli	N	Mean	SD	Median	Min	SE
VA1	127	98.5	3.6	98	86	0.32
VA2	127	97.6	4.1	97	85	0.36
VB1	127	96.8	5.2	96	77	0.46

VA1: vowel and tone *ngang*, VA2: vowel and *huyẻn*, VB1: vowel and *sǎc*

As shown in Table 18, the older listeners obtained fewer correct responses when they listened to the stimuli carrying tone B1 (96.8%). They scored somewhat better on the stimuli VA2 (97.6%) and VA1 (98.5%) than on VB1. To examine whether this difference was significant, a one-way ANOVA was carried out. The result showed that the effect of tone on phoneme scores was significant,  $F(2, 376)=4.9$ ,  $p<0.005$ . Post-hoc analyses (Tukey's HSD) revealed a significant difference between VA1 and VB1 with  $p<0.005$ . There was no significant difference between VA1 and VA2, and between VA2 and VB1.

A detailed look at the perception errors shows that for the stimulus VA1, 27/1778 tokens were misidentified, whereas for VB1 and VA2, respectively 53/1651 and 40/1651 tokens were misidentified. These misidentifications arose from incorrect choices in either tones or vowels. For the stimuli VA1, the older listeners could not discriminate reliably between two short vowels /ǎ/ and /ẻ/. The tone A1 was well recognized with small error rates of 0.8%. For the stimuli VA2, the listeners predominantly misperceived the tone A2. In the adult groups, the older listeners frequently misidentified the tone A2 as tone B2 (tone *nẻng*, “creaky falling”), corresponding to 2.5% of the incorrect responses. This was probably due to the very similar acoustic features of the two tones (Vu, 1984; Brunelle & Jannedy, 2013). With regard to the stimuli VB1, the listeners made mistakes predominantly on vowels rather than on tone. Particularly, they got confused between /o/ and /ẻ/, and /ǎ/ and /ẻ/. Their tonal errors amounted to 2.5%.

More importantly, a comparison between the groups showed relevant results concerning the age of participants and their vowel-plus-tone identification. The 55–66-year and 66–75-year groups scored relatively higher on phonemes (almost 97.5%), whereas the 76–85-year group scored slightly lower, especially in recognizing stimuli carrying the tone B1 (94.5%). This suggests that the degradation of tone-identification scores in the stimuli VB1 might stem from age-related hearing loss, particular for those above 75.

In summary, this part began by describing the norm values for the NAMES test. The results showed a slight discrepancy in PRSs among the age groups within the native listeners (except for phoneme scores obtained by the 76–85-year group). From these findings, the research suggested that the PRSs correlated significantly with the listeners above 75. With respect to the response modes, the phoneme scores were various (roughly 6%) between the verbal and written responses. The listeners obtained significantly higher phoneme scores in the verbal method than in the written method. About the effects of dialects, the native listeners scored somewhat better PRSs than the non-native listeners. Although the disparity was statistically significant, the distance was negligible (around 1.5%). With regard to the tonal effect, the older listeners achieved slightly poorer PRSs when the stimuli carried the high-rising tone B1. They scored better on identifying the remaining tones. However, these differences might be negligible. Due to a lack of persuasive evidence, it is still unclear whether the high pitch level of the tone influences the phoneme recognition of older listeners. Besides tonal identification, the older listeners were more confused when they had to identify the voiceless glottal fricative /h/, a high-frequency phoneme.

The following part of this section will mention the relationship between SRTs and PRSs and duo-tone thresholds.

### **6.3 Correlation between speech audiometry materials and duo-tone audiometry**

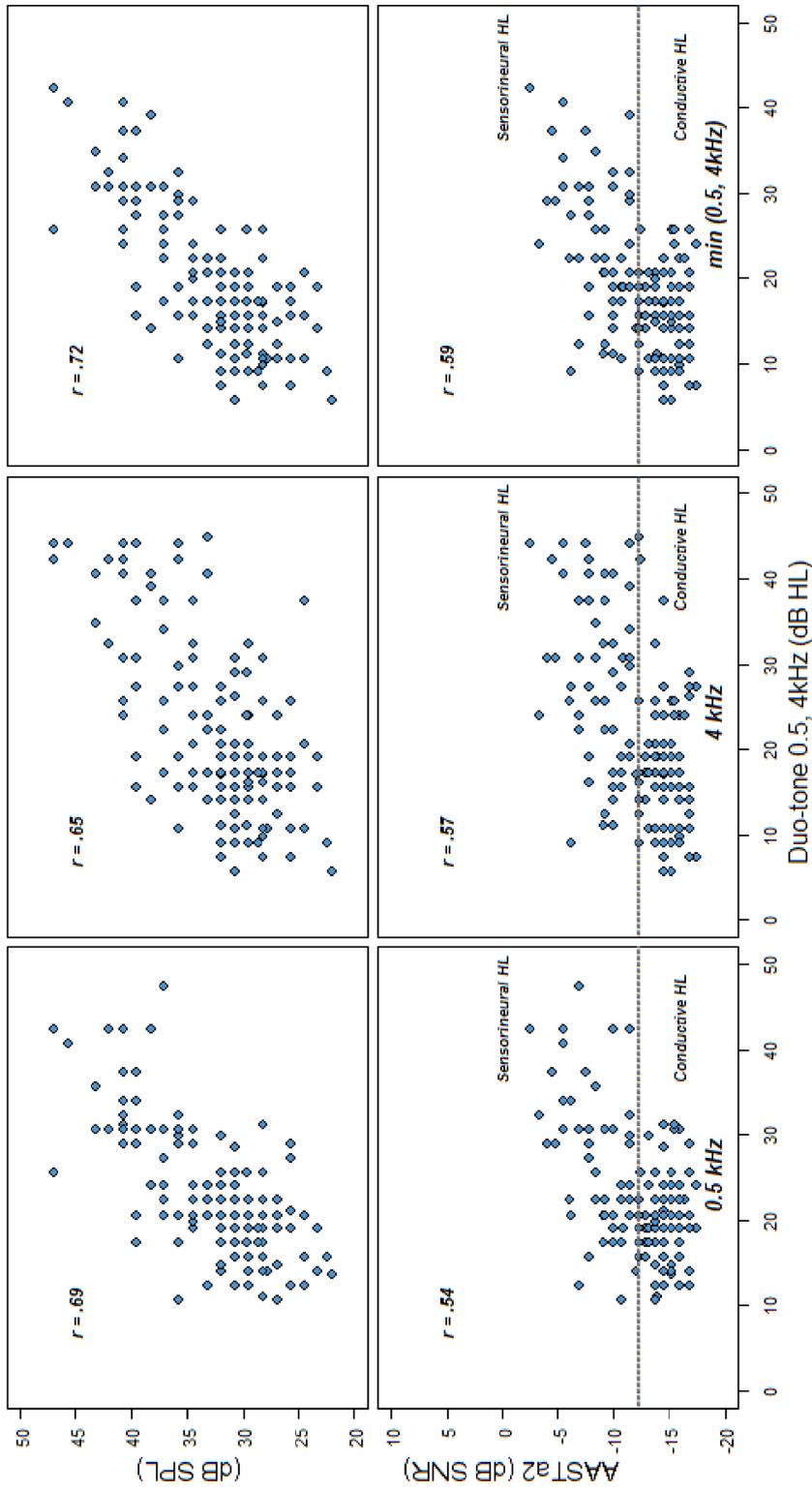
#### **6.3.1 The correlation between SRTs and duo-tone thresholds**

Scatterplots of SRTs versus duo-tone thresholds taken together is presented in figure 26. The two left panels in the figure illustrate the correlation between the SRTs and the duo-tone thresholds at 0.5 kHz. The middle ones describe the associations at 4 kHz, and the right ones show the relation between SRT versus the lowest (best threshold) duo-tone thresholds of 0.5 and 4 kHz. Due to ceiling effects, the maximum duo-tone thresholds of 4 kHz (mostly by older listeners) were excluded from this analysis. The distributions of individual SRT values and duo-tone thresholds are presented in Appendix J.

The figure shows some interesting generalizations. First, the SRTs are found to be significantly correlated with duo-tone thresholds. As the duo-tone threshold increases, there is a corresponding increase in the SRTs. Second, the SRTs of AAST are somewhat strongly associated with the duo-tone threshold in quiet ( $r$  values ranged from 65 to 72), and conversely, weakly correlated in noise ( $r$  valued from 0.55 to 0.59). Third, since the distribution of SRTs and duo-tone thresholds at 0.5 kHz was close to that of the SRTs and the lowest duo-tone thresholds (0.5, 4kHz) but differed from that of the SRTs and duo-tone threshold (4 kHz) in quiet, it can be stated that the listeners leaned towards the duo-tone of 0.5 kHz as the best frequency level for recognizing the speech stimuli of AAST in quiet.

Next, figure 26 also indicates that several older listeners had poorer duo-tone thresholds. But they could obtain better SRTs (especially in noise). In contrast, some others had good duo-tone thresholds but achieved worse speech thresholds. Apparently, there were two different kinds of hearing losses in noise for the native listeners: sensorineural and conductive hearing loss. These types of hearing losses can be found in detail in the correlation between the speech thresholds in quiet and the speech thresholds in noise (see figure 26 and Appendix J3).

Lastly, the duo-tone threshold at 0.5 kHz can predict SRTs better than the duo-tone threshold of 4 kHz in quiet. Correlation coefficients of duo-tone thresholds and speech thresholds in quiet ranged from 0.33 (4 kHz) to 0.50 (0.5 kHz) for native listeners. In contrast, the correlation coefficients were low in noise, ranging from 0.2 (4 kHz) to 0.27 (0.5 kHz).

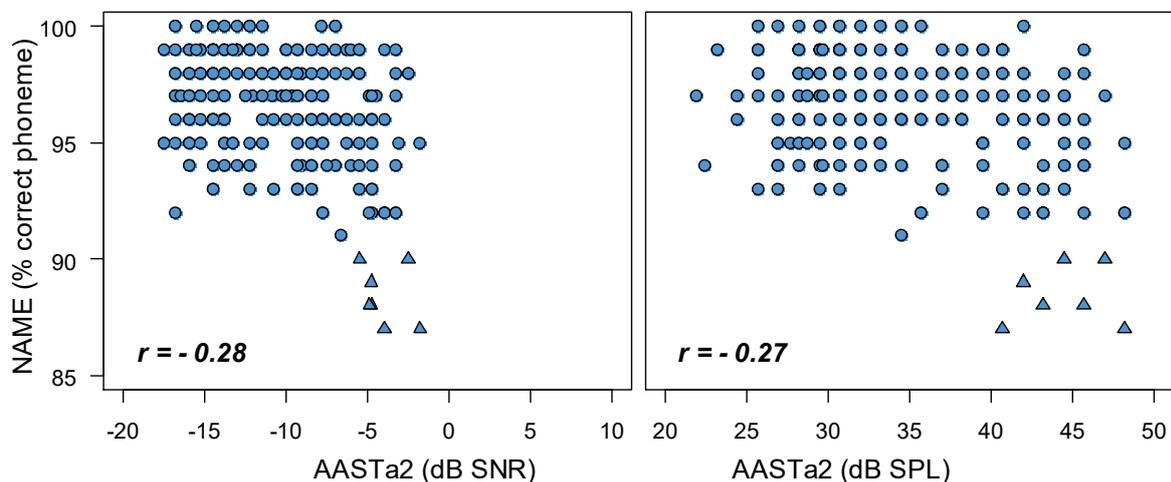


**Figure 26:** Correlations between SRT's and duo-tone thresholds (N=159). The participants included the groups of youths, adults, and older adults

### 6.3.2 Correlation between SRTs (AAST-a2) and PRSs (NAMES)

Figure 27 displays the relationship between the PRSs and the SRTs (quiet and noise). There are significant negative correlations between the two. As speech thresholds decrease, the phoneme scores increase. However, these correlations seem to be too weak. The  $r$ -values are identical ( $r=-0.28$ ). Statistical analyses showed statistical significance for this correlation, although weak ( $p<0.001$ ).

An interesting thing about the figure must be mentioned. Some participants, who were considered to have mild or moderate hearing loss (based on their SRT), achieved high phoneme scores in NAMES. This indicates that the NAMES test is a very easy task for even those with moderate hearing loss.



**Figure 27:** Correlations between the SRTs and PRSs,  $N=247$ . Points in the triangle are considered as outliers, which were left out of the analysis

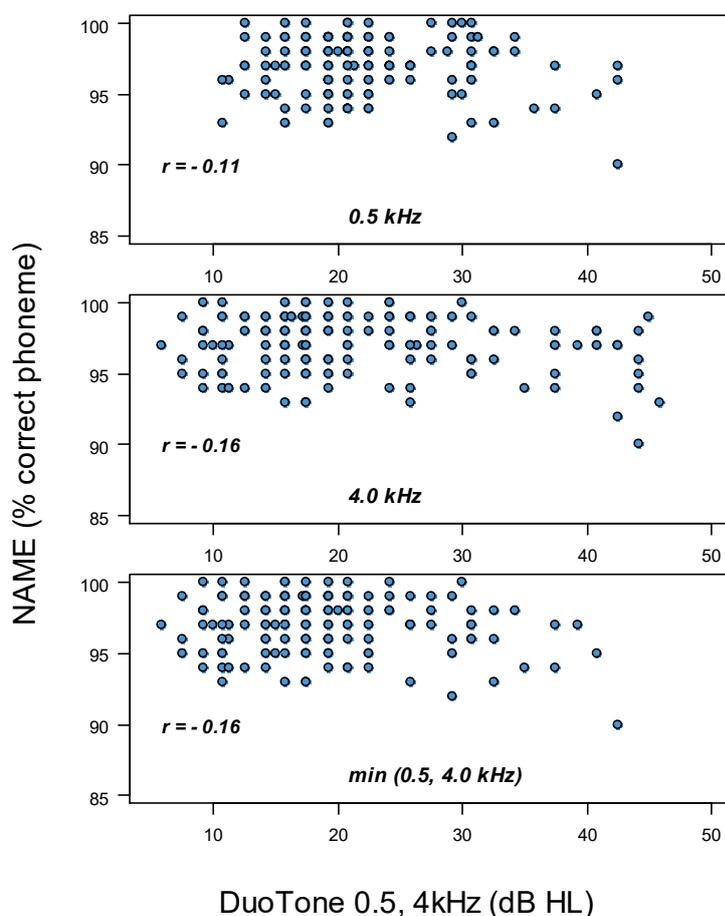
### 6.3.3 Correlation between PRS (NAMES) and duo-tone thresholds

Figure 28 shows the relationship between the PRS and the duo-tone threshold of two different frequencies. The top panel of the figure illustrates the correlation between the PRS and the duo-tone threshold 0.5 kHz. The middle one describes the relationship at duo-tone threshold 4.0 kHz, and the last one shows the relation between PRS and the best duo-tone threshold (lowest threshold) of 0.5 and 4 kHz.

Unlike the strong correlation between the SRTs and the duo-tone thresholds, the correlations between the PRS and the duo-tone thresholds are so weak (as seen in the figure) that it seems that there is no correlation. Pearson's correlation coefficients reveal  $r$  values ranging between -0.1 and -

0.2. The correlations between them can apparently be seen at duo-tone threshold larger than 30 dB HL. In such cases, the listeners could only reach phoneme correctness close to 95 percent.

The correlation between the PRSs and duo-tone thresholds showed similar results. Some participants, who were considered to have moderate or mild hearing loss based on their duo-tone thresholds, seemed to have normal hearing based on their PRS. Due to the simplicity of the NAMES speech stimuli, the PRS could not be predicted through the duo-tone threshold.



**Figure 28:** The correlations between PRS and the duo-tone threshold,  $N=87$ , corresponding to 159 observations

In summary, the above results show a strong relationship between SRTs (a2) and duo-tone thresholds, but very weak or weak associations between SRTs and PRSs, and between PRSs and duo-tone thresholds. Based on the findings, the research suggests that the PRSs in NAMES cannot

predict both the SRTs and duo-tone thresholds, whereas the SRTs of a2 can be used as an interpreter for duo-tone threshold or vice versa.

#### 6.4 Summary of the results

This section has provided the main findings with regard to the normative values, and their dialectal and tonal effects on speech recognition by using two different speech materials: AASTs and NAMES. In terms of normative values, the findings suggested dependence between aging and the speech threshold values in the speech material of AAST. The greater the listener's age the greater their SRT. In contrast, there is no explicit association between aging and PRS in NAMES. All the age groups, except the 76–85-year one, scored the highest on phonemes. This indicated that a possible age effect exists only in the oldest listeners aged between 76 and 85. In terms of dialectal effect, the result showed significant impairment in a non-native listener's SRTs. In contrast to AAST, in NAMES, the PRSs of the native and non-native listeners were close, with only a small variation (1.5%). These results indicated that dialect affects the hearing threshold in AASTs greatly but influences the supra-threshold PRSs in NAMES only slightly. With respect to the effects of the fundamental frequency of tones, SRT values among the subtests of AASTs (except a3 in noise) were relatively equal. In the case of a3 subtest, the SRTs were degraded due to the masking of noise. The tone-plus-vowel identification showed a slight reduction in PRS when the speech stimuli carried the high-rising tone B1. Comparing the two results, it can be said that the fundamental frequency of tones in the SVN might have a negligible influence on the older listeners' speech perception, especially those above 75 years. In relation to learning effects, a minor elevation was seen in the speech thresholds when the test trials were compared.

With respect to the association between the two speech audiometry materials and the duo-tone audiometry, there is a moderate relationship between SRTs (AAST-a2) and duo-tone thresholds in both quiet and noisy conditions. Particularly, the correlation in quiet is somewhat higher than that in noise. In contrast, there is either a weak or no correlation between SRTs and PRSs, and between PRSs and duo-tone thresholds. Lastly, but very interestingly, the observations revealed that both native and non-native listeners tended to shift their perceptions to some degree to the speech cues that were consistent with the speaker's accent. In contrast, the listeners ignored those speech cues that were inconsistent with their own dialect for both AAST and NAMES. These findings suggest that the listener either considered the speech cue as a benchmark for his/her responses or rejected it. Another interesting observation is that the listener sometimes responded to the speech stimuli by

replacing homonyms that occurred in their own language, especially, to the oldest listeners. This can be elucidated by two kinds of processing—top-down and bottom-up—in speech recognition. Comprehensions and perceptions of speech depend on top-down cognitive and bottom-up auditory processing (Goldstein, 2009) in a complementary approach. When incoming signal relating to bottom-up auditory processing is impoverished, top-down processing may compensate for it insofar as it is based on the linguistic context of a wanted message that enables the listener to decode the degraded incoming information (Craik, 2007).

