SPEECH PERCEPTION TEST FOR JORDANIAN ARABIC SPEAKING CHILDREN

By

NADIA MOHAMED ABBAS ABDULHAQ

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

This dissertation is dedicated to my parents Abbas and Ursula.

ACKNOWLEDGMENTS

I would like to thank my parents for their constant support and believe in me. They have been a bottomless source of support, inspiration and encouragement. My sisters Ute Nabila, Najat, Ulla Najwa, and brothers Mejdi and Uwe Maysarah, who where ever in the world they were made sure to keep in touch and encourage me through the tough times and share the joy of my success.

My friend from Palestine, Lubna Shaheen, for all the times she has listened to my complaints and doubts and never let me go without giving me my confidence back. My good friend, Abdelqader Abu Awad, for his support and encouragement from the time I applied to the Fulbright program to the day of my graduation he has always been there for me.

There is a group of people here in Gainesville without whom I would have felt very lonely. I would like to thank all, Debra Anderson, International student coordinator, for being a shoulder to lean on, a good listener and a friend in times of need. She has been my inspiration in many activities on campus that have brought diversity to my experience and made it so much richer in people, memories, and achievements. My friends in Gainesville, Abdullhatif Qamihieh for his readiness to help at any time what ever the task was. Maisa Haj-Tas for being a wonderful understanding roommate and her stimulating discussions and insight into research. Marah Al-Aloul, Ashraf Al-Qdah, Suha Abu Lawi, Shadi Krecht, Yaser Katanani, Enas Katanani and Hadil Faqih, for their support and for being my family away from home through out these past years, and my friends from the

Fulbright program and international students for being a renewing source of hope, inspiration and fun.

This project was made possible by the continuous support by the faculty members of the Departments of Communication Sciences and Disorders and Communicative Disorders at the University of Florida. I would like to thank in particular my first mentor Dr. Patricia Kricos for her guidance, support and generosity. Dr. Griffiths for his advise through the graduate program in general and mentoring my dissertation in particular, Dr. James Hall III, for his guidance, support and encouragement, Dr. Richard Harris for his hospitality, generosity and guidance through the research process, and Dr. Tom Oakland, for his honest opinions, believes in me and support. I would also like to thank Dr. Linda Lombardino, Dr. Alice Dyson, Dr. Keneth Gerhardt, Dr. Mary Kay Dykes, and Dr. Aida Bamia for their support.

Funding for this project was made possible by the Gibson Dissertation Fellowship, offered by the Collage of Liberal Arts and Sciences at the University of Florida. The data collection and field research was made possible by the generous help and collaboration of Samar Al-Aghbar, the Jordanian ministry of education, and elementary schools of Amman, as well as the support of the Middle East Hearing Association and their team members, and Phonak hearing aid distribution office in Amman.

Completion of this degree was made possible by financial support from the department of Communication Sciences and Disorders in the form of teaching assistantship and Grinter Fellowship award for four years, as well as the College of Liberal Arts and Sciences O. Ruth McQuown Scholarship for women, in addition to the

Fulbright scholarship that was the seed funding for this degree, and finally and most importantly, my parents financial support.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SPEECH PERCEPTION TEST FOR JORDANIAN
ARABIC SPEAKING CHILDREN

By

Nadia Mohamed Abbas Abdulhaq

May 2006

Chair: Scott Griffiths
Major Department: Communication Sciences and Disorders

Comprehensive audiologic evaluation includes a variety of tests that provide a determination of the type of hearing loss. Among these tests are tests of word recognition. Many speech perception tests have been developed over the past half century to assess different aspects of speech. Management of hearing loss in developing countries and the USA differs significantly. The prevalence and demographics of the hard of hearing population is also different, and thus imposes different needs. According to the World Health Organization the number of deaf and hard of hearing individuals in developing countries is twice as much as in developed countries. There is a need for early identification and intervention for hearing loss in developing countries. At the same time there is a severe lack of equipment and highly trained professionals to provide such services. The focus of this study is speech audiometric measure for Jordanian Arabic speaking children. The first goal of this study is to develop four Jordanian Arabic 50-word lists appropriate to use for word recognition measure for Jordanian children age 6 to

9 years. The second goal is to investigate the effect of using words recorded in Saudi dialect on the word recognition abilities of Jordanian normal hearing and hard of hearing children.