## 7. DISCUSSION

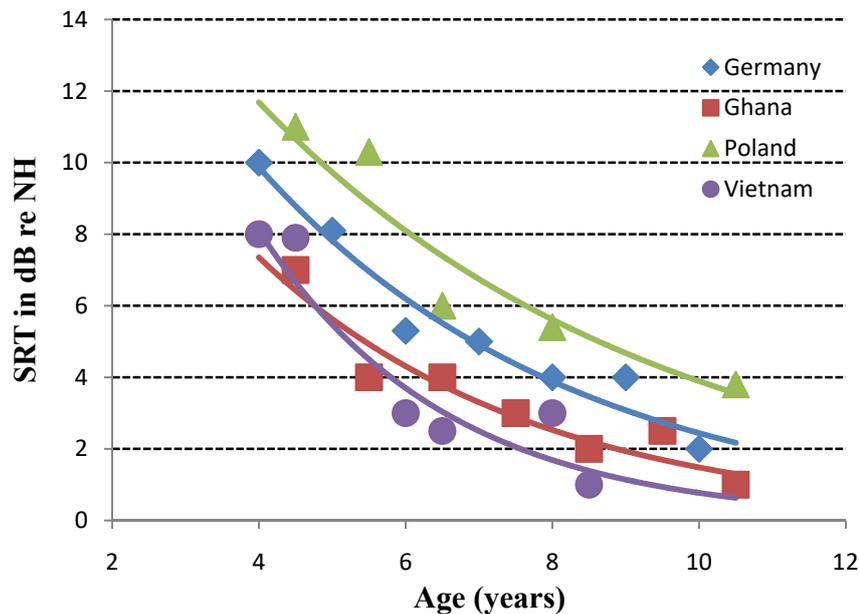
### 7.1 Normative speech threshold values in Vietnamese AAST

The first question in this study sought to determine whether there is an age-related norm value on Vietnamese AAST similar to other languages. As predicted, the results of this study indicate that the SRTs depend on the listener's age, especially in younger children and older adults. Therefore, the two following hypotheses can be accepted:

*H1a.* The norm values depend on the listeners' age, with children and older adults performing significantly poorer in speech threshold values than adults and

*H1b.* The age-related norm values in Vietnamese are analogous to those in other languages

Indeed, the four-year-old children obtained significantly poorer (6 dB) scores on speech threshold than the eight-year-olds. Compared with the adults, the four-year-olds scored considerably poorer (8 dB) speech threshold values. The older listeners (55–65 years) also performed progressively poorer on speech threshold (6 dB) than the adult listeners. The two remaining older listener groups scored poorly on the speech threshold with a difference of 11 dB with the adults. These findings are in line with the results of previous works conducted by Coninx (2005, 2008) in German, and by Offei (2013) in Ghanaian (see Figure 29 for detailed comparisons across languages regarding age-related normative values in AAST). Coninx measured SRTs for 417 German children aged between six and 12 years in NRW (Nordrhein-Westfalen). The results revealed significantly higher SRTs for the four-year-olds (10 dB) than for the 11-year-olds. Similarly, Offei conducted his research on 581 Ghanaian children aged between four and 10. He also found a difference of 7 dB between the younger and the older children. Nevertheless, the previous studies did not investigate the SRTs in AAST for older listeners. Thus, the extent to which their speech threshold differs from that of adults was still not clear. However, the present findings seem to be consistent with Divenyi's work (2005), which found that the speech threshold elevation of older listeners corresponded to 6.32 dB/decade due to aging.



**Figure 29:** Age-related norm values in AAST across languages: German, Ghanaian, Polish (adapted from Coninx, 2016) compared with those in Vietnamese

In the current result, the 66–75-year and 76–85-year groups scored poorly on SRT (5 dB) than the 55–65-year group. The deterioration in speech perception by the older adults might be ascribed to presbycusis or cognitive changes relating to age processing (Gelfand & Piper, 1987; Fortunato et al., 2016; Mukari et al., 2015; Pichora-Fuller, 2003 & 2008). For younger children, the poorest speech recognition stems from a lack of concentration or short attention spans during the test performance (Coninx, 2005).

By comparing the norm values of AAST in German, the average speech threshold in Vietnamese was still higher (2 dB SPL) than those in German and Ghanaian. This difference has also been attributed to the high background noise level in Vietnam in which level of background noise ranges from 38 to 48 dBA.

## 7.2 Learning effects on AAST

It was hypothesized that learning effects are (almost) non-existent in AAST. This research showed negligible differences in SRT values of AAST-a2 across the tests and retests in noisy and quiet conditions. The SRT improved by 2.5 dB SPL (quiet) and by 1 dB SNR (in noise) from the first trial to the second. The difference between the first trial and the third was 2 dB SNR (in noise). However, there was no difference in quiet. These differences had no statistical significance either. The slight improvement in the listener's SRTs observed after the first test trial may have been due to adaptation

to the speech stimuli or the test procedures, or an improved attention level, or all of these. Hence, hypothesis H2 can partially be accepted:

*Learning effects are non-existent in the Vietnamese AAST due to similar speech threshold results across the three test trials.*

This finding is comparable to that of a study conducted by Mohammed (2007), which was slightly different from Offei's work (2013). Mohammed investigated the learning effect on Arabic AAST on 10 children (5–7 years) who spoke Arabic. When the test was performed thrice with the children, he found minor differences in SRTs in both listening conditions. Somewhat differently, Offei examined the learning effect on Ghanaian AAST for two groups of Ghanaian listeners: children and adults. He found that sometimes the learning effects existed and sometimes they did not. Offei concluded that the learning effect was minor and clinically acceptable.

A significant difference in SRT that sometimes occurred in Offei's research disappeared in Mohammed's work and the current study. This could be due to the sample sizes of listeners tested across the studies. Offei's work included hundreds of listeners, including two groups of children and a group of young adults. The current study tested only 11 children while Mohammed's study tested 10 children.

The current study's findings are comparable to those of the aforementioned works. This means the learning effect, in general, appeared between the first and second measurements, but not between the second and third ones. At most, there was a minimal clinically difference. This finding implied that it is better if children are allowed to run a training test to familiarize the speech stimuli or test procedure before starting actual measurements.

### **7.3 Tonal effects on perception of speech in AAST and NAMES**

The result raises the question whether the fundamental frequency (F0) of syllables affects SRTs and PRSs of the older listeners. The SRT and PRSs were measured by using AAST and NAMES for 86 listeners aged between 55 and 85. The speech materials were designed using tonal contrasts in height and contour: high, high-rising, low, and low-falling. The main findings were as follows: Firstly, no significant differences were found among the SRTs across AASTs (except a3's SRT in noise condition). Secondly, a minor difference was found for the PRSs among three lexical tones. In particular, the listeners were less accurate in identifying tones A2 and B1 but more precise in identifying tone A1. Thirdly, the listeners performed significantly worse in SRTs when they listened

to the speech stimuli that carried low or low-falling tones and when these stimuli were presented with a masking noise. In a nutshell, the results were unexpected and suggested that the fundamental frequency of lexical tones might have a small effect on the speech recognition ability of older adults. The low-falling and high-rising tones appeared perceptually difficult, especially for listeners above 75. Hence, the hypothesis H3 can be accepted partially:

*Tonal patterns (F0) in terms of lexical tones in Vietnamese have an effect on the speech recognition of older native listeners with high-frequency hearing loss.*

The results correspond to those of previous research on tonal languages, e.g., Vietnamese (Vũ, 1981; Brunelle, 2009; Kirby, 2009; Brunelle & Jannedy, 2013), Cantonese (Varley & So, 1995), Mandarin (Yang et al., 2015; Liu & Samuel, 2004), and Thai (Kasisopa et al., 2015), which found that tones with low pitch levels or falling contours were more confusing for older listeners, whereas the tones with a high pitch level were identified more accurately. However, there was a contrast when it came to the rising tone. The studies by Yang et al. (2015), and Liu and Samuel (2004) on Mandarin tones found that the high-rising tone (Tone 2) had the highest mean score and was less confused than the low-dipping (Tone 3), and high-falling (Tone 4) tones. In contrast, the high-rising tone (B1) in Vietnamese had somewhat lower mean score as compared with the two remaining tones (A1, A2). This contrast may be due to the listeners' age. The previous results examined the tone identification of the older listeners aged between 50 and 70, while this research tested those aged between 55 and 85. As a function of age, a greater deterioration in an individual's speech recognition could be expected among listeners aged above 75 (Willot, 1991; Studebaker et al., 1997) than those aged between 65 and 75 (Sherbecoe & Studebaker, 2003).

The confusions between tones A2 and B2 have been examined intensively in previous studies (Vũ, 1981; Brunelle, 2009; Kirby, 2009; Brunelle & Jannedy, 2013). All the authors cited the same reason for such confusion: probably due to acoustic similarities between the two tones (starting relatively low and falling smoothly until the endpoints in their F0 contours). As a result, the speech stimuli with tone A2 were often perceptually misidentified as tone B2 in the stimuli of the NAMES test. It has to be noted that such confusion occurred not only for older listeners but also for the younger ones. Similarly, in the speech material of a3 (the stimuli carried tones A2, B2, and C–dipping), these tones caused considerable deterioration in speech thresholds of older listeners, especially in noise. As it is known, the information on phoneme and tone is important to distinguish and recognize the lexical meaning of words in tonal languages. When the speech stimuli were presented with a masking

noise, the speech signals and the maskers overlapped regarding spectrum and temporality (Stickney et al., 2004). As a result of the combination of acoustic similarities, overlapping of the maskers, and speech signals, the listeners' speech recognition for a3 stimuli showed a severe decline with a difference of 4dB (SNR) as compared with other stimuli.

As it has been reported above, the primary features of Vietnamese lexical tones lie in fundamental frequency (F0), their harmonics, and duration. According to Pham (2003), and Brunelle (2003), the F0 and the harmonics in Vietnamese tones are located in the low-frequency range, between 50 and 500 Hz, regardless of dialects. This is likely to be the reason for the older listeners to have a good ability in the low-frequency area, which was sufficient for them to recognize the F0 values from the speech tests. The findings of the current study do not support the hypothesis that the older listeners who are considered to have high-frequency hearing loss have poorer speech recognition in high-rising tones than in low-falling tones, regardless of the speech materials. The degradation in phoneme scores in relation to high-rising tones by the older listeners (75–85 years) might have been caused by hearing impairment or cognitive ability reduction relating to their age processing, but was not apparently linked to high-frequency hearing loss.

This study examined only the tone identification for normal-hearing older listeners. Future work may focus on other populations, in particular, older listeners with mild, moderate, and severe hearing loss, to ascertain whether there are any differences between these groups and the group of normal-hearing listeners.

#### **7.4 Normative phoneme scores in Vietnamese NAMES**

The fourth objective of this study was to validate the use of nonsense consonant-vowel-consonant-vowel-consonant ( $C_1V_1C_2V_2C_3$ ) syllables to measure a listener's phoneme identification ability. The research also sought to determine whether there was a dependency in terms of age and norm values similar to other languages. The result of the current study shows that there was a slight deterioration relevant to the listener's age in the NAMES test. Particularly, the phoneme scores of the listeners under 65 years were somewhat equal—roughly 97% correct responses. The phoneme scores of the remaining groups, especially the 76–85-year group, were slightly poor—around 94.5% correct responses. A comparison of the standard deviations in phoneme scores across the age groups show that inter-subject variability was larger for listeners above 75 than those under 75. This indicates that the older listeners (76–85) seem to produce a wider range of PRSs than the younger ones.

Contrary to expectation, this study did not find any statistically significant differences among listeners under 75. The significant difference was found only for the 76–85-year group as compared with other groups. Hence, hypothesis Ha4 can be partially accepted:

*The normative values depend on the listener's age, with older adults performing significantly poorer regarding phoneme scores than adults.*

Hypothesis H4b can be completely accepted:

*The age-related norm values for NAMES in Vietnamese are similar to those in other languages.*

The results of the current study are consistent with those of Studebaker et al. (1997) and Billings et al. (2015). These studies have used University Auditory Test Number 6 (NU-6) as speech stimuli. Studebaker et al. (1997) investigated the association between speech performances and chronological age. They conducted the study on 140 normal-hearing participants, who were native English listeners aged between 20 and 90. Each participant heard the speech stimuli containing 50 monosyllabic words. The result showed that the phoneme scores obtained by listeners over 80 years differed significantly from those achieved by listeners aged below 80. From this finding, the authors suggested that instead of aging, it was probably hearing loss that caused the poor performance by the older adults (over 80 years). Similarly, Billings and her colleagues examined PRSs of three groups of listeners: young normal-hearing (18 to 35 years), and normal-hearing older listeners (60 to 80 years). The authors found a difference of 5% in phoneme scores between the young and older listeners.

However, the findings of the current study are not consistent with the results of Townsend and Bess (1980). They examined the effect of age on word scores for two groups of listeners: a younger adult group of 56 listeners (15 to 35 years) and an older adult group of 139 listeners (55 to 85 years). The NU-6 word list was presented to the listeners. The result showed that both listener groups obtained exactly the same word score, corresponding to 93% correct words. A possible explanation for the discrepancy across the studies might stem from the distribution of age in each group. For example, in the study by Townsend and Bess, all the listeners were grouped only into two age groups (15–35 years and 55–85 years) with a wider age range (20–30 years) within each. In Studebaker's study and the current research, the listeners were divided into small age groups with a narrower age range (10 years) within each. According to Bergman (1980), Willot (1991), and Studebaker et al. (1997), speech recognition performances remain until the 60s and deteriorate at the 70s or later with a small value. Hence, the age distribution of Townsend's and Bess's work might not have been an optimal approach, especially for the group of older adults.

With respect to the normative phoneme score obtained by listeners under 40 years, the findings are in line with those of Kuk et al. (2010). However, the findings are not consistent with those of Mackersie, Boothroyd, and Minniear (2001). Kuk et al. found that the younger listeners scored highly on phoneme identification, corresponding to 98% of correct responses when the speech material was presented in a male voice and 97% when it was done in a female voice. Nine native listeners of English, aged between 20 and 39 years, and with normal hearing took part in their study. A list of 115 CVCVC nonsense syllables was presented to each individual with a presentation level of 68 dB SPL in quiet. The second experiment of this study was conducted on a group of older listeners with hearing impairment. Thus, it was still unclear whether the older listeners with normal hearing could obtain similar phoneme scores on the CVCVC nonsense syllables compared with younger listeners. Mackersie, Boothroyd, and Minniear measured phoneme scores for 22 normal-hearing adults, who were English listeners, by using the speech material of Computer-assisted Speech Perception Assessment (CASPA) Test. The speech stimuli consisted of 20 lists of CVC words, with 10 words per list. The result suggested that the phonemes score in the study were in the range of 58–74% across 10-word lists with an average phoneme score correctness of 65 percent.

The higher performance reported by Kuk et al. and the current study, as compared with the remaining studies, can be attributed to the presentation level: 68 dB (SPL) in Kuk's work, 80 dB SPL in the current study, and 50 dB SPL in the study of Mackersie and her colleagues. The second reason for this discrepancy might be methodology. For example, the speech stimuli were delivered monaurally in Mackersie and her colleagues' work, binaurally in the current study. The methodology of Kuk's study is unclear.

### **7.5 Effects of dialects on speech perception in relation to AAST and NAMES**

The aim of this study was to determine whether dialect had any significant influence on speech perception by measuring SRTs on AAST and PRSs on NAMES in children, adults, and older adults. The speech stimuli were presented in the SVN. As predicted by the results of SRTs and PRSs, the non-native listeners obtained progressively poorer scores than the native listeners of SVN. In a statistical sense, the results showed significant differences across listeners' dialects in both speech tests. For AAST, the variation of SRTs between the native and the non-native listeners was extremely great, ranging between 2 dB and 6 dB (SPL) in quiet and between 1 dB and 5 dB (SNR) in noise (except the speech thresholds for older listeners). In NAMES, the differences in PRSs between the native and non-native listeners were negligible—roughly 2 percent of the correct responses.

Hence, the following hypotheses can be accepted.

*H5a. There are significant differences in the speech threshold values in AAST between the native and non-native listeners of a dialect.*

*(The native listeners achieve better SRT values than the non-native listeners)*

*H5b. There are significant differences in phoneme recognition scores in NAMES between the native and non-native listeners of a dialect.*

*(The native listeners obtain better phoneme scores than the non-native listeners)*

**Regarding AAST**, the findings on SRT do not support the previous research (Schneider, 1992; Nissen et al., 2013). Nissen et al. examined the effects of dialects in a Mandarin speech test on 32 Mandarin listeners aged between 18 and 50 who lived in the US. Half of the participants spoke Mainland Mandarin while the other half spoke Taiwan Mandarin. Speech material included 26 tri-syllabic words in Mainland Mandarin and 28 tri-syllabic words in Taiwan Mandarin. The authors found a less-than-2dB difference in mean thresholds between the native and non-native listeners. Based on the finding, the authors interpreted that such difference might not affect the clinical interpretation. In the same vein, Schneider investigated the effects of dialect on Spanish speech material on 12 Spanish children aged between six and seven years. The stimuli were chosen from three Spanish dialects. He found differences in SRTs ranging between 1.9 and 2.4 dB for speech materials relevant to the different Spanish dialects. In terms of clinical findings, Schneider also interpreted that the difference in SRTs among material dialects ranging from 1.9 to 2.4 dB would not alter clinical assessment. In this study, as it has been mentioned, the difference in SRTs among the dialects was 2 to 6 dB. This difference is substantial enough to influence clinical evaluation.

A possible explanation for the different results of the current study and that by Schneider and Nissen et al. is the listeners' age. The current research used a wider range of ages (six-year-olds, 20–30-year-olds, 60–70-year-olds), whereas two of these studies explored a much narrower age range, for example, six to seven years (Schneider), and 18 to 50 years (Nissen et al.). Secondly, the number of listeners in the study might affect the research findings to some degree, in particular, 12 children in Schneider's work, and 32 younger adults and adults in Nissen's study. The current study recruited 183 listeners who originated from Northern, Central, and Southern Vietnam. Thirdly, Nissen and his colleagues measured participants' SRTs in the US, not in China or Taiwan. Due to language contacts,

the listeners who spoke Mainland Mandarin could have become familiarized with Taiwanese Mandarin or vice versa.

Aside from the speech threshold differences of the native and non-native listeners, this research also found statistically significant differences in perception processing (measured in reaction time) between the dialects. Indeed, the native listener showed significantly quicker reaction time than the two groups of non-native listeners. Hence, our finding supports hypothesis H5c:

*There are significant differences in the response time in AAST between the native and the non-native listeners of the dialect.*

*(The native listeners need a shorter response time than the non-native listeners)*

These findings are in line with Adank and McQueen's work (2007), which showed a delay in word processing of roughly 100 ms when the native speakers of Dutch listened to an unfamiliar accent (East Flemish) compared with a familiar accent (Local Dutch).

**Regarding NAMES**, our findings on recognition scores are consistent with some published studies (Weisleder & Hodgson, 1989; Le et al., 2007; Shi & Canizales, 2012; Nissen et al., 2013), which showed significant difference in word recognition scores between the native and non-native listeners due to dialectal variation. Weisleder and Hodgson examined the effect of dialect on word recognition test for 16 college students who spoke Spanish. The participants came from different countries: Panama, Spain, Colombia, Mexico, and Venezuela. They listened to the speech stimuli presented in four different presentation levels: 8, 16, 24, and 32 dB HL. The authors found that at a low presentation level, the listeners of Mexican origin had significantly better recognition scores than others. The threshold difference disappeared when the intensity level increased. In the same vein, Le et al. (2007) examined the effect of dialect and word familiarity on speech test for 21 first-year psychology students at Western Sydney University aged between 18 and 39 years. The stimuli consisted of 36 English words recorded in three male voices spoken in different African-English dialects. The result also revealed that the listeners obtained progressively poorer scores when the speech stimuli differed from their own dialect. Similar to the above studies, Shi and her colleague also assessed the effect of dialect on the Spanish word recognition test for 40 Spanish/English bilinguals. The researchers found that the Highland listeners obtained a better recognition score than the Caribbean Coastal listeners. From this result, the researchers suggested that the dialectal effect was still present in the speech audiometry testing. In general, the researcher came to the same

conclusion: the phonological and phonetic variations of an unfamiliar dialect significantly affect word recognition. Thus, clinicians should be sensitive to a listener's phonetic features when they measure hearing abilities, and try to find an adequate speech test based on the individual's language background.

In contrast to the findings of these researchers, Nissen and his colleagues (2013) also found statistically significant differences among word recognition scores of the listeners' dialects. However, due to the minor discrepancies, they suggested that these might not clinically affect audiometric measurement. As mentioned earlier, the average phonemic scores in the current research were slightly different for the native and non-native listeners—roughly 1.5%. This difference is negligible, and might not influence the clinical finding as compared with those of Nissen and his colleagues.

The current study and Nissen's work are similar in terms of the phoneme scores on the supra-threshold speech in the listeners' dialects, but slightly different from those of Weileder and Hodgson (1989), Le et al. (2007), and Shi and Canizales (2012). Some possible explanations for this difference can be elucidated. First, it may stem from the methodology used in the studies. In the study by Weileder and Hodgson (1989), speech stimuli were presented monaurally to the listeners. In the current study, these were presented binaurally. Second, it may be caused by the presentation level of the stimuli. In the study by Weileder and Hodgson, the speech stimuli were presented at four different levels (8, 16, 24, and 32 dB HL). In the research by Shi and Canizales, the intensity level was 40 dB HL. In contrast, the speech stimuli were presented at the fixed intensity level of 80 dB SPL in the current research. Third, according to the Perceptual Assimilation Model (Best, 1995), adult listeners find it difficult to identify speech stimuli not presented in their own dialect. They will assimilate "the non-native phonemes to native phoneme where possible", and if not, "focus on auditory or articulatory information", especially in the case of meaningless stimuli. Our observation on listener responses is in agreement with this model of perception. The listeners reflected the influences between their dialect and the speaker's dialect through their responses. So, they either imitated the acoustic cues from the speaker's voices or ignored the signals not relevant to their own dialect. In addition, the speech signals in the current research were nonsense syllables, which are presumably more novel and challenging to listeners. Therefore, both native and non-native listeners put in more effort and were more cautious during their speech performances. This can also be elucidated through the Ease of Language Understanding (ELU) model (Rönneberg, 2014), which proposes that the listeners would modify their attention from relative effortlessness to more effort

towards incoming speech signals when the listening condition or the task requirement was sufficiently challenging. The above-mentioned reasons may have resulted in high phoneme scores for both native and non-native listeners in this study than those in other studies.

Comparing the SRTs and PRSs with respect to dialectal effects, the test procedures for AAST and for NAMES were totally different. However, this research was consistent in the results, in which the non-native listener groups obtained poorer speech thresholds/suprathreshold phoneme scores than the native group. However, due to the difference in methodology and speech materials between AAST and NAMES, the extent of differences in dialectal effects is not the same (larger difference in SRTs, smaller in PRSs). Nevertheless, these results reflected that the effect of dialect on audiological assessments is real and substantial enough to influence clinical assessments, especially as speech material is AAST. Furthermore, from our observation, the non-native listeners faltered while responding to the first stimuli in both AAST and NAMES. So, they more often asked for a reiteration of the stimuli. In contrast, the native listeners went through their performances relatively easily and smoothly with fewer requests for stimulus repetition.

This study only focused on the effects of dialects on younger adults. Children and older adults were not included in this part of the research. So, the extent to which the phoneme scores of the children and older listeners who spoke CVN and NVN was affected is not known. Further research may investigate this factor in these age groups.

### **7.6 Interdependencies of SRT, duo-tone threshold, and PRS**

The purpose of the present study was to examine whether there are any interdependencies between (1) the SRTs in AAST-a2 and duo-tone thresholds, (2) the PRSs in NAMES and duo-tone thresholds, and (3) the SRTs in AAST and the PRSs in NAMES.

*H6a. There are strong associations between duo-tone and AAST thresholds in quiet and in noise.*

*(The better the speech threshold value, the lower the duo-tone threshold value)*

This hypothesis can be accepted. The results of this study showed a strong correlation between the duo-tone threshold and the capacity to recognize speech material in quiet, whereas a poor correlation was found between pure-tone threshold and the ability to recognize the speech stimuli in noise. The results of the current study are in agreement with previous studies (Bosman & Smoorenburg, 1995; Wilson et al., 2005; Brandy & Lutman, 2005; and Vermiglio et al., 2012), which have also

shown a weaker relationship between duo-tone thresholds and speech recognition in noise and a stronger one in quiet.

In a quiet condition, the magnitude of the relations between duo-tone 0.5 kHz and the SRTs are larger as compared with those between duo-tone 4 kHz and the SRTs. This finding implies that the duo-tone threshold at 0.5 kHz is the best frequency for speech threshold prediction. In contrast, in noise, the magnitude of the correlation between duo-tone 0.5 kHz and the SRTs are smaller as compared with those between duo-tone 4 kHz and the SRTs. This may indicate that the duo-tone threshold 4 kHz is the best frequency for speech threshold prediction in noise. These findings are also in agreement with Bosman and Smoorenburg's results (1995).

Based on these associations, it can be asserted that the listeners' speech threshold can be predicted well from their duo-tone threshold values. However, the optimum is to predict those in quiet rather than noise (Vermiglio et al., 2012).

*H6b. There is an association between duo-tone threshold and NAMES scores.*

*(The higher the phoneme score, the better the duo-tone threshold)*

*H6c. There are strong relationships between NAMES scores and AAST thresholds.*

*(The better the speech threshold value, the higher the phoneme score)*

These hypotheses have to be rejected. The research found tenuous relationships between the duo-tone and the two speech materials, with  $r$ -values ranging from -0.3 to -0.1. These tenuous relationships indicate that the PRS cannot be predicted from the SRT or the duo-tone threshold.

The findings of the current research are different from Bosman's and Smoorenburg's finding, which showed a higher correlation between pure-tone thresholds at seven-octave frequencies (0.125 to 8 kHz) and phoneme scores (CVC syllable). In that study, the authors examined the relationship between these factors in Dutch for four groups of listeners, including a group of normal-hearing young listeners and three groups of hearing-impaired listeners. The research showed a significantly high correlation between phoneme scores and pure-tone thresholds for all listener groups with  $r$  values at 0.98, 0.97, and 0.76 over three subsets of pure-tone thresholds (0.125 to 8 kHz; 0.5, 1, 2 kHz; and 2, 4 kHz), respectively. A possible explanation is that the listeners were measured differently in the two studies. The normal-hearing listeners were recruited in our study, whereas both normal-hearing and hearing-impaired listeners were tested by Bosman and Smoorenburg. This can

be easily seen from a comparison of the magnitude of correlations for the normal-hearing listeners between these two studies. The results of the two studies found equivalences with respect to  $r$ -values: a weak correlation between the phoneme score and the pure-tone threshold for the normal-hearing listeners.

One again, there is no or a very weak relationship between the PRS scores and the duo-tone thresholds and the SRTs. Therefore, to precisely know the ability of listeners, the only approach is that the tester or audiologist should measure the phoneme scores because these cannot be predicted by other auditory tasks.

### **7.7 Summary of discussion**

Our discussion is about, normative values of the AAST and NAMES, learning effects on the AAST, dialect effects on speech recognition of the non-native listeners, effects of tonal patterns on speech perception of the older native listeners, and associations between the two speech materials and duo-tone (0.5 and 4kHz).

Regarding norm values of the AAST, the results of the current study support the hypothesis that the normative values depend on the listener's age in which the younger children and older adults obtained significantly poorer speech thresholds than the adults. The age-related normative values in this study are in line with previous studies. However, the normative value in the Vietnamese language is higher (2 dB) than those in German due to high levels of ambient noise in Vietnam. Learning effects still exist in the AAST Vietnamese between the first trial and the second ones for both quiet and noisy condition. Nevertheless, these differences are negligible and no statistical significance either.

Regarding normative values of the NAMES, we hypothesized that the norm values depend on listener's age, in which the older adults perform significantly poorer phoneme scores than the adults do. This hypothesis, however, can partly be accepted. The listeners below 75 years old obtained more or less similar in terms of phoneme recognition scores, whereas older listeners above 75 years old achieved poorer phoneme recognition scores. The result suggested that the NAMES is an easy speech test for listeners with normal hearing even those with mild hearing loss. A comfortable presentation level of 80 dB (SPL) seems to contribute to ceiling effects in performances of the NAMES.

Regarding dialect effects on speech performances, we hypothesized that the native listeners of the Southern Vietnamese achieve the better SRTs and PRSs than the non-native listeners do. The results

of the current study support the above hypothesis. The effect of dialect on audiological assessments is real. However, the extents of the differences are not analogous between AAST and NAMES, the larger differences in speech thresholds (AAST) and the smaller in phoneme scores (NAMES).

Regarding effects of tonal patterns on speech perception of older adult listeners, it was hypothesized that tonal patterns (F0) in terms of lexical tones in Vietnamese have an effect on speech recognition. This hypothesis can be only partly accepted. Tones with low pitch levels and falling contours become difficult for older listeners to identify or recognize, especially in noisy condition. The deterioration in tones with high pitch levels and rising contours does not stem from the high-frequency hearing loss but deteriorations of speech perception in older adults.

Lastly, regarding interdependencies of speech audiometric tests and duo-tone audiometry, the results of the study support the hypothesis that there is a strong association between AAST thresholds and duo-tone thresholds in quiet and noisy condition. It means that the SRTs can be predicted from their duo-tone thresholds. The optimum is to predict in quiet. Unlike the strong relationship between SRTs and duo-tone thresholds, the PRS scores showed very weak associations with duo-tone thresholds as well as SRTs.

## 8. APPLICATION STUDY OF AAST IN HEARING-IMPAIRED CHILDREN

### 8.1 Introduction

So far, the Adaptive Auditory Speech Test (AAST) in German, developed by Coninx (2005), has proved to be suitable for measuring hearing impairment to decide on the recipients of hearing aid (HA) among children (Coninx, 2005; Nekes, 2016) and cochlear implant (CI) (Hoffman, 2013) among adults. These studies have shown that hearing-impaired listeners show a large variety of SRTs in their performances in quiet and noisy conditions. For example, in Coninx's work, ranges of SRTs of 82 children (five and 12 years old) were more variable for 20 and 50 dB SPL. Similarly, in Nekes's research, the mean SRT of mild hearing-impaired groups (4 to 10 years) was  $32 \pm 8.9$  dB SPL, whereas the mean aided SRT of moderate hearing-impaired groups was  $42.8 \pm 9.3$  dB (SPL) in a quiet condition. Nekes's findings also revealed that the SRT of CI recipients was significantly better than that of HA recipients with severe and profound hearing loss.

AAST in Vietnamese was developed with five subtests. The normative SRT values were assessed for listeners aged between four and 85. The findings indicated that AAST in Vietnamese is a valid audiometry speech material to quantify speech recognition for normal-hearing listeners in both quiet and noisy conditions. The SRT values of hearing-impaired listeners were also used to compare the normative values of normal-hearing listeners, as well as the SRTs in other languages in case of hearing-impaired patients.

The characteristics of a speech material are usually examined with respect to SRTs, the slope of the intelligibility function, the correlations between speech recognition thresholds and pure-tone thresholds (Neumann et al., 2012; Weissgerber et al., 2012).

The aim of this work was (1) to examine the applicability of AAST in quiet for diagnostic purposes for hearing-impaired children with HA, (2) to assess the slopes of psychometric functions for two subtests of AAST, (3) to investigate the effect of tonal patterns of disyllabic words on the speech perception of HA recipients, (4) to investigate the relationship of SRTs with duo-tone thresholds.

Verbal communication in daily life is often masked by background noise. The speech test in noise is considered as the best method to replicate the listening condition. Therefore, the speech test in noise is widely used for children with hearing disorders (Weissgerber, 2012). However, in the case of children with severe or profound hearing loss, their linguistic abilities are extremely limited. An

assessment of speech recognition is only feasible in a quiet condition. This study conducted a speech test in quiet on hearing-impaired children aged between 10 and 14.

## 8.2 Methods

### Participants

Two subtests of AAST were administered to 12 children aged between 10 and 14. All the test subjects were studying at the Thuan An Center for Disabled Children at the time of the data collection. The degree of hearing loss of these 12 children was between 100 and 110 dB HL. The test took place in an acoustic room for deaf children at the centre. Duo-tone audiometry (0.5 and 4 kHz) was performed in free fields for aided condition. The data from seven hearing-impaired children (15 ears) were included in the analysis. The data for five children out of the 12 had to be excluded because of incomplete measurements due to lack of speech perception (e.g. a misunderstood task). The measurements of seven children were used in the data analysis. The duo-tone threshold ranged from 24 to 47 dB HL for aided conditions across individuals. The average aided duo-tone threshold (0.5, 4 kHz) was roughly 36.8 for the hearing-impaired children, including < 25 dB HL (one child), 25 to 40 dB HL (three children), and 41 to 55 dB HL (three children).

### Materials

The study used two subtests of AAST-VN<sup>3</sup>, and AAST-aTP. As mentioned before, the speech stimuli of AAST Vietnamese bore tones A1 and B1 with high pitch levels and rising contours. The speech stimuli of AAST aTP carried tones A1, B1, and A2 with heterogeneous pitch levels and pitch contours: high level, high rising, and low falling.

### Test Procedure

The hearing measurement was conducted in a sound-treated room at Thuan An Center. The calibration of the equipment was made before the commencement of the hearing measurements. To become familiar with the test procedure, the tester introduced the speech stimuli of AAST as well as the response mode to the listeners.

The measurements were taken in hearing devices, and the acoustic signals were presented via loudspeakers (Bose Companion 2 series III) to the children. After the performances of the duo-tone task (only one time), the AAST Vietnamese and aTP were delivered twice for each speech material. The data for hearing-impaired listeners were computed separately for each individual due to the different degrees of hearing impairment.

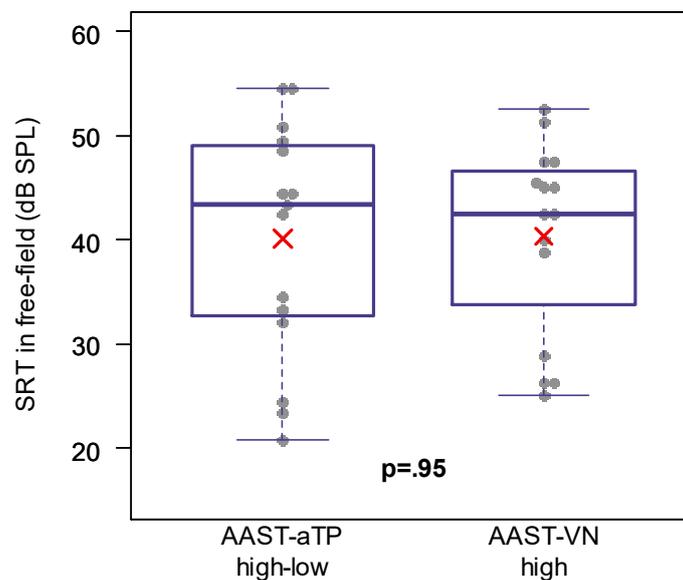
---

<sup>3</sup> The AAST-a2 is regarded as a standard speech test for Vietnamese, named as AAST-VN in this application study.

### 8.3 Results

The results of the SRT measurements for the seven hearing-impaired children are given in figure 30. The poorest speech threshold of AAST-VN was found for listeners with average aided duo-tone thresholds in the range of 25 to 60 dB HL. In contrast, listeners whose duo-tone threshold was less than 25 dB HL obtained better SRTs with a range of 25 to 29 dB SPL. Due to the different degrees of hearing impairment among individuals, the study showed large inter-individual differences in a quiet condition for hearing-impaired listeners as compared with the threshold values of normal-hearing children (see the norm values in the section of results).

The mean SRTs in both subtests of AAST were equal—roughly 40 dB SPL with standard deviation  $\pm 11.3$  dB for AAST-aTP and  $\pm 9.3$  dB for AAST-VN. Although a large range of SRTs was observed for AAST-aTP compared with AAST-VN, no significant difference was found between AAST-aTP and AAST-VN as a function of the tonal pattern effects,  $p$  value=0.95. This revealed that the differences in the tonal patterns of the two subtests of AAST did not affect speech recognition in the group of hearing-impaired children.



**Figure 30:** SRT values for hearing-impaired children, N=8 (N=15 ears)

The relationships between duo-tone threshold (0.5 and 4kHz) and speech thresholds of the two subtests of AAST are plotted in Figure 31. Significant correlations ( $r=0.84$  for AAST-VN and  $r=0.94$  for AAST-aTP) were found for bilateral duo-tone thresholds and SRTs of the two subtests of AAST in quiet thresholds. As duo-tone thresholds increase, so do AAST thresholds in quiet.

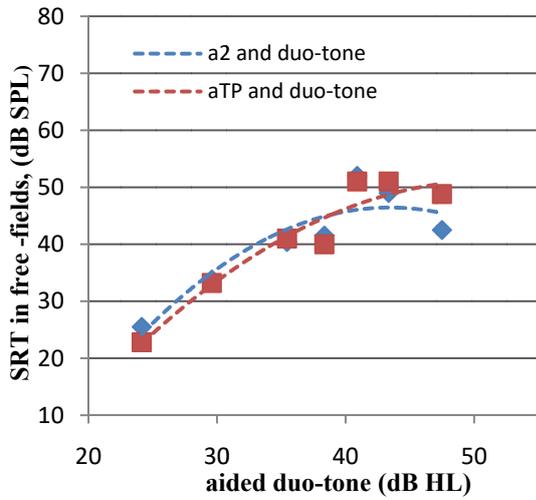


Figure 31: SRT values of AAST in quiet, correlated to aided duo-tone thresholds of HA users

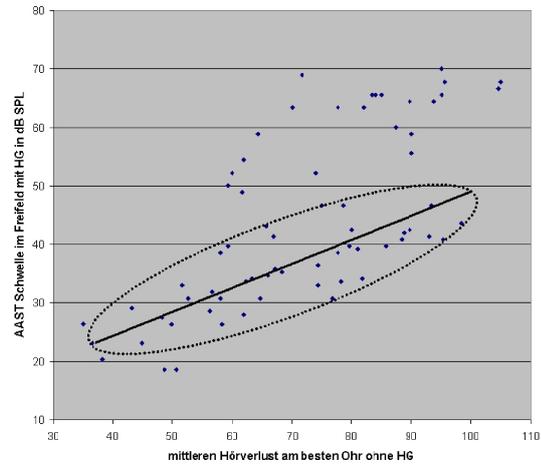


Figure 32: SRT values of AAST in quiet, correlated to unaided PTA of HA users (Coninx, 2005)

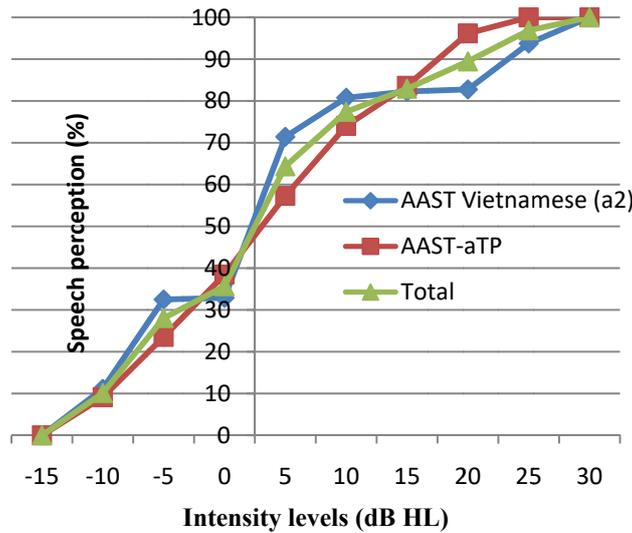


Figure 33: Psychometric curves for hearing-impaired children with hearing aids in quiet

Psychometric curves are calculated for each subtest of AAST in figure 33. The psychometric curves of both subtests are different from each other, roughly 7.5%/dB for AAST-VN and 4.0%/dB for AAST-aTP. The slope of AAST-VN is somewhat different from that of AAST-aTP at a presentation level of 5 dB HL: 71% intelligibility for AAST-VN and 57% intelligibility for AAST-aTP. This difference might indicate that AAST-aTP is somewhat difficult than AAST-VN in terms of intelligibility of hearing-impaired children. In average, the slope is roughly 6.0%/dB at 35% and 65% intelligibility for SRTs in both subtests.

## 8.4 Discussion

### **The SRT, psychometric curves, and correlations between SRT values and duo-tone thresholds**

The mean SRT (AAST-VN) for a sample of seven HA recipients was  $40 \pm 9.3$  dB (SPL). The average threshold of this group was substantially different (roughly 10 dB SPL) from those of the normal-hearing children, aged eight. The study also showed a larger standard deviation for the hearing-impaired children ( $\pm 9.3$ ) as compared with the normal-hearing group ( $\pm 2.5$ ). Unfortunately, no reference data was available for HA users in Vietnamese. Therefore, comparisons between the current and previous findings were not possible. In the current research, five listeners showed profound hearing loss (100–110 dB HL). However, they obtained different speech thresholds of AAST in both subtests. This indicated that the amplification of the hearing devices is not linear. They are active differently at the level of the speech thresholds.

An improvement of the SRT value is insignificant when the gradient of the speech intelligibility function is unidentified. To interpret the clinical improvement of a client, the gradient of the psychometric function can help estimate the improvement relevant to speech intelligibility (Dietz, 2015). The gradient of psychometric curves of the Vietnamese AAST is roughly 8.2%/dB for normal-hearing subjects. The gradient of the psychometric curve was 6%/dB for seven hearing-impaired children with aided SRTs in a free field condition. The gradient of the group of hearing-impaired listeners was gentler than those of the normal-hearing listeners in present study. However the mean slope values of AAST on the hearing-impaired children was in line with those conducted by Weißgerber et al. (2012) which found a mean slope value of 7%/dB in aided condition for children aged 4 to 10 years old.