Twenty individuals age 6 through 9 years participated in this study. A Jordanian speaker recorded 250 Arabic words selected for familiarity to elementary aged Jordanian children. The raw score data of all participants at the 10 intensity levels were compiled for each of the 250 words. Four equally difficult lists of fifty words each were derived from this experiment. In the second experiment, 3 Jordanian and 3 Saudi male speakers were selected to record 33 words. These were played back at a constant comfortable level for 10 children with normal hearing and 10 children with hearing impairment who were asked to identify the recorded words. Dialect produced a significant difference in performance for children with normal hearing, but not for children with hearing impairment. These results will lead to the development of improved techniques for assessing auditory performance in Arabic-speaking children.

CHAPTER 1
INTRODUCTION

Background of the Study

An audiological evaluation typically includes measures of tympanometry, acoustic reflexes, otoacoustic emission, and pure-tone (air and bone conduction) threshold and speech audiometry. Results of these tests are used to diagnose hearing loss and determine the type of hearing loss. Hearing testing has come a long way from estimating hearing from the distance a person can hear a voice. Since the early 1900s, pure-tone audiometry has been instrumental in measuring hearing sensitivity. The work of Harvey Fletcher on the perception of speech in relation to sound level and noise effect laid the foundation for decades of speech perception research (Fletcher, 1995). Speech perception measures in the present format in the United States of America (USA) have been used since the late 1940s and early 1950s (Egan, 1948; Haskins, 1949; Hirsh, Davis, Silverman, Reynolds, Eldert, & Benson, 1952). Many speech perception tests have been developed and improved over the past half century to assess different aspects of speech; e.g., speech reception thresholds, speech pattern identification, and speech reception in noise (Elliot & Katz, 1980; Erber, 1974; Goldman, Fristoe, & Woodcock, 1970; Jerger, Lewis, Hawkins, & Jerger, 1980; Ross & Lerman, 1970; Tillman & Carhart, 1966).

Speech reception tests that mostly are used clinically include the Central Institute for the Deaf W-22 (CID W-22) by Hirsh et al. (1952), Northwestern University Auditory Test No. 6 (NU-6) by Tillman & Carhart (1966), Phonetically Balanced Kindergarten

Test (PBK-50) by Haskins (1949), and Northwestern University Children's Perception of Speech (NU-CHIPS) by Elliot & Katz (1980). Common characteristics among these tests are that they are based on monosyllabic frequently used words which are familiar to the target population (at least at the time of test development) and were developed based on measures of correct responses as a function of intensity of presentation. The articulation function is a common representation of speech audibility result, which forms an ogive or S-shaped curve and indicates the degree a person's hearing ability improves given increases in the intensity of the material presented (Carhart, 1951). The shape of the ogive and the 50% point (threshold point) are affected by the speech material presented as well as the speaker. That is, different speech materials produce different ogives. If different speakers recorded the same speech material, the result would yield different ogives (Beattie, Edgerton, & Svihovec, 1977; Beattie, Svihovec, &Edgerton, 1975; Carhart, 1965; Doyne & Steer, 1951; Hirsh, Reynolds, & Joseph, 1954; Kruel, Bell, & Nixon, 1969; Wilson & Carter, 2001; Wilson & Oyler, 1997). Results from adult speech recognition test (NU-6) and children's tests (PBK-50 and WIPI) were compared (Sanderson-Leepa & Rintelman, 1976). The adult test was found to be inappropriate for use with young children age 3.5 to 9.5 years while WIPI was most appropriate for young children ages 3.5 to 5.5, and NU-6 was appropriate but more difficult than the PBK-50 for children age 7.5 to 11.5 the. That is, speech material needs to be age appropriate, and words need to be familiar to the target group. Carhart (1965), and Kreul et al. (1969) recommend the use of a limited set of recorded materials to establish normative measures because different recordings by the same speaker can result in different articulation functions.

Studies of speech perception in other languages (e.g., Arabic, Polish, Korean, and Chinese) have followed Egen's (1948) and Hirsh et al. (1952) lead in developing speech reception material (Alusi, Hinchcliffe, Ingham, Knight, & North, 1974; Zakrzewski, Jassem, Pruszewicz, & Obrebowski, 1976; Ashoor & Prochazka, 1982; Ashoor & Prochazka 1985; Harris, Kim, & Eggett, 2003a; Harris, Kim, & Eggett, 2003b; Nissen, Harris, Jennings, Eggett, & Buck, 2005a; Nissen, Harris, Jennings, Eggett, & Buck, 2005b). Special attention was paid to the familiarity of words, the equivalence of word lists, and in some cases the phonetic balance of word lists. The resulting lists have had similar articulation functions to the English word lists in shape and slope. Thresholds of speech reception and word recognition were more varied across languages.

Rationale and Purpose

Hearing Loss in Developing Countries

The management of hearing loss in developing countries and the USA differs significantly. The prevalence and demographics of the hard of hearing population also differs and thus imposes different needs for services. According to the World Health Organization (WHO), the number of deaf and hard of hearing individuals in developing countries is twice that in developed countries (Smith, 2001). According to the American Academy of Audiology an average of 3 in every 1,000 healthy newborns in the United States has severe sensorineural hearing loss. In Jordan 6 in every 1,000 healthy newborns have hearing loss (Al-Masri, 2003). For comparison purposes, the WHO reported that 4 to 5 children under the age of 18 in every 1,000 children have sensorineural hearing loss in the South-East-Asian Region (Smith, 2001). This number is reported to be inaccurate due to the lack of epidemiology surveys in developing countries; the actual numbers are projected to be double what has been reported (Smith, 2001). Globally, the majority of children are living in developing counties, which indicates that the majority of children with hearing loss live in those countries as well (Jauhiainen, 2001).

The WHO identified the major causes of hearing loss to be chronic otitis media, genetics, maternal and perinatal problems, and ototoxicity (Smith, 2001). The rates of complications due to otitis media are 1/1000 in developed countries and 60/1000 in developing countries. The incidence of deaths due to ear infections complications are 1/100,000 in developed countries and 1 /100 in developing countries (WHO, 1998). In spite of the higher prevalence of hearing loss in developing countries, services and technology generally are limited or lacking, especially in rural areas (WHO, 1998; Jauhiainen, 2001). The prevalence of Ear, Nose and Throat doctors ranges from 1/ 30,000 to 1/150,000 in developed countries and 1/2,000,000 in the less developed countries in Africa (excluding South Africa and Egypt) (WHO, 1998).

Early identification and intervention for hearing loss in developing countries are need. At the same time there is a severe lack of equipment and highly trained professionals to provide such services. In spite of this general statement, services in some developing counties are more advanced than others, and some individuals have the means to afford world class services (Jauhiainen, 2001). The focus of this study is on audiology in Jordan, specifically speech audiometric measure for Jordanian Arabic speaking children.

Hearing Loss in Jordan

Pilot data from screenings conducted by the Middle East Hearing Association (MEHA) suggest estimates of the hard of hearing and deaf population in Jordan to be 64,000, with 2,200 infants with severe hearing loss born yearly (Al-Masri, 2003). Some institutions provide limited services, including the Speech and Hearing Clinic at the

University of Jordan, Al-Ahliyya Amman University, Holyland Institute for the Deaf, King Hussein Medical Center, and some private otolaryngology and audiology clinics and hearing aid providers. However, diagnostic and rehabilitative services and professional training are seriously lacking. There are initiatives by the Jordanian government, the royal family in particular, to develop such services. Under the patronage of His Royal Highness, Prince Firas Raad, MEHA was established in 1998 in cooperation with the Canadian International Scientific Exchange Program, and a new center was opened recently to provide services to the hearing impaired and deaf population. Projects being implemented through this organization include newborn hearing screening, genetic hearing loss research, audiologic evaluation, hearing aid fitting, audiologic rehabilitation and follow up for children. Equipment needed for such services is available and some basic audiometric measures are being performed. An important part of evaluation, speech perception audiometry, is missing from the diagnostic battery. Speech perception materials suitable for testing Jordanian Arabic speaking children are unavailable. The purpose of this study is to develop a speech perception test for Jordanian Arabic speaking children.

Present Condition of Arabic Speech Audiometry

Carhart (1951) emphasized the importance of using familiar words that are in the listener's native language. Tests developed in Arabic have been recorded using Maroccan (Messouak, 1956), Iraqi (Alusi, 1974), Egyptian (Soliman, 1976; Soliman, Abd El-Hady, Saad, & Kolkaila, 1987; and Soliman, Fathallah, & El-Mahalawi, 1987), and Saudi (Ashoor et al., 1982; and Ashoor et al. 1985) dialects. The question arises as to whether the different dialects have an effect on speech reception. Although all Arabic countries share one standard Arabic language that is taught formally in school in the form of the written language (Altoma, 1969; Al-Kahtani, 1997; Fatihi, 2001; Ferguson, 1956) dialectal differences exist. Studies conducted to measure the dialectal effect on speech reception in Arabic could not be located. Thus, one may question the effect of the dialect related to the words used in that dialect as well as the articulation and voice characteristics of the speaker.