The aided duo-tone threshold could give at least an estimate of speech threshold and vice versa at least for the duo-tone threshold between 20 and 40 dB HL. In contrast, the patients with duo-tone threshold above 40 dB HL showed heterogeneous SRT values in both subtests of AAST. The correlations between the aided SRT values and the aided duo-tone thresholds in the group of hearing-impaired children ( $r=0.84$ ) were significantly higher than those in the normal-hearing groups (0.69) in a quiet condition. This difference might have arisen from the different degrees of hearing loss among hearing-impaired children, or a relatively small sample size in this study.

### **Comparison of AAST Vietnamese with AAST German**

The mean SRT of AAST-VN was  $40 \pm 9.3$  dB (SPL) for the profound hearing-impaired children. This SRT value in this study is in line with the SRT values of AAST-German in recent studies conducted

by Nekes (2016), and Coninx (2005). Sandra Nekes (2016) conducted a study on 277 hearing-impaired children with hearing aids (4 to 10 years). The research found SRT values of  $42.8 \pm 9.3$  and  $45.9 \pm 14.1$  dB (SPL) for moderate and profound hearing impairment with hearing aid. The result of the current research seems to be comparable to the findings of Coninx who found SRTs between 40 dB and 47 dB SPL for children (5 to 12 years) with profound hearing impairment (see figure 32 in ellipse). Compared with Nekes's finding, the current research showed a slight difference in terms of mean SRT (2 to 5 dB SPL). A possible explanation for the difference might be ascribed to the patient's age. The patients' ages ranged from four to 10 years in the research of Nekes, whereas it ranged from 10 to 14 years in the current study. Furthermore, the current research includes data from only one patient with mild hearing impairment.

### **Effects of the tonal pattern on speech perception by hearing-impaired children**

The present study was designed to examine the effect of tonal patterns of syllables on speech recognition by hearing-impaired children with profound hearing loss. The speech stimuli of AAST-VN carried high-level and high-rising tones. Aside from high-level and high-rising tones, the speech stimuli of AAST-aTP included low-falling tones. The current study found equal speech threshold values between AAST-VN and AAST-aTP with mean SRTs of roughly 40 dB SPL. A wider range of SRTs was found for AAST-VN as compared with AAST-aTP. However, this negligible difference may not be derived from different tonal patterns between the subtests. The finding, while preliminary, suggests that the tonal patterns of syllables do not affect speech recognition of hearing-impaired children when they use the subtests of AAST-VN and AAST-aTP. Due to the small sample size of participants, further work is needed to examine the speech threshold not only using the subtests of AAST-VN and AAST-aTP but also using AAST-a3, a speech test with low-falling tones. In addition, further study on the current topic is also suggested for CI recipients to make a comparison with HA recipients. Furthermore, the benefits of hearing devices in recognizing speech for hearing-impaired children with and without hearing devices needs to be investigated when using either hearing aids or a cochlear implant.

### **8.5 Conclusion**

AAST-VN was considered to be suitable for assessing hearing-impaired children with HAs. Comparable outcomes were found between AAST-VN and AAST-German for hearing-impaired children. This suggests that AAST-VN may become a reliable speech material for the evaluation of patients with different degrees of hearing loss. More research will have to be carried out to get more referent data about the characteristic of the Vietnamese AAST for hearing-impaired patients.

## 9. GENERAL DISCUSSION

### 9.1 Evaluation of speech materials

#### Practical uses of the speech materials

The normative values of the two speech materials in this research show the reliability of these materials for routine clinical implementation. These materials also show reliability in high background noise. The speech materials can be easily implemented in a laptop installed with BELLS, and connected to other standard devices, such as headphones, a soundcard, and a microphone.

For AAST, the speech stimuli were presented in random, and there can be infinite iterations of the speech test. Hence, the learning effect is minimized, as reported in the result section. There were minimal effects of cognitive and linguistic skills, for example, active vocabulary, working memory, and speech production (Coninx, 2016). The test performance in the current study was fast and motivating. Furthermore, the slopes were 8.2%/dB for quiet and 8.4%/dB for noise at 50% speech intelligibility. The slopes of Vietnamese AAST seemed to be somewhat shallower than those of German and Ghanaian, which have slopes of about 10.2%/dB in a quiet condition. However, the small difference of slope values between these languages is trivial. Hence, the slope of about 8%/dB for normal-hearing listeners is acceptable. In addition, age-related normative values and the normative values in Vietnamese AAST are more or less equal to those in other languages. Besides, the proportions of word confusion are minor for six speech stimuli of Vietnamese AAST across the age groups. Taken together, it can be confirmed that the Vietnamese AAST can be a reliable speech audiometry test for hearing measurement in the Vietnamese language.

For NAMES, the speech stimuli do not depend on the vocabulary, and the morpho-syntactic skills of individuals. Moreover, the result of NAMES provides not only phoneme identification scores but also very detailed information relevant to phoneme pattern errors showed by listeners. In addition, the testing time is short, roughly two minutes per 20 words. Furthermore, several participants showed interest in the novelty of the speech stimuli of NAMES. A few listeners showed difficulty in articulation due to misunderstood tasks or even the novelty of acoustic stimuli in the early testing performances. However, after the second trial, they felt comfortable going through the performances and showed the ease of articulation. Taken together, what these findings mean is that the NAMES test can be suitable for audiological assessments and phoneme identification testing in the Vietnamese language.

The results of NAMES were based on the listeners' articulation or phoneme-grapheme correspondences. Hence, several issues must be taken into account while administering NAMES. Firstly, the clinicians or researchers who administer the speech test must ideally be trained in phonetic transcription. With this linguistic skill, the clinician can determine the listener's responses accurately and give proper phoneme scores. Secondly, to ensure that the clinicians can identify phoneme responses accurately, the listeners' responses must be loud enough. More importantly, the clinicians should observe the patient's mouth to capture the extra-visual information of the target phoneme. Additionally, NAMES results are scored by testers (clinicians or researchers), not by software (as AAST). Hence, more subjective biases from clinicians might exist in the administration. To ensure high reliability, the same clinician should record the scores while collecting clinical or research data. Finally, since the speech test is based on verbal repetition, the participants need to have sufficient production skills to give the verbal responses. Thus, the speech material might be difficult for younger children who may lack articulation skills.

Officially, there are around 1.4 million hearing-impaired people in Vietnam. Among those, more than 180,000 are children (under 18 years) (Nelson, 2015). The actual number may be much higher. However, hearing care for such people is quite limited. For example, newborn hearing screening is not widespread. At the earliest, intervention programs for children begin at two years. Audiology assessment is also restricted (Madell, 2013). Furthermore, Vietnam lacks audiological materials like audiometers, test booths, and hearing devices, especially CIs. In early 2005, CIs were introduced in Vietnam. But the number of CI users is still limited. The recipients are young children. The price of a CI is roughly US\$50,000 and has to be borne by the children's families. They were quite expensive compared with the average per capita incomes in 2015 (US\$1,685). So, very few families can afford a CI for their child. To assess hearing, audiologists use only pure tone. No speech materials are available to assess hearing, though word recognition tests have been established (Ngô, 1977; Nguyễn, 1986; Nguyễn, 2005).

### **The difference of 2 dB in AAST-a2**

By comparing the speech thresholds of German AAST and AAST-a2, it was found that the average speech thresholds in Vietnamese (for young listeners) were still higher (2 dB SPL) than those in German. The average speech threshold for normal-hearing German subjects was around 28 dB (SPL) in a quiet room, whereas the best threshold values averaged roughly 29.5 dB (SPL) for AAST-a2. The difference of 2 dB might be caused due to language, speaker-dependent factors (Theunissen

et al., 2009), or the level of background noise. The current data were collected in Vietnam where noise—for example, traffic and different environmental sounds—was omnipresent. The speech audiometry test was done in the quietest room. However, sometimes, considerable background noise interfered. According to Chan et al. (2005), the background noise was measured at 50 dBA for classrooms in Hong Kong. Similarly, the ambient noise level in Vietnamese classrooms is extremely high, especially in large cities.

To ascertain the extent of the reliance of speech threshold on ambient noise, the SRT for 18 German listeners was measured at five different ambient noise levels: 30, 35, 45, 55, and 65 dBA (see Appendix K). The results revealed that the speech threshold worsened gradually as ambient noise increased. The speech thresholds were detrimentally affected when the background noise went up to 45 dBA. The speech thresholds at the noise level of 35 dBA were exactly equal to those at a noise level of 30 dBA (normal ambient noise level in a quiet room). As it was mentioned in the section of methods, the ambient noise was measured at 38 to 48 dBA in Vietnam, which led to a difference of 2 dB (SPL) for Vietnamese as compared with German. Therefore, the extreme background noise level in Vietnam could be an exclusive explanation for SRT differences in this study.

In addition, the authors of this study also tried to examine whether the speech threshold in Vietnamese was equivalent to those in German when the hearing measurements took place at similar background noise levels. Fifteen Vietnamese adults were involved in a test held in a silent room at the Department of Special Education & Rehabilitation, the University of Cologne. As expected, the speech threshold in Vietnamese was exactly equal to those in the German language “AAST-DE” (see Appendix L), which adapts to normative values for normal-hearing individuals with an SRT of  $25 \pm 5$  (dB SPL) in AAST.

### **Word scoring and phoneme scoring in NAMES**

There are two scoring methods in a word recognition test: word scoring and phoneme scoring. In the current study, the word recognition scores were also computed. However, the result was primarily reported in phoneme scores because this showed an advantage. First, a larger number of phonemes reduce variability (Boothroyd, 1968; McCreery, 2010) whereas a smaller number of words increase variability and results in a larger extent of hearing loss (Billings et al., 2015). For example, when the phoneme scores were calculated, the average scores ranged from 94.5% to 97.5% correct phonemes (3% differences, all listeners had normal hearing). In contrast, when the results were calculated on the basis of words, the average scores ranged from 75% to 87.5% correct words (12%

differences, all listeners with mild hearing loss). Second, phoneme scores may give us a better understanding of a listener's ability to identify phonemes through their phoneme perception errors (Kramer, 2008). Based on this research, phoneme scoring is considered to be a more proper approach than word scoring.

### **Application of speech-in-noise in NAMES**

In future assessments, it is possible to present the speech stimuli at different presentation levels in several signal-to-noise ratios. Phoneme scores may be achieved differently at various intensity levels or in the presence of noise. Using a wider range of presentation levels was considered to be the optimal method (Boothroyd, 1986 and 2008; Ullrich & Grimm, 1976; Beattie & Zipp, 1990; Guthrie & Mackersie, 2009). It might provide the clinician with a complete assessment of a listener's PRS. Furthermore, speech in daily life is often masked by ambient noise. So, it is necessary to replicate this hearing condition for a speech audiometric test (Neumann et al., 2012; Weissgerber et al., 2012), which also avoids a ceiling effect, as it appeared in the current study.

### **Behaviors of participants in the performance of NAMES**

Based on our observation during the test performance, the listeners tended to pay more effort, attention, and vigilance to nonsense speech stimuli in the NAMES test. They considered speech signals as a benchmark and could adjust their perception to the speakers' accent as far as it was possible. Consequently, the listeners sometimes imitated the intonation patterns of speech stimuli or even the speaker's accent. This observation was consistent with the Perceptual Assimilation Model (Best, 1995) and previous studies (Evans & Iverson, 2005; Brunelle, 2009).

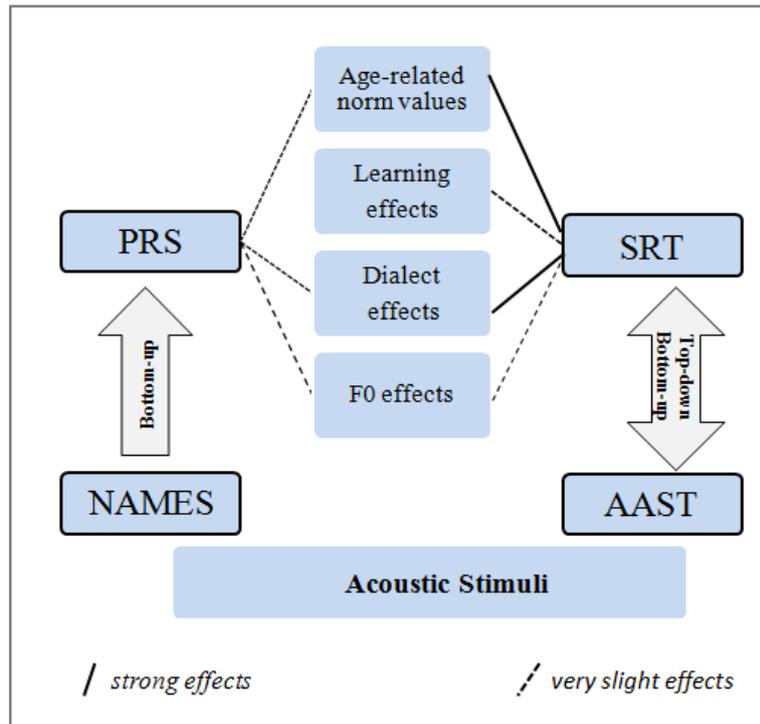
### **Top-down and bottom-up processing in the performances of AAST and NAMES**

The present study, on the other hand, reflects a great interest in the assessment of speech recognition by using different kinds of speech materials. The two speech materials have a similar phonetic content (based on frequently used phonemes in the language). However, these speech materials varied regarding their semantic content. The speech stimuli of AAST are disyllabic noun phrases with a different lexical content than the speech stimuli of NAMES, which are also disyllabic structures but without any semantic content. These two speech materials were used to assess the contribution of bottom-up and top-down processing on the speech recognition performance of children up to older adults.

The semantic information in the acoustic signal of a word can be referred to as a top-down process to compensate for an insufficiency of auditory perception. Dirks et al. (2001) stated that a reduction of lexical or semantic information included in a speech signal negatively influences the speech recognition of listeners, especially older adults, because the benefits of top-down processing are restricted and removed.

In contrast, findings of the current research indicated that the overall results of the meaningless NAMES test were significantly better and less variable than the overall results of AAST. A small difference (a slight effect) was found upon evaluating age-related normative values and the effects of dialectal and tonal patterns (F0) by using the meaningless CV-CVC syllables. In contrast, a larger difference (a strong effect) was found upon evaluating such issues using the speech test of AAST. The effects of bottom-up and top-down processing on speech performances are illustrated in Figure 34. By comparing the results of the speech recognition performances using both AAST and NAMES, it can be demonstrated that the speech recognition abilities of listeners do not seem to be influenced by semantic contents of speech stimuli in the current study.

Differences in the characteristics of speech materials can account for the contrasting conclusion as compared with the literature on roles of lexical contents on speech perception. AAST is a SRT test, whereas NAMES is a supra-threshold speech recognition test. AAST measures hearing abilities at a threshold level with an adaptive procedure. A stimulus is delivered at 65 dB. When it is identified correctly, the next stimulus is presented at a lower intensity level (reduced by one step or 5 dB). When the stimulus is not identified correctly, the following stimulus is increased (by two step or 10 dB). In contrast, NAMES administers speech stimuli at a constant supra-threshold level of 80 dB SPL. With this comfortable intensity level, even a normal-hearing older adult can easily identify the stimulus being presented.



**Figure 34:** Effects of top-down lexical semantic and bottom-up acoustic processing of AAST and NAMES

## 9.2 Implications of the study

**Regarding the field of audiology**, the findings of this study may have significant implications for audiology in Vietnam.

First, norm values of AAST and NAMES

The research assessed the hearing ability of both normal-hearing listeners and 12 hearing-impaired children. The values of speech thresholds for AAST and speech supra-thresholds for NAMES were obtained and evaluated in aided condition (i.e. with hearing aids). Preliminary results confirmed that both speech materials are suitable to assess hearing abilities. However, the range of values in these two speech tests was not enough for the classification of the degree of hearing loss. Results of the present research showed a very large range of AAST thresholds for normal hearing (40% listeners, SRT<30 dB SPL), mild hearing loss (50%, SRT=31–40 dB SPL), and moderate hearing loss (10%, SRT=41–55 dB SPL). In contrast, there was a smaller range of NAMES phoneme scores for normal hearing (98.5% listeners), and mild hearing loss (1.5%). The listeners with a poorer speech threshold still achieved relatively high phoneme scores. Our findings confirmed that the choice of speech test used to evaluate speech perception has a significant effect on the outcomes. Therefore, clinicians or

researchers of hearing assessments need to be aware that norm values for different speech materials may influence the classifying degrees of hearing loss for a person.

#### Second, linguistic backgrounds of listeners

The results showed that the native listeners of Southern Vietnamese scored significantly better SRTs and PRSs compared with non-native listeners. This indicated that linguistic backgrounds of clients significantly affect audiological assessments. Understanding the significance of a speech recognition deficit under dialectal effects may help audiologists or speech pathologists to be more aware of speech materials being used for audiological assessments. When selecting a speech material for audiological evaluations, use of a protocol with speech audiometry materials appropriate for clients is recommended to avoid misdiagnosis or inflated assessment scores.

#### Third, tonal distributions in speech materials

As reported in the section of results, the average SRT in a4 was somewhat lower, indicating a better result compared with other speech materials, both in quiet and noisy conditions. In particular, each speech material should involve tones in different pitch heights and contours to ensure the naturalness of speech materials. For example, the speech material a3 carries only tones with low pitches and falling contours. When the speech stimuli were presented with masking noise, the listeners were more confused. In the case of the speech material a4, when each syllable involved tonal contrasts such as high-low, the high one could act as an acoustic cue (acoustic salience), drawing the older adult's attention to the stimuli in a complex hearing condition, for example, in masking noise. For optimal performances, the design of a new speech audiometry material should also concern harmonic distribution of tonal patterns across speech stimuli.

Furthermore, regarding NAMES, the tone A2 (a low-falling tone) was identified as B2 (a creaky falling tone). These two tones have similar pitch contours making them acoustically indistinguishable. As a result, listeners are perceptually indistinguishable between such tones.

These above disadvantages remind audiologist that an issue associated with effects of the low-falling tones need to be considered for any assessment of hearing or speech performance in noisy condition in the Vietnamese language.

#### Fourth, phonological error patterns

NAMES test assesses not only phoneme identification abilities but also phonological error patterns of clients. Hence, the results of NAMES can give clinicians specific information on error patterns

that the patients exhibited in their speech performances. By detecting these mistakes, the clients can modify their speech with support of clinicians or speech pathologists.

**Regarding the field of linguistics**, the results of this research provided some important insights into speech perception of non-native listeners of a dialect, characteristics of speech perception by older adults as compared with younger adults in native listener groups, errors patterns of speech perception in Southern Vietnamese, and tonal perceptions in older adults.

Firstly, perceptual differences in segmental and suprasegmental phonemes across dialects

As reported in the section on literature, the dialects were phonetically different from each other. Indeed, the present study provided evidence to prove that phonological differences exist and influence phoneme identification. The non-native listeners made more errors when the perceived syllables were followed by the coda glide /-j/ or the velar coda /-ŋ/, as well as when the syllables carried the tone A2 in Southern Vietnamese. For the onset palatal /j-/, the non-native listeners tended to interchange it for /v-/ or /z-/. In contrast, the native listeners gave a single form of answer, /j-/. For the coda velar /-ŋ/, the interchange between /-ŋ/ and /-n/ was prevalent among the non-native listeners. There was a one-to-one correspondence between a speech stimulus and a response to /-ŋ/ in the native listeners' speech. With respect to suprasegmental phonemes, a larger extent of tonal misidentification for tone A2 was found among the non-native listeners, whereas a lesser extent of tonal misidentification was found for tone A2 in the groups of native listeners.

Secondly, social effects on linguistic behaviours

The research reflected an interesting observation that the short vowels /ă/ and /ǎ/ and the rounded vowels /o/ and /ɔ/ were more easily confused by young adults. In contrast, the older listeners showed less confusion in identifying such vowels. The indistinguishability between these vowels occurred in young adults but rarely appeared in older adults. The differences in perception and production of such vowels might reveal a change of linguistic behavior between young and older adults in Southern dialect. From this finding, teachers or educators should be aware of speech perception and production skills in children, and especially confusable phonemes due to similar acoustic properties, for example, between /o/ and /ɔ/, and between /ă/ and /ǎ/.

### 9.3 Limitations of the study

The research has some unexpected limitations that might affect the findings to some extents. Firstly, the study did not include cognitive measures. Some listeners have slightly poorer speech thresholds due to a decline in cognitive capacity and not because of hearing loss. Without the cognitive measure, we could not fully assess the effects of cognitive ability on deterioration in speech recognition of listeners, especially older adults whose education level was limited. Future studies should examine the cognitive factor, working memory capacity, as well as attention.

Secondly, high level of background noise influences test performances, especially for the AAST task. Both measures took place in kindergartens, and school libraries where noise interfered. Some measurements were run in a musical room of the university but background noise still existed. As a result, the speech threshold values of AAST-a2 were higher (roughly 2 dB SPL) than those of AAST German. The presence of background noise constituted a great challenge to older adults, as speech stimuli were presented in masking noise.

Thirdly, the data of NAMES from non-native listeners did not include data from older adults due to the small sample size. Therefore, dialectal effects might not be fully appreciated with nonsense syllables. Further research needs to be considered.

Lastly, the data of hearing-impaired children were obtained on a small sample (N=7). It is difficult to make a full assessment regarding the applicabilities of the speech material of AAST for hearing impairment with different degrees of hearing loss. The NAMES test was also used to assess this group of hearing-impaired children. But most children had impaired hearing and insufficient production skills. Therefore, the NAMES data from the group were not included in this study. A larger sample of hearing-impaired children with a wider range of degree of hearing loss needs to be studied for further research.

### 9.4 Future research

This research has thrown up many questions that need further investigation. First, the speech tests are reliable tools for audiological and speech assessments. The speech tests have been used to measure the hearing ability of normal-hearing listeners. However, a question remains on the relevance of the method to judge the speech threshold or phoneme scores in different populations. It would be necessary to conduct the test on hearing-impaired populations with varying degrees of hearing loss: mild to severe hearing loss with CI in various age groups. For AAST, younger children aged four can be included in the measures. However, NAMES seems to be appropriate for hearing-

impaired children above 10 years. Second, participants recruited for the NAMES test were aged 15 years and older. This speech test should be evaluated for children to assess their phonological awareness skills compared with young adults and adults. From this assessment, the phonemes that a listener could not produce can be identified and a good provision for that made during the test performances. Third, the measured values of supra-threshold PRSs are high in a quiet condition, which reflects the presence of near-ceiling effects for NAMES. An application of speech-in-noise is required to validate the phoneme scores in a noisy condition. Finally, the current study was based on the native speaker form of Southern Vietnamese. Thus, the SRTs were variable for the non-native listeners. Hence, additional speech materials of AAST or NAMES have to be designed for Northern and Central Vietnamese dialects.

## 10. SUMMARY AND CONCLUSION

The first part of this study was concerned with designs of speech audiometry materials based on Southern Vietnamese. Speech materials of AAST (Adaptive Auditory Speech Test) and NAMES were constructed. AAST included disyllabic noun phrases containing lexical contents. Five subtests of AAST were digitally adjusted to make it homogeneous regarding psychometric functions, speech intelligibility, and speech recognition thresholds. The average psychometric slope for five subtests of AAST was 8.2%/dB in quiet and 8.4%/dB in noise. These slope values seemed to be close to the slope values in the German language. NAMES comprised disyllabic structures without lexical contents. Four sublists of NAMES were digitally modified to make them congruent to the durations of syllables, and the total root mean square (RMS) levels across acoustic signals.

The second part, which was the main content of this thesis, dealt with normative values, as well as the effects of dialects and tonal patterns on speech perception. This part also examined the correlations between speech audiometry materials and duo-tone audiometry. The speech materials of AAST and NAMES were used to assess the above-mentioned issues. The data analyses gave the following results:

Regarding the normative values, AAST found a strong correlation between the listeners' age and speech threshold values, as predicted. Mean SRTs were significantly worse for young children aged four years and older adults above 75, and significantly better for young and adult listeners. The difference in the SRTs of the young children and the adults was 8 dB (SPL), and of the adults and the oldest adults was 11 dB (SPL). In contrast, NAMES found a weak correlation between the listeners' age and their PRSs. The PRSs did not vary significantly with increasing age for listeners aged between 15 and 75. The factor of age might have impacts on the speech reception of the oldest listeners beyond 75 years with a deterioration of 3% correct phonemes as compared with adults.

Regarding the dialectal effects, the speech material of AAST threw up significant differences in word recognition due to dialectal variations. The non-native listeners performed poorly on speech threshold values and reaction times than the native listeners of Southern Vietnamese. In contrast, NAMES found negligible impacts on speech performance for the non-native listeners compared with the native listeners, a difference of less than 2% correct phonemes. The influence levels were not similar. Depending on the speech materials used to diagnose hearing, a lower extent of effects was found for nonsense disyllabic tests presented at a fixed presentation level, and a greater extent of

effects for sensible disyllabic tests presented in an adaptive procedure. However, the results of this research support the hypothesis that dialectal variations substantially influence speech audiometry testing. Dialectal effects on speech recognition are “real”. For hearing evaluation, the clinicians should try to find an adequate speech audiometry material, as well as be sensitive to phonological properties of dialects shown by patients.

Regarding the effects of tonal patterns on older adults, AAST found no significant differences in quiet but did so in noise among the five subtests. Speech stimuli of a3 carried tones with low pitch levels and falling pitch contours. As these speech stimuli were presented in masking noise, the speech-in-noise seemed to be more challenging for older listeners. Therefore, their average SRT value for AAST-a3 was significantly worse as compared with the remaining subtests. The NAMES test showed a link between pitch levels or pitch contours of tones and response correctness. Speech stimuli carrying the tone ngang (A1) were identified better than those bearing the tones huyệ̀n (A2), and sặ́c (B1). The finding might indicate that older adults obtained better scores for tones with flat contours rather than tones with complex contours, such as low falling (A2) and high rising (B1). Since the F0 values of high-rising tone B1 were less than 500 Hz, the poorest score in the perception tone B1 stemmed from age-related changes in speech recognition but not due to high-frequency hearing loss in the older listeners, as assumed. The findings of this research partly supported the hypothesis that in terms of lexical tones, tonal patterns (F0) in Vietnamese have an effect on the speech recognition of the older native listeners of SVN above 75 years. As mentioned before, the decline in tonal identification scores is due to age-related declines in speech reception, and not high-frequency hearing loss.

Regarding the correlation between speech audiometry materials and duo-tone audiometry, the result illustrated relatively strong correlations between speech thresholds and duo-tone thresholds. However, no or weak association was found between PRSs and duo-tone thresholds. There was also no correlation between SRT and PRS. These findings implicated that speech thresholds in quiet and noise can be well predicted from duo-tone thresholds, whereas phoneme scores cannot be predicted from duo-tone thresholds or SRTs.

The third part of the thesis dealt with the application of AAST for hearing-impaired children. The research determined the speech threshold values, the psychometric curves, the effect of tonal patterns of syllables on speech perception, and the correlations between aided speech thresholds and aided duo-tone thresholds. The preliminary results of the measurement in seven hearing-impaired

children showed some interesting observations. The SRT values of the severe hearing-impaired children were comparable to those obtained in AAST German, roughly 40 dB (SPL) in free field condition using AAST Vietnamese. There was a strong significant correlation between the speech threshold values and aided duo-tone thresholds in aided condition (i.e. with hearing aids). As compared with psychometric curves of normal hearing, the slope value of hearing-impaired children was two times higher. However, the difference might have been an unexpected observation. Further studies need to be conducted on larger samples of hearing-impaired listeners. Regarding the effects of tonal pattern on speech perception, there were no significant differences in speech threshold values between AAST Vietnamese and AAST-aTP. The finding might suggest that tonal patterns do not affect speech reception of hearing-impaired children with hearing aid.

In summary, the two speech audiometric tests have been designed and evaluated regarding norm values, dialectal effects, effects of tonal patterns of syllables on speech receptions, and correlations between speech audiometric tests and duo-tone audiometry. The findings of this thesis render the useful information regarding the norm values, the dialects or linguistic properties in speech audiometry materials in Vietnamese audiology. Outcomes of the current study insisted that AAST and NAMES are reliable speech materials, easy to use, and robust to background noise. These two speech audiometric tests complement each other in evaluating impairments of hearing and languages. It is expected that these speech audiometric tests will serve as an effective clinical tool for improving the quality of speech audiometry testing in Vietnam.

## 11. APPENDIX

### Appendix A: Frequency of occurrence of phonemes in Vietnamese

#### 1. The consonant phoneme frequency in the SVN

Consonant phoneme	Written text		Spoken text	
	Frequency	Percent	Frequency	Percent
/-ŋ/	26418	11.60	20473	10.80
/k/	17589	7.72	15150	7.99
/-n/	17387	7.64	14975	7.90
/d/	13054	5.73	13460	7.10
/-k/	11647	5.12	6909	3.64
/t/	11394	5.00	7588	4.00
/t <sup>h</sup> /	9776	4.29	7844	4.14
/l/	9756	4.28	10210	5.38
/v/	9184	4.03	7289	3.84
/m/	8818	3.87	8357	4.41
/h/	8793	3.86	6500	3.43
/-t/	8691	3.82	6512	3.43
/b/	7711	3.39	5545	2.92
/c/	7481	3.29	7092	3.74
/ɲ/	7213	3.17	4959	2.62
/-m/	7201	3.16	5047	2.66
/z/	6670	2.93	5830	3.07
/n/	5889	2.59	9201	4.85
/ɣ/	5661	2.49	4470	2.36
/ŋ/	5590	2.46	3480	1.84
/ʃ/	5458	2.40	4026	2.12
/t̚/	4977	2.19	3553	1.87
/z̚/	3158	1.39	3211	1.69
/-p/	2772	1.22	2014	1.06
/f/	2687	1.18	2426	1.28
/s/	1591	0.70	2213	1.17
/ʎ/	1125	0.49	1269	0.67
Total	227691	100.00	189603	100

## 2. The vowel phoneme frequencies in the SVN

Vowel phoneme	Written text		Spoken text	
	Frequency	Percentage	Frequency	Percentage
/a/	31516	16.55	25300	15.05
/ɔ/	14812	7.78	16786	9.99
/i/	14690	7.71	12695	7.55
/o/	14258	7.49	14424	8.58
/ə̃/	11087	5.82	10885	6.48
/i_ə/	10642	5.59	8908	5.30
/ɛ/	9334	4.90	5377	3.20
/ä/	9106	4.78	6107	3.63
/u/	8780	4.61	6903	4.11
/u_ə/	7796	4.09	6616	3.94
/e/	7622	4.00	5926	3.53
/ə/	7267	3.82	6708	3.99
/u/	6452	3.39	6721	4.00
/u_ə/	3975	2.09	2773	1.65
/-j/	23297	12.23	23829	14.18
/-w/	9811	5.15	8112	4.83
<b>Total</b>	<b>190445</b>	<b>100</b>	<b>168070</b>	<b>100</b>

## 3. Tonal distribution frequencies in the SVN, spoken text

Tone	Frequency	Percentage
<i>Ngang</i> (A1)	16050	29.17
<i>Huyền</i> (A2)	12456	22.64
<i>Sắc</i> (B1)	12653	23.00
<i>Nặng</i> (B2)	6841	12.43
<i>Hỏi</i> (C1-C2)	7021	12.76
Total	55021	100

**Appendix B:** Sublists of NAMES test with nonsense two syllables structures

Test No	List A11		List A22		List B11		List B22	
	Word	Transcription	Word	Transcription	Word	Transcription	Word	Transcription
1	<i>tá lòi</i>	/ta5 lɔj2/	<i>lò tái</i>	/lɔ2 taj5/	<i>đi thát</i>	/di1 t'ət5/	<i>thi đát</i>	/t'i1 dət5/
2	<i>thì táng</i>	/t'i2 tǎŋ1/	<i>thì táng</i>	/t'i2 tǎŋ1/	<i>đà căm</i>	/da2 kəm1/	<i>cà đăm</i>	/ka2 dăm1/
3	<i>đi háam</i>	/di2 ham5/	<i>đi háam</i>	/di2 ham5/	<i>tó căc</i>	/tɔ5 kak5/	<i>tá cóc</i>	/ta5 kɔk5/
4	<i>vì băng</i>	/zi2 băŋ1/	<i>bì vǎng</i>	/bi2 zǎŋ/	<i>tố bát</i>	/to5 bat5/	<i>tá bốt</i>	/ta5 bot5/
5	<i>mi thát</i>	/mi1 t'ət5/	<i>mi thát</i>	/mi1 t'ət6/	<i>la hồng</i>	/la1 hoŋ2/	<i>hồ lang</i>	/ho2 laŋ1/
6	<i>li vǎng</i>	/li1 zǎŋ 2/	<i>li vǎng</i>	/li1 zǎŋ2/	<i>đi thóc</i>	/di2 t'ɔk5/	<i>thì đóc</i>	/t'i2 dɔk5/
7	<i>mồ cói</i>	/mo2 cɔj5/	<i>cồ mới</i>	/ko2 mɔj5/	<i>lô căm</i>	/lɔ5 kəm1/	<i>lô căm</i>	/lɔ5 kəm1/
8	<i>lô căng</i>	/lo1 kaŋ2/	<i>cà lông</i>	/ka2 loŋ1/	<i>mà lát</i>	/ma2 lət5/	<i>mà lát</i>	/ma2 lət5/
9	<i>đô bóc</i>	/do1 hɔk5/	<i>hồ đóc</i>	/ho1 dɔk5/	<i>va đǎng</i>	/za1 dǎŋ1/	<i>đà vǎng</i>	/da1 zǎŋ1/
10	<i>là bắc</i>	/la2 bək5/	<i>là bắc</i>	/la2 bək5/	<i>mô lái</i>	/mo1 laj2/	<i>lô mài</i>	/lo1 maj2/
11	<i>bó đóc</i>	/hɔ5 dɔk5/	<i>bó đóc</i>	/hɔ5 dɔk5/	<i>đi moi</i>	/di2 mɔj1/	<i>mì đôi</i>	/mi2 dɔj1/
12	<i>đà cai</i>	/da2 kaj1/	<i>ca đài</i>	/ka1 daj2/	<i>thì hồng</i>	/t'i2 hoŋ1/	<i>thì hồng</i>	/t'i2 hoŋ1/
13	<i>la mắm</i>	/la1 məm1/	<i>ma lắm</i>	/ma1 ləm1/	<i>bá thài</i>	/ha1 t'aj2/	<i>thá bài</i>	/t'a5 haj2/
14	<i>thà bói</i>	/t'a2 boj1/	<i>bô thài</i>	/bo2 t'aj2/	<i>đà cắc</i>	/da2 kak1/	<i>cà đắc</i>	/ka2 dək5/
15	<i>đó cắt</i>	/do5 kət5/	<i>đó cắt</i>	/do5 kət5/	<i>bì lóc</i>	/bi1 lɔk5/	<i>lì bóc</i>	/li1 bɔk5/
16	<i>tho vài</i>	/t'ɔ1 zaj2/	<i>và thoi</i>	/za2 t'ɔj1/	<i>ho vít</i>	/hɔ1 zit5/	<i>vo hít</i>	/zɔ1 hit5/
17	<i>vô đǎng</i>	/zo1 dǎŋ2/	<i>đô vǎng</i>	/do1 zǎŋ2/	<i>đô vôi</i>	/do1 zɔj2/	<i>đô vôi</i>	/dɔ2 zɔj1/
18	<i>tì đác</i>	/ti2 dak5/	<i>đi tác</i>	/di2 tak5/	<i>bo vǎng</i>	/bɔ1 zǎŋ2/	<i>vo bǎng</i>	/zɔ1 băŋ2/
19	<i>va bít</i>	/za1 bit5/	<i>va bít</i>	/za1 bit5/	<i>bì thai</i>	/bi2 t'aj1/	<i>thì bai</i>	/t'i2 baj1/
20	<i>ho thóc</i>	/hɔ1 t'ɔk5/	<i>tho bóc</i>	/hɔ1 t'ɔk5/	<i>vô tǎng</i>	/zɔ5 tǎŋ2/	<i>tó vǎng</i>	/tɔ5 zǎŋ2/

**Appendix C:** Descriptive statistics of normative values for native listeners

1. Descriptive statistics of the normative values in quiet across age groups. All values are calculated in dB SPL.

		Four year-olds	Six year-olds	Eight year-olds
<b>Children</b>	n (subjects/ears)	24 / 44	29 / 51	21 / 37
	Mean	37.2	31.8	31.0
	SD (SE)	5.12 (0.77)	2.51 (0.35)	2.54 (0.42)
	Median	35.8	32	30.7
	Min	27.5	24.4	23.7
	Max	47.5	35.7	34.5
		<b>15 to 20</b>	<b>21 to 30</b>	<b>31 to 40</b>
<b>Youths &amp; Adults</b>	n (subjects/ears)	22 / 42	24 / 45	20 / 38
	Mean	29.9	29.4	30.6
	SD (SE)	2.45 (0.38)	3.41 (0.51)	2.92 (0.47)
	Median	30.2	29.7	30.7
	Min	24.4	21.9	25.7
	Max	34.5	34.5	35.7
		<b>55 to 65</b>	<b>66 to 75</b>	<b>76 to 85</b>
<b>Older Adults</b>	n (subjects/ears)	20 / 26	20 / 35	20 / 28
	Mean	35.8	40.4	40.6
	SD (SE)	3.31 (0.65)	3.18 (0.54)	3.79 (0.72)
	Median	34.7	40.7	40.8
	Min	29.5	33.2	33.5
	Max	40.7	45.7	47

2. Descriptive statistics of the normative values in noise across age groups. All values are calculated in dB SNR.

		Four year-olds	Six year-olds	Eight year-olds
<b>Children</b>	n (subjects/ears)	24 / 43	29 / 55	21 / 39
	Mean	-9.1	-11.7	-12.8
	SD/SE	1.89 / 0.29	2.22 / 0.3	2.63 / 0.42
	Median	-9.5	-11.5	-13
	Min	-12.5	-16.8	-16.8
	Max	-4.5	-7	-7.8
		<b>15 to 20</b>	<b>21 to 30</b>	<b>31 to 40</b>
<b>Youths &amp; Adults</b>	n (subjects/ears)	22 / 44	24 / 47	20 / 39
	Mean	-13.9	-14.1	-13.8
	SD / SE	2.26 / 0.34	2.45 / 0.36	2.30 / 0.37
	Median	-13.8	-14.5	-14.5
	Min	-17.5	-17.5	-17.5
	Max	-9.3	-7.9	-7.8
		<b>55 to 65</b>	<b>66 to 75</b>	<b>76 to 85</b>
<b>Older Adults</b>	n (subjects/ears)	20 / 35	20 / 37	20 / 29
	Mean	-9.6	-6.5	-6.4
	SD / SE	2.66 / 0.45	2.10 / 0.34	2.50 / 0.45
	Median	-10	-5.5	-5.5
	Min	-16	-12.3	-10.5
	Max	-4.8	-3.3	-2.5

**Appendix D:** Statistical values of AAST-a2 for non-native listeners

1. Descriptive statistics of AAST values in quiet across dialectal groups. All values are calculated in dB SPL.

Groups		North	Central	South
<b>Children</b>	n (subjects/ears)	20 / 36	19 / 36	27 / 47
	Mean	34.4	38.5	32.2
	SD	4.0 / 0.74	2.95 / 0.49	2.02 / 0.29
	Median	34.5	39.5	32
	Min	24.7	32	28.2
	Max	42	44.5	35.7
<b>Adult</b>	n (subjects/ears)	21 / 40	21 / 35	24 / 43
	Mean	33.4	33.6	29.7
	SD	4.63 / 0.73	4.09 / 0.69	3.1 / 0.47
	Median	34.5	33.2	30.7
	Min	25.7	24.4	23.2
	Max	40.7	40.7	34.5
<b>Older Adult</b>	n (subjects/ears)	18/26	18/29	17/30
	Mean	40.6	38.8	37.1
	SD	5.71 / 1.12	3.61 / 0.6	3.54 / 0.65
	Median	41.4	38.2	37.3
	Min	30.7	33.2	29.5
	Max	48.7	45.7	43.2

2. Descriptive statistics of AAST values in noise across dialectal groups. All values are calculated in dB SNR

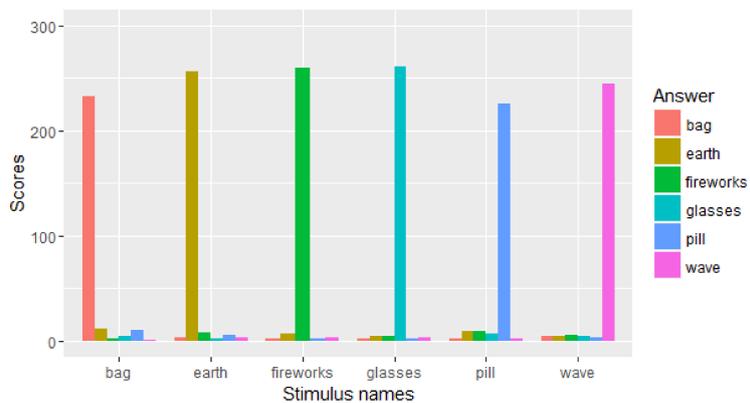
Groups		North	Central	South
<b>Children</b>	n (subjects/ears)	20 / 34	19 / 37	27 / 54
	Mean	-8.2	-6.1	-11.7
	SD	2.73 / 0.47	1.54 / 0.25	2.24 / 0.3
	Median	-8.2	-6.1	-11.5
	Min	-15.3	-9.3	-16.8
	Max	-3.3	-3.3	-7
<b>Adult</b>	n (subjects/ears)	20 / 40	19 / 39	24 / 46
	Mean	-8.8	-10.5	-14.3
	SD	2.64 / 0.42	3.39 / 0.54	2.3 / 0.34
	Median	-8.5	-10	-14.5
	Min	-15.5	-16	-17.5
	Max	-4	-4	-8.5
<b>Older Adult</b>	n (subjects/ears)	18/28	18/29	17/32
	Mean	-6.3	-7.4	-8.3
	SD	2.7 / 0.51	2.3 / 0.43	2.2 / 0.39
	Median	-6.3	-7.8	-8.2
	Min	-10.9	-11.5	-12.3
	Max	-2.5	-2.5	-4.5

**Appendix E:** Descriptive statistical values of SRT across subtests of AAST

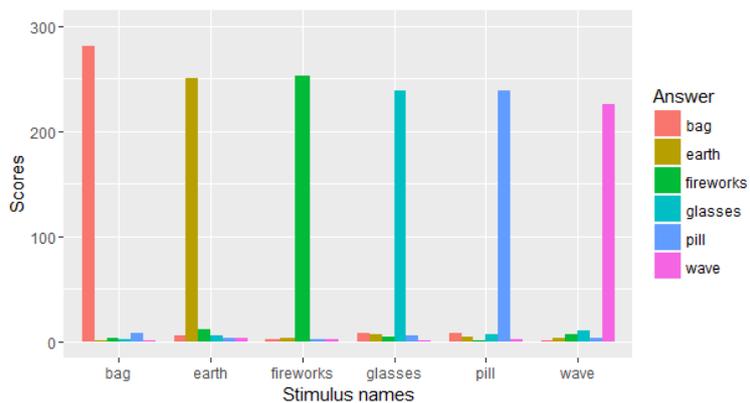
Conditions		<b>a1</b>	<b>a2</b>	<b>a3</b>	<b>a4</b>	<b>aTP</b>
<b>Quiet</b>	n (subjects/ears)	21 / 36	19 / 32	23 / 39	23 / 43	22 / 39
	Mean (dB SPL)	37.4	37.7	37.7	36.1	37.5
	SD / SE	4.22 / 0.7	4.01 / 0.71	4.5 / 0.72	4.38 / 0.67	4.79 / 0.77
	Median	35.8	38.2	38.4	35.8	37.5
	Min	30.7	29.5	29.2	26.2	30.5
	Max	45.7	45.7	45.8	45.5	47.0
<b>Noise</b>	n (subjects/ears)	21 / 37	19 / 36	23 / 30	23 / 41	22 / 36
	Mean (dB SNR)	-8.3	-7.7	-4.4	-8.3	-7.4
	SD / SE	3.04 / 0.5	2.71 / 0.45	1.69 / 0.31	3.17 / 0.5	3.4 / 0.57
	Median	-8.5	-7.8	-4.5	-8.5	-7.0
	Min	-15.5	-14.5	-8.5	-14.5	-13.5
	Max	-3.5	-2.5	-1.9	-3.3	-2.5

**Appendix F:** Word confusion matrix of native listeners in AAST-a2

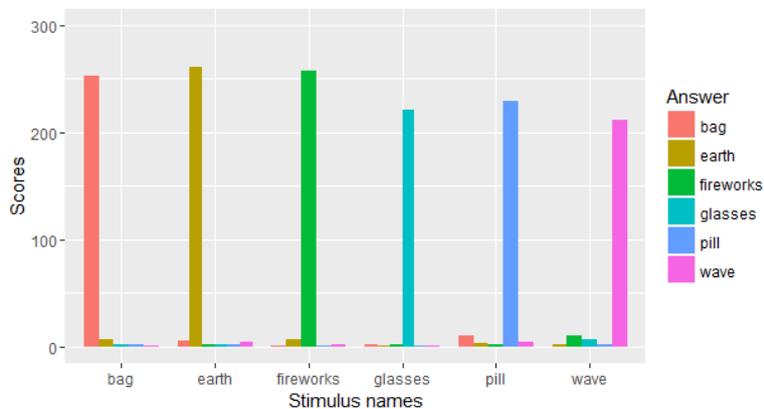
1. Word confusion matrix by the children



2. Word confusion matrix by the young and adults

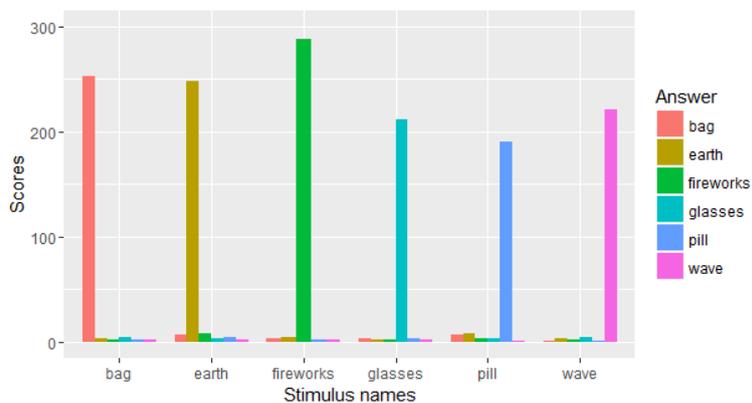


3. Word confusion matrix by the older adults

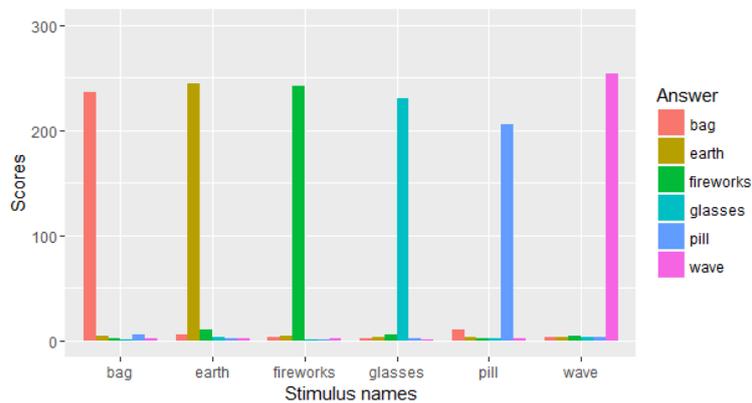


**Appendix G:** Word confusion matrix of non-native listeners in AAST-a2

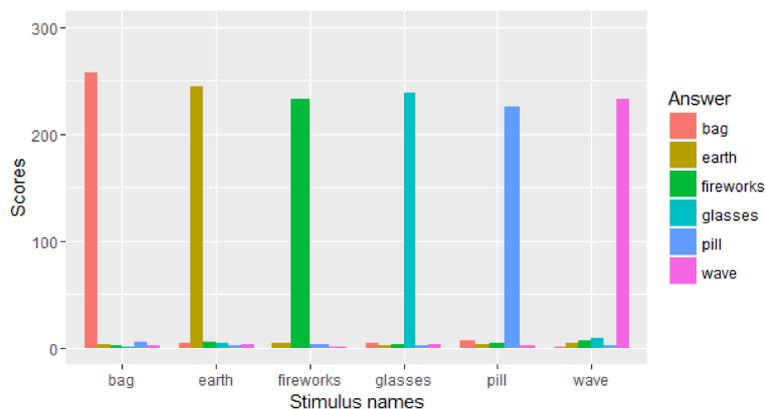
1. Word confusion matrix of the northern listeners



2. Word confusion matrix of the central listeners



3. Word confusion matrix of the southern listeners



**Appendix H: Error rates and response matrix of native listeners for NAMES**

## 1. Error rates and response matrix taken from the youth and the adults

Response												
Onset	wrong	/k/	/d/	/t/	/l/	/j/	/m/	/h/	/b/	/t'/	number	% error
/k/	12	556	0	0	0	0	0	0	0	0	12/568	2.1
/d/	2	0	983	2	0	0	0	0	7	0	9/994	1.1
/t/	1	2	0	423	0	0	0	0	0	0	3/426	0.7
/l/	1	0	0	0	709	0	0	0	0	0	1/710	0.1
/j/	12	0	0	0	0	698	0	0	0	0	12/710	1.7
/m/	0	1	0	0	0	0	423	0	2	0	3/426	0.7
/h/	9	0	0	0	0	0	0	557	0	2	11/568	1.9
/b/	0	0	1	0	0	0	1	0	566	0	2/568	0.4
/t'/	9	0	0	0	0	0	0	1	0	700	10/710	1.4

Response										
Nucleus	wrong	/a/	/o/	/ɔ/	/i/	/ǐ/	/ǎ/	number	% error	
/a/	8	1554	0	0	0	0	0	8/1562	0.5	
/o/	0	0	773	75	0	4	0	79/852	9.3	
/ɔ/	12	1	4	1113	0	6	0	23/1136	2.0	
/i/	21	0	0	0	973	0	0	21/994	2.1	
/ǐ/	26	1	0	0	0	628	55	82/710	11.5	
/ǎ/	3	0	0	0	0	3	420	6/426	1.4	

Response							
Tone	wrong	B1	A2	A1	number	% error	
B1	30	1811	0	5	35/1846	1.9	
A2	65	2	1778	1	68/1846	3.7	
A1	16	0	1	1971	17/1988	0.9	

Response								
Coda	wrong	/k/	/ŋ/	/j/	/t/	/m/	number	% error
/k/	1	701	0	0	8	0	9/710	1.3
/ŋ/	17	0	693	0	0	0	17/710	2.4
/j/	0	0	0	710	0	0	0/710	0.0
/t/	1	27	0	0	398	0	28/426	6.6
/m/	20	0	1	0	0	263	21/284	7.4

## 3. Error rates and response matrix taken from the older adults

Response												
Onset	wrong	/k/	/d/	/t/	/l/	/j/	/m/	/h/	/b/	/tʰ/	number	% error
/k/	33	<b>479</b>	0	0	0	0	0	0	0	0	33/512	6.4
/d/	50	0	<b>840</b>	0	0	0	0	0	5	1	56/896	6.3
/t/	10	1	0	<b>372</b>	0	1	0	0	0	0	12/384	3.1
/l/	5	0	0	0	<b>635</b>	0	0	0	0	0	5/640	0.8
/j/	9	0	0	0	0	<b>631</b>	0	0	0	0	9/640	1.4
/m/	30	0	0	0	0	0	<b>350</b>	0	4	0	34/384	8.9
/h/	<b>170</b>	1	0	0	0	0	0	<b>329</b>	0	12	183/512	35.7
/b/	10	0	1	0	0	0	1	0	<b>500</b>	0	12/512	2.3
/tʰ/	7	0	0	0	0	0	0	0	0	<b>633</b>	7/640	1.1

Response										
Nucleus	wrong	/a/	/o/	/ɔ/	/i/	/ǐ/	/ǎ/	number	% error	
/a	16	<b>1392</b>	0	0	0	0	0	16/1408	1.1	
/o/	10	0	<b>742</b>	<b>13</b>	0	3	0	26/768	3.4	
/ɔ/	17	0	1	<b>1002</b>	0	3	1	22/1024	2.1	
/i/	16	0	0	0	<b>880</b>	0	0	16/896	1.8	
/ǐ/	18	0	0	0	0	<b>615</b>	<b>7</b>	25/640	3.9	
/ǎ/	8	0	0	0	0	0	<b>376</b>	8/384	2.1	

Tone	wrong	B1	A2	A1	number	% error
<b>B1</b>	42	<b>1622</b>	0	0	42/1664	2.5
<b>A2</b>	41	0	<b>1623</b>	0	41/1664	2.5
<b>A1</b>	14	0	1	<b>1777</b>	15/1792	0.8

Response								
Coda	wrong	/k/	/ŋ/	/j/	/t/	/m/	number	% error
/k/	6	<b>631</b>	0	0	3	0	9/640	1.4
/ŋ/	24	0	<b>616</b>	0	0	0	24/640	3.8
/j/	7	0	0	<b>633</b>	0	0	7/640	1.1
/t/	4	<b>4</b>	0	0	<b>376</b>	0	8/384	2.1
/m/	14	0	0	0	0	<b>242</b>	14/256	5.5

## Appendix I: Error rates and response matrix across dialectal groups

## 1. Error rates and response matrix of the Northern listeners

## Response

Onset	wrong	/k/	/d/	/t/	/l/	/j/	/m/	/h/	/b/	/t'/	number	% error
/k/	0	<b>160</b>	0	0	0	0	0	0	0	0	160/160	0
/d/	1	0	<b>279</b>	0	0	0	0	0	0	0	1/280	0.4
/t/	2	0	0	<b>118</b>	0	0	0	0	0	0	2/120	1.7
/l/	3	0	0	0	<b>197</b>	0	0	0	0	0	3/200	1.5
/j/	19	0	0	0	0	<b>181</b>	0	0	0	0	19/200	9.5
/m/	2	0	0	0	0	0	<b>118</b>	0	0	0	2/120	1.7
/h/	16	0	0	0	0	0	0	<b>144</b>	0	0	16/160	10.0
/b/	1	0	0	0	0	0	0	0	<b>159</b>	0	1/160	0.6
/t'/	1	0	0	0	0	0	0	0	0	<b>199</b>	1/200	0.5

## Response

Nucleus	wrong	/a/	/o/	/ɔ/	/i/	/ě/	/ǎ/	Number	% error
/a/	1	<b>439</b>	0	0	0	0	0	1/440	0.2
/o/	5	0	<b>235</b>	0	0	0	0	5/240	2.1
/ɔ/	18	0	0	<b>302</b>	0	0	0	18/320	5.6
/i/	30	0	0	0	250	0	0	50/280	10.7
/ě/	10	0	0	0	0	<b>190</b>	0	10/200	5
/ǎ/	0	0	0	0	0	0	<b>120</b>	0/120	0

## Response

Tone	wrong	B1	A2	A1	number	% error
<b>B1</b>	23	<b>497</b>	0	0	23/520	4.4
<b>A2</b>	33	0	<b>487</b>	0	33/520	6.3
<b>A1</b>	8	0	0	<b>552</b>	8/560	1.4

## Response

Coda	wrong	/k/	/ŋ/	/j/	/t/	/m/	number	% error
/k/	4	<b>196</b>	0	0	0	0	4/200	2
/ŋ/	38	0	<b>162</b>	0	0	0	38/200	19
/j/	1	0	0	<b>199</b>	0	0	1/200	0.5
/t/	4	0	0	0	<b>116</b>	0	4/120	3.3
/m/	10	0	0	0	0	70	10/80	12.5

## 2. Error rates and response matrix of the Central listeners

## Response

Onset	wrong	/k/	/d/	/t/	/l/	/j/	/m/	/h/	/b/	/t'/	number	% error
/k/	4	<b>164</b>	0	0	0	0	0	0	0	0	4/170	2.4
/d/	0	0	<b>294</b>	0	0	0	0	0	0	0	0/294	0.0
/t/	2	0	0	<b>124</b>	0	0	0	0	0	0	2/126	1.6
/l/	1	0	0	0	<b>209</b>	0	0	0	0	0	1/210	0.5
/j/	22	0	0	0	0	<b>188</b>	0	0	0	0	22/210	10.5
/m/	5	0	0	0	0	0	<b>120</b>	0	1	0	6/126	4.0
/h/	4	1	0	0	0	0	0	<b>163</b>	0	0	5/168	3.0
/b/	0	0	0	0	0	0	0	0	<b>168</b>	0	0/168	0.0
/t'/	1	0	0	0	0	0	0	0	0	<b>209</b>	1/210	0.5

## Response

Nucleus	wrong	/a/	/o/	/ɔ/	/i/	/ǎ/	/ä/	number	% error
/a/	3	<b>459</b>	0	0	0	0	0	3/462	0.6
/o/	1	0	<b>247</b>	4	0	0	0	5/252	2.0
/ɔ/	3	0	7	<b>326</b>	0	0	0	10/336	3.0
/i/	23	0	0	0	<b>271</b>	0	0	23/294	7.4
/ǎ/	14	0	0	0	0	<b>195</b>	1	15/210	6.7
/ä/	0	0	0	0	0	0	<b>126</b>	0/126	0.0

## Response

Tone	wrong	B1	A2	A1	number	% error
<b>B1</b>	13	<b>533</b>	0	0	13/546	2.4
<b>A2</b>	31	0	<b>515</b>	0	31/546	5.7
<b>A1</b>	5	0	0	<b>583</b>	5/583	0.9

## Response

Coda	wrong	/k/	/ŋ/	/j/	/t/	/m/	number	%error
/k/	0	<b>209</b>	0	0	1	0	1/210	0.5
/ŋ/	58	0	<b>152</b>	0	0	0	58/210	27.6
/j/	0	0	0	<b>210</b>	0	0	0/210	0.0
/t/	6	7	0	0	<b>113</b>	0	13/126	5.8
/m/	4	0	0	0	0	<b>80</b>	4/84	4.8

## 3. Error rates and response matrix by the Southern listeners

## Response

Onset	wrong	/k/	/d/	/t/	/l/	/j/	/m/	/h/	/b/	/t'/	number	% error
/k/	45	1035	0	0	0	0	0	0	0	0	45/1080	4.2
/d/	52	0	1823	2	0	0	0	0	<b>12</b>	1	67/1890	3.5
/t/	11	3	0	795	0	1	0	0	0	0	15/810	1.9
/l/	6	0	0	0	1344	0	0	0	0	0	6/1350	0.4
/j/	21	0	0	0	0	1329	0	0	0	0	21/1350	1.6
/m/	30	1	0	0	0	0	773	0	6	0	37/810	4.6
/h/	<b>179</b>	1	0	0	0	0	0	886	0	14	194/1080	18.0
/b/	10	0	2	0	0	0	2	0	1066	0	14/1080	1.3
/t'/	16	0	0	0	0	0	0	1	0	1333	17/1350	1.3

## Response

Nucleus	wrong	/a/	/o/	/ɔ/	/i/	/ǝ/	/ǎ/	number	error
/a/	24	2946	0	0	0	0	0	24/2970	0.8
/o/	10	0	1515	<b>88</b>	0	7	0	105/1620	6.5
/ɔ/	29	1	5	2115	0	9	1	45/2160	2.1
/i/	37	0	0	0	1853	0	0	37/1890	2.0
/ǝ/	44	1	0	0	0	1243	<b>62</b>	107/1350	7.9
/ǎ/	11	0	0	0	0	3	796	14/810	1.7

## Response

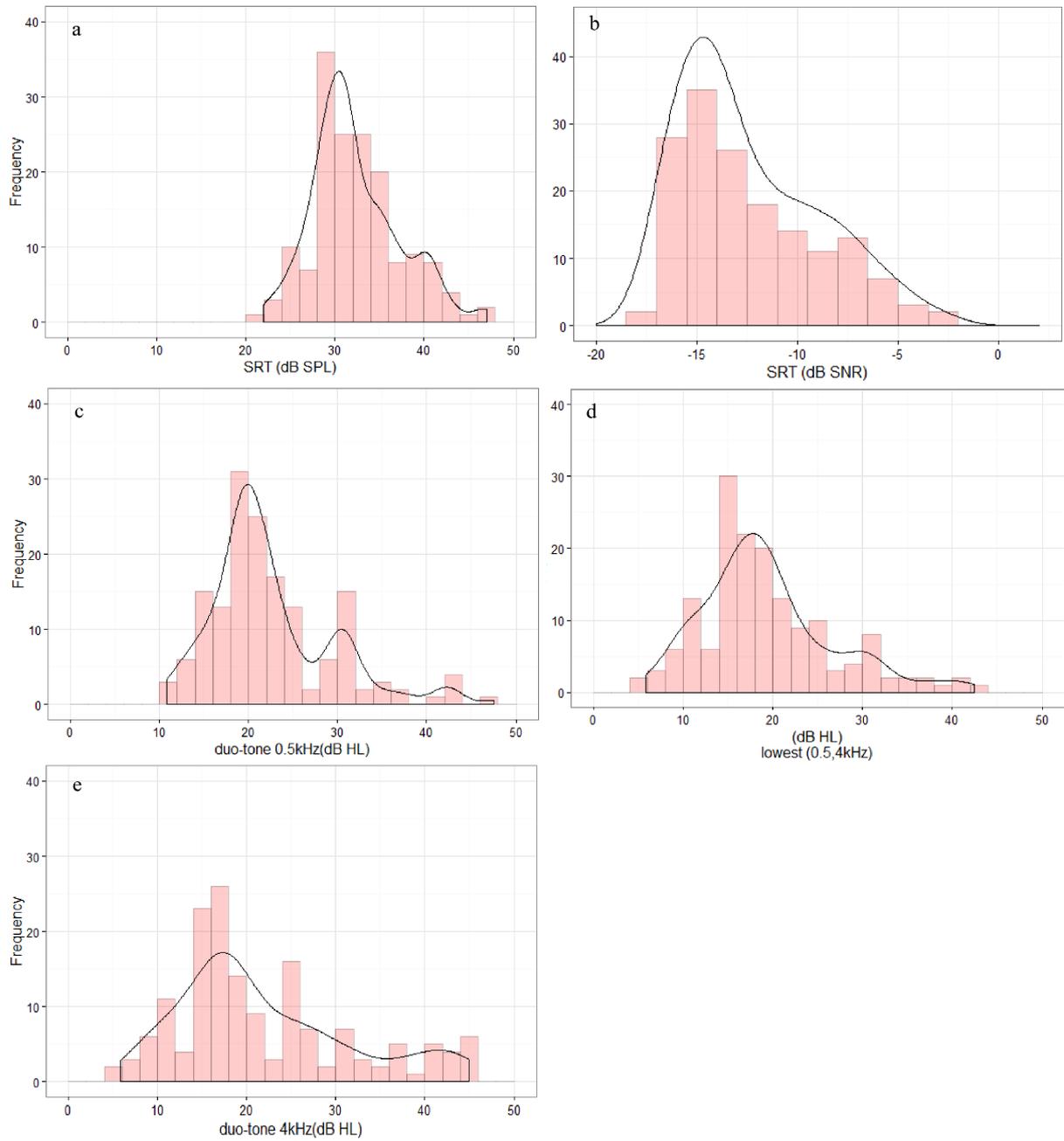
Tone	wrong	B1	A2	A1	number	% error
<b>B1</b>	72	3433	0	5	77/3510	2.2
<b>A2</b>	106	2	3401	1	109/3500	3.1
<b>A1</b>	30	0	2	3748	32/3780	0.8

## Response

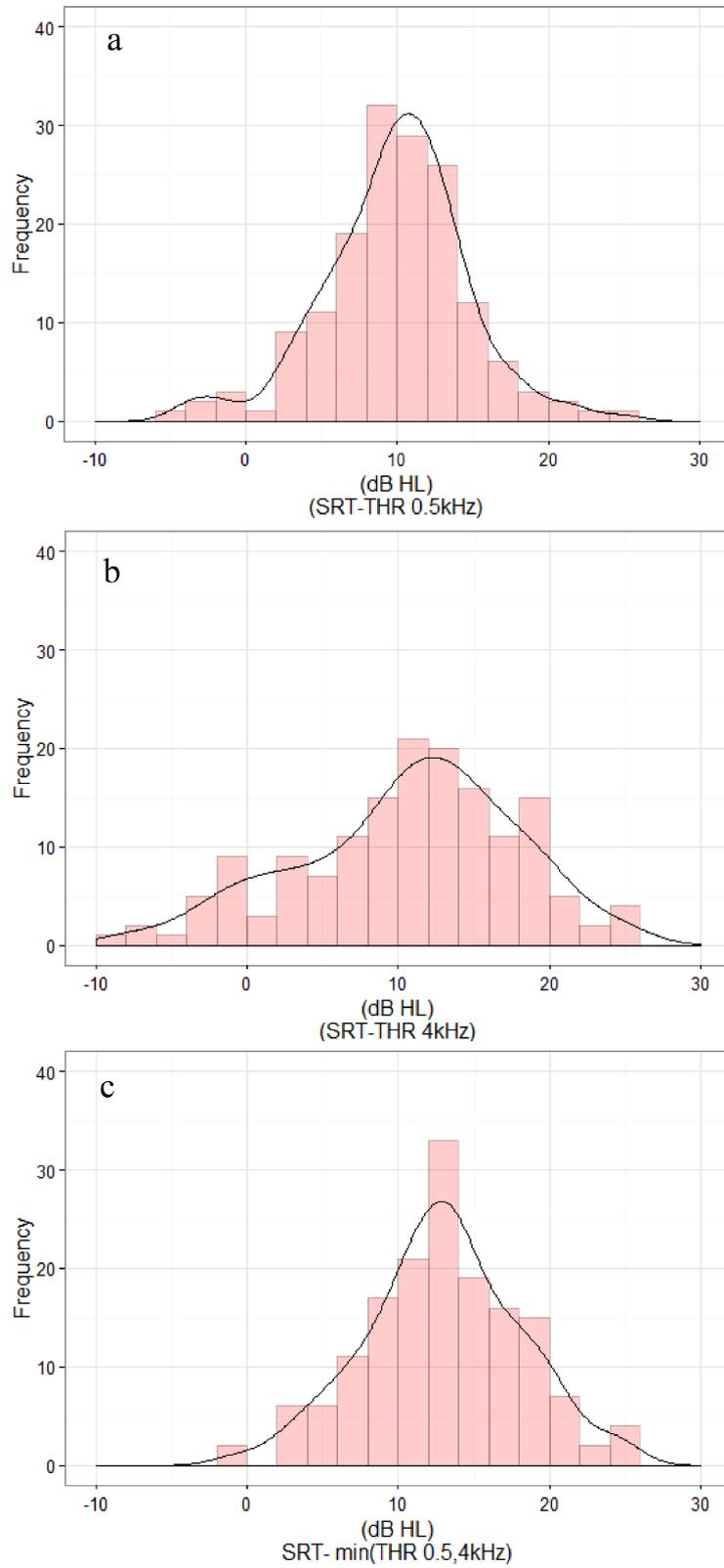
Coda	wrong	/k/	/ŋ/	/j/	/t/	/m/	number	% error
/k/	7	1332	0	0	<b>11</b>	0	18/1350	1.3
/ŋ/	41	0	1309	0	0	0	41/1350	3.0
/j/	7	0	0	1343	0	0	7/1350	0.5
/t/	5	<b>31</b>	0	0	774	0	36/810	4.4
/m/	34	0	1	0	0	505	35/540	6.5

**Appendix J:** Distribution of individual values in SRTs and duo-tone thresholds

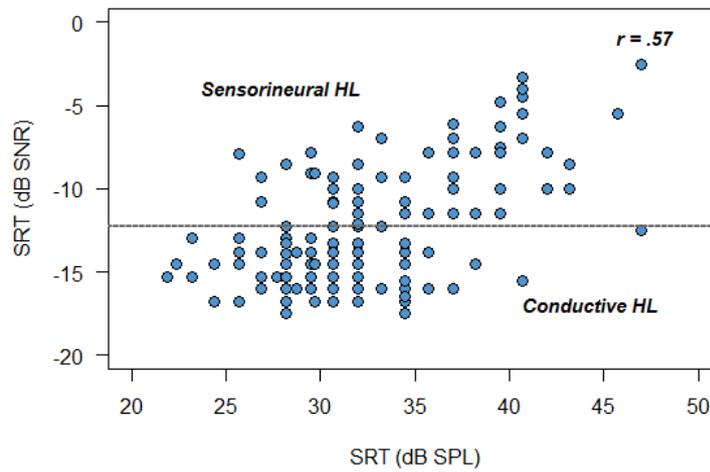
1. Distribution of individual values in correlation between speech thresholds and duo-tone thresholds



2. Distribution of individual values (speech thresholds minus duo-tone thresholds)

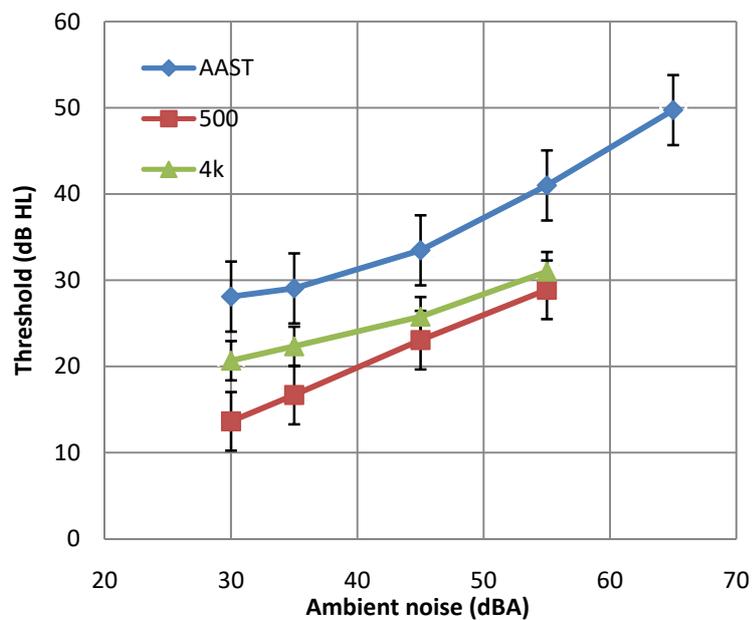


3. Correlation between SRTs quiet and SRTs noise



**Appendix K:** SRT values and duo-tone thresholds in different ambient noise levels

The exacerbation of duo-tone and speech threshold corresponds to enhancement of ambient noise level from a group of 22 German adult listeners. The speech thresholds are intact as ambient noises are between 30 and 35 dBA, whereas the speech threshold are explicitly elevated in the noise levels over 40 dBA. Collected at IfAP, April 2016



**Appendix L:** A comparison between AAST-Vietnamese and AAST-German

Descriptive statistical values of AAST in Vietnamese compare with AAST in German. AAST-a2 is used to collect data in Vietnam, while AAST-VN (new version of a2) is used in Germany. The results showed a congruity in the speech thresholds between AAST-DE and AAST-VN as the data of AAST-VN were gathered in Germany; otherwise, there was a small difference roughly 1.5 dB due to ambient noise in Vietnam. All values are calculated in dB SPL.

<b>Groups of Adult listeners</b>	AAST-DE	<b>AAST-VN</b>	AAST-a2
		(Germany)	(Vietnam)
n (ears)	19	35	45
Mean	27.7	28	29.4
SD (SE)	2.2 (0.52)	1.61(0.28)	3.41 (0.51)
Median	27.5	27.5	29.7
Min	22.5	23.8	21.9
Max	31.3	30	34.5

**Appendix M:** Children with the performances of AAST



## 12. REFERENCES

- Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1925–1928). Retrieved from [http://pubman.mpdl.mpg.de/pubman/item/escidoc:59662:2/component/escidoc:59663/Adank\\_2007\\_effect.pdf](http://pubman.mpdl.mpg.de/pubman/item/escidoc:59662:2/component/escidoc:59663/Adank_2007_effect.pdf)
- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47(sup2), S53–S71. <https://doi.org/10.1080/14992020802301142>
- Alves, M. (1995). Tonal Features and the Development of Vietnamese Tones Mark Alves. *Working Papers in Linguistics: Department of University of Hawaii at Manoa*, 27, 1–13.
- Alves, M. J. (2010). A Look at North-Central Vietnamese, (February), 100–104.
- Beattie, R. C., & Zipp, J. A. (1990). Range of intensities yielding PB Max and the threshold for monosyllabic words for hearing-impaired subjects. *The Journal of Speech and Hearing Disorders*, 55(3), 417–26. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2381183>
- Bergman, M. (1980). *Aging and the Perception of Speech*. Baltimore: University Park Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press. Retrieved from <http://www.haskins.yale.edu/Reprints/HL0996.pdf>
- Billings, C. J., Penman, T. M., McMillan, G. P., & Ellis, E. M. (2015). Electrophysiology and Perception of Speech in Noise in Older Listeners. *Ear and Hearing*, 36(6), 710–722. <https://doi.org/10.1097/AUD.0000000000000191>
- Blanchet, C., Pommie, C., Mondain, M., Berr, C., Hillaire, D., & Puel, J.-L. (2008). Pure-tone threshold description of an elderly French screened population. *Otology & Neurotology*, 29(4), 432–440. <https://doi.org/10.1097/MAO.0b013e3181719746>
- Blicher, D. L., Diehl, R. L., & Cohen, eslie B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37–49.
- Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. *Glott International*. Retrieved from <http://www.praat.org/>
- Boothroyd, A. (1968). Developments in Speech Audiometry. *International Audiology*, 7(3), 368–368. <https://doi.org/10.3109/05384916809074343>

- Boothroyd, A. (2008). The Performance/Intensity Function: An Underused Resource. *Ear and Hearing*, 29(4), 479–491. <https://doi.org/10.1097/AUD.0b013e318174f067>
- Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology: Official Organ of the International Society of Audiology*, 34(5), 260–84. <https://doi.org/10.3109/00206099509071918>
- Brewer, C., Carlin, M. F., Durrant, J. D., Frank, A., Givens, G. D., Gorga, M. P., & Kamara, C. (1988). Guidelines for determining threshold level for speech. *ASHA*, 30(3), 85–9. <https://doi.org/10.1044/policy.GL1988-00008>
- Brunelle, M., & Jannedy, S. (2013). The cross-dialectal perception of Vietnamese tones: Indexicality and convergence. In D. Hole & E. Löbel (Eds.), *The Linguistics of Vietnamese* (pp. 9–34). Berlin: Mouton de Gruyter.
- Brunelle, M. (2009). Northern and Southern Vietnamese Tone Coarticulation : a Comparative Case Study. *Journal of the Southeast Asian Linguistics Society (JSEALS)*, 1, 49–62.
- Brunelle, M. (2015). *Proceedings of ELM-2014 Volume 1*. (J. Cao, K. Mao, E. Cambria, Z. Man, & K.-A. Toh, Eds.), *JSEALS - Journal of the Southeast Asian Linguistics Society* (Vol. 3). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-14063-6>
- Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37(1), 79–96. <https://doi.org/10.1016/j.wocn.2008.09.003>
- Brunelle, M., & Jannedy, S. (2007). Social effects on the perception of Vietnamese tones. *Proceedings of the 16th International Congress of Phonetic Sciences*, (August), 1461–1464.
- Bùi, K. T. (2009). Tiếng Việt Sài Gòn - TP. Hồ Chí Minh là một cực quy tụ và lan tỏa của tiếng Việt toàn dân [Vietnamese of Saigon is a convergence and pervasion of Vietnamese]. *Sài Gòn University*, (1).
- Cao, X. H. (1998). Hai vấn đề âm vị học của phương ngữ Nam Bộ [Two phonological issues in the southern dialect]. *Ngôn Ngữ*, 1, 48–53.
- Cao, X. H. (1999). *Tiếng Việt: mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa* [Vietnamese: issues of phonology, syntax, and semantics]. Education Publisher.
- Cao, X. H., & Lê, M. T. (2005). Tiếng Sài Gòn và cách phát âm của các phát thanh viên HTV (The Saigon dialect and the pronunciation of announcers on HTV). In K. T. Nguyễn (Ed.), *Tiếp Xúc Ngôn Ngữ ở Việt Nam (Language contacts in Vietnam)* (pp. 153–226). Hồ Chí Minh: Nhà Xuất bản Khoa học Xã hội.
- Carhart, R. (1952). Speech audiometry in clinical evaluation. *Acta Oto-Laryngologica*, 41(1–2), 18–48. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14932955>

- Carhart, R. (1965). Problems in the Measurement of Speech Discrimination. *Archives of Otolaryngology - Head and Neck Surgery*, 82(3), 253–260. <https://doi.org/10.1001/archotol.1965.00760010255007>
- Carhart, R. (1951). Basic Principles of Speech Audiometry. *Acta Oto-Laryngologica*, 40(1–2), 62–71. <https://doi.org/10.3109/00016485109138908>
- Cervera, T. C., Soler, M. J., Dasi, C., & Ruiz, J. C. (2009). Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 63(3), 216–226. <https://doi.org/10.1037/a0014321>
- Chan, K. K., Li, C., Ma, E. M., Yiu, E. L., & McPherson, B. (2015). Noise levels in an urban Asian school environment. *Noise and Health*, 17(74), 48. <https://doi.org/10.4103/1463-1741.149580>
- Coninx, F. (2008). Hörscreening bei Kindern im Alter von 4–6 Jahren mit dem Adaptiven Auditiven SprachTest AAST. 25. *Wissenschaftliche Jahrestagung Der Deutschen Gesellschaft Für Phoniatrie Und Pädaudiologie E. V. Kongress Der Union Europäischer Phoniater*. Göttingen: Dt. Ges. Für Phoniatrie Und Pädaudiologie.
- Coninx, F. (2006). Development and testing Adaptive Auditory Speech Tests (AAST). In *Kongress der Deutsche Gesellschaft für Audiologie(DGA). 9th year 2005*. Köln.
- Coninx, F. (2005). Konstruktion und Normierung des Adaptiven Auditiven Sprach-Test (AAST). In *22. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie, Berlin*. Retrieved from <http://www.egms.de/static/de/meetings/dgpp2005/05dgpp045.shtml>
- Coninx, F. (2016). *AAST und weitere kindgerechte Hörtestverfahren*. 19. *Jahrestagung der Deutschen Gesellschaft für Audiologie e.V.* Hannover. Retrieved from [http://www.auritec.de/fileadmin/user\\_upload/Coninx\\_CIC\\_Hannover\\_20160309.pdf](http://www.auritec.de/fileadmin/user_upload/Coninx_CIC_Hannover_20160309.pdf)
- Coninx, F., Senderski, A., Kochanek, K., Lorens, A., & Skarzynski, H. (2009). Screening with AAST in 6-7 year old children in elementary schools in Poland. In *The IX European Federation of Audiology Societies (EFAS) Congress* IX European Federation of Audiology Societies (EFAS) Congress. Spain. Retrieved from <http://auditio.com/congresos/efas2009/programe/abstract/164.htm>
- Cooke, M., Lecumberri, M. L. G., Scharenborg, O., & Van Dommelen, W. A. (2010). Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages. *Speech Communication*, 52(11–12), 954–967. <https://doi.org/10.1016/j.specom.2010.04.004>
- Craik, F. I. M. (n.d.). The role of cognition in age-related hearing loss. *Journal of the American Academy of Audiology*, 18(7), 539–47. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18236642>
- Crews, M. (1990). *Effects of regional dialect on the word recognition scores of children using the phonetically balanced kindergarten test and the nonsense syllable test*. Master thesis. The University of Montana. Retrieved from <http://scholarworks.umt.edu/etd/7795/>

- Dietz, A. (2015). *The Development of the Finnish Matrix Sentence Test*. Oldenburg: Verlag der Carl von Ossietzky Universität Oldenburg.
- Divenyi, P. L., Stark, P. B., & Haupt, K. M. (2006). Decline of speech understanding and auditory thresholds in the elderly. *J Acoust Soc Am*, *118*(2), 1089–1100.
- Divenyi, P. L., Stark, P. B., & Haupt, K. M. (2005). Decline of speech understanding and auditory thresholds in the elderly. *The Journal of the Acoustical Society of America*, *118*(2), 1089. <https://doi.org/10.1121/1.1953207>
- Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America*, *115*(1), 352. <https://doi.org/10.1121/1.1635413>
- Fortunato, S., Forli, F., Guglielmi, V., De Corso, E., Paludetti, G., Berrettini, S., & Fetoni, A. R. (2016). A review of new insights on the association between hearing loss and cognitive decline in ageing. *Acta Otorhinolaryngologica Italica : Organo Ufficiale Della Società Italiana Di Otorinolaringologia E Chirurgia Cervico-Facciale*, *36*(3), 155–66. <https://doi.org/10.14639/0392-100X-993>
- Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42* Suppl 1, S77-85. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12918613>
- Gelfand, S. A., Silman, S., & Piper, N. (1987). Consonant recognition in quiet and in noise with aging among normal hearing listeners. *The Journal of the Acoustical Society of America*, *80*(6), 1589–98. <https://doi.org/10.1121/1.394323>
- Goldstein, E. B. (2009). *Sensation and Perception. Perception* (8th ed.). Pacific Grove: Wadsworth Inc. Retrieved from <http://www.amazon.com/dp/0495601497>
- Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *The Journal of Rehabilitation Research and Development*, *42*(4s), 9. <https://doi.org/10.1682/JRRD.2005.01.0006>
- Hanson, C. (2014). *Development of Speech Recognition Threshold and Word Recognition Materials for Native Vietnamese Speakers*. Brigham Young University - Provo. Retrieved from <http://scholarsarchive.byu.edu/etd>
- Hart, L. A. (2008). Development of Thai Speech Audiometry Materials for Measuring Speech Recognition Thresholds. *Master Thesis*. Retrieved from <http://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2514&context=etd>
- Hazan, V., & Simpson, A. (n.d.). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Language and Speech*, *43*(Pt 3), 273–94. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11216296>

- Helzner, E. P., Cauley, J. A., Pratt, S. R., Wisniewski, S. R., Zmuda, J. M., Talbott, E. O., ... Newman, A. B. (2005). Race and Sex Differences in Age-Related Hearing Loss: The Health, Aging and Body Composition Study. *Journal of the American Geriatrics Society*, 53(12), 2119–2127. <https://doi.org/10.1111/j.1532-5415.2005.00525.x>
- Hoàng, T. C. (2009). *Phương ngữ học tiếng Việt [Vietnamese Dialectology]*. Hà Nội, Vietnam: Đại học Quốc gia Hà Nội.
- Hull, R. H., Saunders, G., & Martin, D. L. (2012). *Hearing and Aging* (San Diego). Plural Publishing.
- Huỳnh, C. T. (2014). *Tiếng Sài Gòn [The speech of Saigon]*. Nhà Xuất bản Chính trị Quốc gia.
- Huỳnh, C. T. (2007). *Từ điển từ ngữ Nam Bộ [The Dictionary of Southern Vietnamese]*. Publisher of Social Science.
- Huỳnh, C. T. (1999). *Hệ thống ngữ âm của phương ngữ Sài Gòn - so với phương ngữ Hà Nội và một số phương ngữ khác ở Việt Nam [The phonetic system of the Saigon dialect, compared with the Hanoi and other dialects of Vietnamese]*. Trường Đại học Khoa học Xã hội và Nhân văn Tp. Hồ Chí Minh. Việt Nam.
- Hwa-Froelich, D., Hodson, B. W., & Edwards, H. T. (2002). Characteristics of Vietnamese Phonology. *American Journal of Speech-Language Pathology*, 11(3), 264. [https://doi.org/10.1044/1058-0360\(2002/031\)](https://doi.org/10.1044/1058-0360(2002/031))
- Jenny, M., & Sidwell, P. (2014). *The Handbook of Austroasiatic Languages (2 vols). The Handbook of Austroasiatic Languages* (Vol. 1). Brill. <https://doi.org/10.1163/9789004283572>
- Jenstad, L. (2001). Speech Perception and Older Adults: Implications for Amplification. *Time*, 57–70. Retrieved from [http://verve.phonak.com/de/com\\_2006proceedings\\_jenstad.pdf](http://verve.phonak.com/de/com_2006proceedings_jenstad.pdf)
- Jerger, J. (1972). Audiological findings in aging. *Advances in Otorhino-Laryngology*, 20, 115–124.
- Kirby, J. (2010). Dialect experience in Vietnamese tone perception. *The Journal of the Acoustical Society of America*, 127(6), 3749. <https://doi.org/10.1121/1.3327793>
- Kramer, S. (2008). *Audiology Science to Practice*. San Diego: Plural.
- Kuk, F., Lau, C.-C., Korhonen, P., Crose, B., Peeters, H., & Keenan, D. (2010). Development of the ORCA Nonsense Syllable Test. *Ear and Hearing*, 31(6), 779–795. <https://doi.org/10.1097/AUD.0b013e3181e97bfb>
- Le, J. T., Best, C. T., Tyler, M. D., & Kroos, C. (2007). Effects of non-native dialects on spoken word recognition. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 2, pp. 1417–1420).
- Lê, T. H., Nguyễn, A. V., Vinh, T. H., Bùi, V. H., & Lê, D. (2011). A study on vietnamese prosody. *Studies in Computational Intelligence*, 351, 63–73. [https://doi.org/10.1007/978-3-642-19953-0\\_7](https://doi.org/10.1007/978-3-642-19953-0_7)

- Lee, F. S., Matthews, L. J. J., Dubno, J. R. R., & Mills, J. H. H. (2005). Longitudinal study of pure-tone thresholds in older persons. *Ear and Hearing, 26*(1), 1–11.  
<https://doi.org/10.1097/00003446-200502000-00001>
- Lee, J. Y. (2015). Aging and Speech Understanding. *Journal of Audiology & Otology, 19*(1), 7.  
<https://doi.org/10.7874/jao.2015.19.1.7>
- Leslie A. Guthrie, & L.Mackersie, C. (2009). A comparison of Presentation Levels to Maximize Word Recognition Scores, *20*(6), 381–390.
- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech, 47*(2), 109–138.  
<https://doi.org/10.1177/00238309040470020101>
- Lunner, T., & Sundewall-Thorén, E. (n.d.). Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology, 18*(7), 604–17. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18236647>
- Mackersie, C. L., Boothroyd, A., & Minniear, D. (2001). Evaluation of the Computer-assisted Speech Perception Assessment Test (CASPA). *Journal of the American Academy of Audiology, 12*(8), 390–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11599873>
- Madell, J. (2013). Audiology in Vietnam. Hearing Health & Technology Matters. Retrieved from <http://hearinghealthmatters.org/hearingandkids/2013/audiology-in-vietnam/>
- Mai, N. C., Vũ, Đ. N., & Hoàng, T. P. (1997). *Cơ sở ngôn ngữ học và tiếng Việt [Vietnamese linguistics]*. Hanoi: Education Publisher.
- Maniwa, K., Jongman, A., & Wade, T. (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *The Journal of the Acoustical Society of America, 123*(2), 1114.  
<https://doi.org/10.1121/1.2821966>
- Martin, F. N., & Clark, J. G. (2008). *Introduction to Audiology*. Allyn & Bacon.
- McCreery, R., Ito, R., Spratford, M., Lewis, D., Hoover, B., & Stelmachowicz, P. G. (2010). Performance-Intensity Functions for Normal-Hearing Adults and Children Using Computer-Aided Speech Perception Assessment. *Ear and Hearing, 31*(1), 95–101.  
<https://doi.org/10.1097/AUD.0b013e3181bc7702>
- Mendel, L. L., & Owen, S. R. (2011). A study of recorded versus live voice word recognition. *International Journal of Audiology, 50*(10), 688–693.  
<https://doi.org/10.3109/14992027.2011.588964>
- Meyer, B. T., Jürgens, T., Wesker, T., Brand, T., & Kollmeier, B. (2010). Human phoneme recognition depending on speech-intrinsic variability. *The Journal of the Acoustical Society of America, 128*(5), 3126. <https://doi.org/10.1121/1.3493450>

- Meyer, J., Dentel, L., & Meunier, F. (2013). Speech Recognition in Natural Background Noise. *PLoS ONE*, 8(11), e79279. <https://doi.org/10.1371/journal.pone.0079279>
- Michaud, A. (2005). Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica*, 61(2–3), 119–146. <https://doi.org/10.1159/000082560>
- Mika, K. (2013). Vietnamese Dialect Maps on Vocabulary. *Asian Geolinguistic Society of Japan, Monograph Series*, (1).
- Mohamed, M. M. T. (2010). *Constructing and Norming Arabic Screening Tool of Auditory Processing Disorders: Evaluation in a Group of Children at Risk For Learning Disability*. Ph.D thesis. Universität zu Köln. Retrieved from <http://kups.ub.uni-koeln.de/3123/>
- Moore, D. R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R. H., & Munro, K. J. (2014). Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0107720>
- Mukari, S. Z.-M. S., Wahat, N. H. A., & Mazlan, R. (2014). Effects of Ageing and Hearing Thresholds on Speech Perception in Quiet and in Noise Perceived in Different Locations. *Korean Journal of Audiology*, 18(3), 112. <https://doi.org/10.7874/kja.2014.18.3.112>
- Murphy, C. F. B., Rabelo, C. M., Silagi, M. L., Mansur, L. L., & Schochat, E. (2016). Impact of educational level on performance on auditory processing tests. *Frontiers in Neuroscience*, 10(MAR), 1–8. <https://doi.org/10.3389/fnins.2016.00097>
- Nekes, S. (2016). *Equivalent Hearing Loss in Children*. Univesität zu Köln. Retrieved from <http://kups.ub.uni-koeln.de/7092/>
- Nelson, L. H. (2015). Deaf Education Services in Southern Regions of Vietnam: A Survey of Teacher Perceptions and Recommendations. *Deafness & Education International*, 17(2), 76–87. <https://doi.org/10.1179/1557069X14Y.0000000048>
- 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>
- Ngô, N. L. (1977). *Speech audiometry using in diagnosing and evaluating occupational deaf*. Hanoi University of Medicine.
- Nguyen, A.-T. T., & Ingram, J. C. L. (2007). Acoustic and perceptual cues for compound-phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America*, 122(3), 1746. <https://doi.org/10.1121/1.2747169>
- Nguyễn, H. T. T. (2005). *Developing a Speech Perception Test in Vietnamese*. University of Iowa. Retrieved from <https://books.google.de/books?id=-j9cNwAACAAJ>
- Nguyễn, H. K. (1986). *Construction of Vietnamese speech tests and researching -applying technique in measuring speech audiometry*.

- Nguyễn, Q. D. (2014). Frequency of Phonemes in Southern Vietnamese Texts and its Contribution to Developing Speech Tests for Patients with Hearing Impairments. In *Vortrag bei der Tagung "Vietnamese Language and Culture" an der Universität Köln*. Retrieved from [https://vietnamese.uni-koeln.de/?page\\_id=116](https://vietnamese.uni-koeln.de/?page_id=116)
- Nguyễn, T. G. (2013). Ba cách xác định từ và hình vị tiếng Việt [Three Methods of Identifying Words and Morphemes in Vietnamese]. *Tạp Chí Khoa Học ĐHQGHN*, 29(4), 1–7.
- Nguyễn, V. L., & Edmondson, J. A. (1998). Tones and voice quality in modern Northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, (28), 1–18.
- Nissen, S. L., Harris, R. W., Channell, R. W., Richardson, N. E., Garlick, J. A., & Eggett, D. L. (2013). The Effect of Dialect on Speech Audiometry Testing. *American Journal of Audiology*, 22(2), 233. [https://doi.org/10.1044/1059-0889\(2013/12-0077\)](https://doi.org/10.1044/1059-0889(2013/12-0077))
- 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. <https://doi.org/10.3109/14992027.2010.542491>
- Offei, Y. N. (2013). *Educational audiology in Ghana – developing screening tools for hearing in infants and children. Inaugural - Dissertation*. Köln. Retrieved from <http://kups.ub.uni-koeln.de/5211/>
- Pagialonga, A., Tognola, G., & Grandori, F. (2014). A user-operated test of suprathreshold acuity in noise for adult hearing screening: The SUN (SPEECH UNDERSTANDING IN NOISE) test. *Computers in Biology and Medicine*, 52, 66–72. <https://doi.org/10.1016/j.combiomed.2014.06.012>
- Phạm, A. H. (2005). Vietnamese tonal system in Nghi Loc : A preliminary report, (August), 183–201.
- Phạm, A. H. (2003). *Vietnamese Tone. Outstanding Dissertations in Linguistics*. Routledge. <https://doi.org/10.4324/9780203500088>
- Phạm, A. H. (2009). The identity of non-identified sounds: glottal stop, prevocalic /w/ and triphthongs in Vietnamese \*. *Proceedings of the 3rd Toronto Workshop on East Asian Languages in Toronto Working Papers in Linguistics (TWPL)*, 34, 1–17.
- Phạm, A. H. (2008). The Non-Issue of Dialect in Teaching Vietnamese. *Journal of Southeast Asian Language Teaching*, 14.
- Phạm, B., & McLeod, S. (2016). Consonants, vowels and tones across Vietnamese dialects. *International Journal of Speech-Language Pathology*, 18(2), 122–134. <https://doi.org/10.3109/17549507.2015.1101162>
- Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, 42 Suppl 2, 2S26-32. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12918626>

- Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *International Journal of Audiology*, 47(Suppl 2), S72–S82. <https://doi.org/10.1080/14992020802307404>
- Profile.id. (2011). Australia - Language spoken at home. Retrieved from <http://profile.id.com.au/australia/language>
- Ramkissoon, I. (2001). Speech Recognition Thresholds for Multilingual Populations. *Communication Disorders Quarterly*, 22(3), 158–162. <https://doi.org/10.1177/152574010102200305>
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7(July), 31. <https://doi.org/10.3389/fnsys.2013.00031>
- Schneider, B. A. (2011). How age affects auditory-cognitive interactions in speech comprehension. *Audiology Research*, 1(1S), 1–5. <https://doi.org/10.4081/audiore.2011.e10>
- Schneider, B. S. (1992). Effect of Dialect on the Determination of Speech-Reception Thresholds in Spanish-Speaking Children. *Language, Speech, and Hearing Services in Schools*, 23(2), 159–162.
- Sherbecoe, R. L., & Studebaker, G. A. (2003). Audibility-Index Predictions of Normal-Hearing and Hearing-Impaired Listeners' Performance on the Connected Speech Test. *Ear and Hearing*, 24(1), 71–88. <https://doi.org/10.1097/01.AUD.0000052748.94309.8A>
- Shi, L.-F., & Canizales, L. A. (2013). Dialectal Effects on a Clinical Spanish Word Recognition Test. *American Journal of Audiology*, 22(1), 74. [https://doi.org/10.1044/1059-0889\(2012/12-0036\)](https://doi.org/10.1044/1059-0889(2012/12-0036))
- Stam, M., Smits, C., Twisk, J. W. R., Lemke, U., Festen, J. M., & Kramer, S. E. (2015). Deterioration of Speech Recognition Ability Over a Period of 5 Years in Adults Ages 18 to 70 Years. *Ear and Hearing*, 36(3), e129–e137. <https://doi.org/10.1097/AUD.0000000000000134>
- Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am*, 116(2), 1081–91.
- Studebaker, G. A., Gray, G. A., & Branch, W. E. (n.d.). Prediction and statistical evaluation of speech recognition test scores. *Journal of the American Academy of Audiology*, 10(7), 355–70. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10949940>
- Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., & Gray, G. A. (1997). Age-related changes in monosyllabic word recognition performance when audibility is held constant. *Journal of the American Academy of Audiology*, 8(3), 150–62. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9188072>
- Styler, W. (2013). Using Praat for Linguistic Research. *Savevowels*, 1–70. Retrieved from <http://savethevowels.org/praat/>

- 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, L. C. (1991). A Vietnamese reference grammar. Honolulu: University of Hawai'i Press.
- Townsend, T. H., & Bess, F. H. (1980). Effects of Age and Sensorineural Hearing Loss on Word Recognition. *Scandinavian Audiology*, 9(4), 245–248. <https://doi.org/10.3109/01050398009076359>
- Trần, D. D., Castelli, E., Serignat, J.-F., Trịnh, V. L., & Lê, X. H. (2005). Influence of F0 on Vietnamese syllable perception. *Interspeech*, (figure 1), 1697–1700.
- Tsukada, K., Nguyen, T. T. A., & Roengpitya, R. (2006). Cross-language Perception of Word-final Stops by Native Vietnamese Listeners : Preliminary Results on the Role of Specific , Non-native Phonetic Experience. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 118–123.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Ullrich, K., & Grimm, D. (1976). Most comfortable listening level presentation versus maximum discrimination for word discrimination material. *Audiology : Official Organ of the International Society of Audiology*, 15(4), 338–47. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1275818>
- Hoffmann, V. (2013). *Auditive Satzverarbeitung Bei Postlingual Ertaubten Erwachsenen Mit Cochlea-Implantaten. Inaugural - Dissertation*. Köln. Retrieved from <http://d-nb.info/1046939599/34>
- Varley, R., & So, L. K. H. (1995). Age effects in tonal comprehension in Cantonese.
- Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The Relationship between High-Frequency Pure-Tone Hearing Loss, Hearing in Noise Test (HINT) Thresholds, and the Articulation Index. *Journal of the American Academy of Audiology*, 23(10), 779–788. <https://doi.org/10.3766/jaaa.23.10.4>
- Võ, X. H. (2009). *Giáo trình ngữ âm tiếng Việt hiện đại (Text book of modern Vietnamese phonology)*. Quy Nhon: The University of Quy Nhon.
- Vũ, T. P. (1982). The acoustic and perceptual nature of tone in Vietnamese. In *Papers in South East Asian Linguistics*. Canberra: Australian National University.
- Vũ, T. P. (1981). *The Acoustic and Perceptual Nature of Tone in Vietnamese*. Australian National University.
- Vuong, H. Le, & Hoang, D. (1994). *Giáo trình ngữ âm tiếng Việt (Text book of Vietnamese phonology)*. Hà Nội, Vietnam: Đại học Sư phạm Hà Nội.
- Wagener, K. C., Brand, T., & Kollmeier, B. (2006). Evaluation des Oldenburger Kinder-Reimtests in Ruhe und im Störgeräusch. *HNO*, 54(3), 171–178. <https://doi.org/10.1007/s00106-005-1304-4>

- Wagener, K., & Kollmeier, B. (2005). Evaluation des Oldenburger Satztests mit Kindern und Oldenburger Kinder- Satztest. *Zeitschrift Für Audiologie*, *44*(134–143).
- Wagener, K. C. (2005). Moderne Sprachverständlichkeitstests für Kinder OlKi. *DGA Jahrestagung*, 1–4.
- Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. *International Journal of Audiology*, *44*(3), 144–56. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15916115>
- Weisleder, P., & Hodgson, W. R. (1989). Evaluation of four Spanish word-recognition-ability lists. *Ear and Hearing*, *10*(6), 387–92. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2606290>
- Weißgerber, T., Baumann, U., Brand, T., & Neumann, K. (2012). German Oldenburg Sentence Test for Children: A Useful Speech Audiometry Tool for Hearing-Impaired Children at Kindergarten and School Age. *Folia Phoniatrica et Logopaedica*, *64*(5), 227–233. <https://doi.org/10.1159/000342414>
- Wiley, T. L., Chappell, R., Carmichael, L., Nondahl, D. M., & Cruickshanks, K. J. (2008). Changes in hearing thresholds over 10 years in older adults. *Journal of the American Academy of Audiology*, *19*(4), 281–92; quiz 371. <https://doi.org/10.3766/jaaa.19.4.2>
- Wilson, R. H., Burks, C. a, & Weakley, D. G. (2005). Word Recognition in Multitalker Babble Measured with Two Psychophysical Methods. *Journal of the American Academy of Audiology*, *16*(8), 622–630. <https://doi.org/10.3766/jaaa.16.8.11>
- worldatlas.com. (2016). The Most Spoken Languages In America. Retrieved from <http://www.worldatlas.com/articles/the-most-spoken-languages-in-america.html>
- Yang, X., Wang, Y., Xu, L., Zhang, H., Xu, C., & Liu, C. (2015). Aging effect on Mandarin Chinese vowel and tone identification. *The Journal of the Acoustical Society of America*, *138*(4), EL411-EL416. <https://doi.org/10.1121/1.4933234>
- Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-Noise Tests for Multilingual Hearing Screening and Diagnostics1. *American Journal of Audiology*, *22*(1), 175. [https://doi.org/10.1044/1059-0889\(2013/12-0061\)](https://doi.org/10.1044/1059-0889(2013/12-0061))