Recordings by different speakers result in different word recognition scores, and the use of different speech material has the same effect (Beattie et al., 1977; Beattie et al., 1975; Carhart, 1965; Doyne & Steer, 1951; Hirsh et al., 1954; Kruel et al., 1969; Wilson & Carter, 2001; Wilson & Oyler, 1997). This suggests a need to develop speech audiometric material based on one recording to ensure reliable results. Recordings reported in other studies (e.g. Alusi et al., 1974; Ashoor & Prochazka, 1985) have not been available for wide use. One intended outcome of this study is to make available a high quality digital recording of speech material appropriate for use in testing word recognition loss of Jordanian children. Audiologic testing in Jordan currently is based on non-speech related audiometry, including pure-tone audiometry, otoacoustic emissions, and auditory evoked potentials.

This study has two main goals. One is to develop four Jordanian Arabic 50-word lists appropriate to use in word recognition measures for Jordanian children ages 6 through 9. Another goal is to investigate the effect of using words recorded in Saudi dialect on word recognition abilities of Jordanian children who display normal hearing and hearing loss.

Research Questions

Experiment One

1. Is there a significant difference in word recognition abilities of Jordanian Arabic speaking children, given increases in intensity of presentation?

2. Can four parallel word lists be developed in Jordanian Arabic language (e.g., their psychometric qualities of the word lists do not differ significantly)?

Experiment Two

1. Do word recognition abilities of Jordanian Arabic speaking children differ when listening to words presented in Jordanian versus Saudi dialects?

2. Do word recognition abilities of Jordanian Arabic speaking children who display normal and hearing disabilities differ when hearing words presented in a Jordanian versus Saudi dialect?

3. Does word recognition ability differ when hearing speakers using the same dialect?

Hypothesis

Experiment One

1. An increase in presentation level will increase word recognition scores.

2. Four lists of 50-words with similar characteristics can be created to be used in measures of word recognition abilities.

Experiment Two

1. Jordanian children will produce higher word recognition scores when listening to words presented in Jordanian dialect than when listening to words presented in Saudi dialect.

2. Jordanian and Saudi dialects will have the same effect on word recognition scores of Jordanian children with normal hearing and with hearing loss.

3. Word recognition will differ for two talkers with the same Arabic dialect.

CHAPTER 2
REVIEW OF THE LITERATURE

Importance of Speech Audiometry

Non-speech audiometric procedures provide valuable information needed for description of hearing loss and estimation of amplification benefit. These tests, however, cannot measure the effect of hearing loss on speech (Carhart, 1951; Cramer & Erber 1974; Davis, 1948; Doyne & Steer, 1951; Erber, 1974; Hirsh, et al. 1952).

Speech audiometry requires a language related test material, and thus may be influenced by the phonetic, melodic, and intonational differences between languages (Carhart, 1951). Different languages require speech tests that consider the features unique to the language. Distinctive features of languages can result in different auditory requirements and affect the auditory capacities evaluated. To provide comprehensive audiologic services, it is necessary to use speech audiometry for measurement of hearing and the outcome of management.

Speech perception is important in facilitating cognitive development and normal language acquisition (Cramer & Erber 1974). Hearing loss is an obstacle of language and speech acquisition. Pure-tone thresholds provide information about detection of sound at specific frequencies, yet provide little information about the perception of complex signals such as speech. Knowledge of a person’s ability to perceive speech can provide information as to the extent hard of hearing children are able to communicate effectively and, in addition, how they are likely to learn language. Speech audiometry is a clinical approach in which well-defined speech samples are presented using a calibrated system to measure an important aspect of hearing ability. Measurement of speech perception different intensities results in an articulation function taking the shape of an ogive or S-shaped curve, that indicates the degree a person's hearing ability improves with an increase of the intensity of the material presented (Carhart, 1951).

Over 50 years ago, Carhart (1951) reported on the importance of speech audiometry in hearing assessment. He considered it to be the most useful contribution to hearing testing after the introduction of pure-tone audiometry. Carhart emphasized the value of well-defined speech audiometry to provide finer classification of hearing loss, and its importance in providing measures for educational and rehabilitation purposes. Another early hearing researcher, Hallowell Davis (1948), acknowledged that there is no simple interpretation of the pure-tone audiogram to express the patient's ability to hear speech. It is important to measure the patient's ability to discriminate speech just as it is important to measure the hearing loss in decibels.

Fletcher noted, "The process of speaking and hearing are very intimately related, so much so that I have often said that, we speak with our ears. We can listen without speaking but we can not speak without listening" (p.A1 Fletcher, 1995). Speech perception and language acquisition are two closely related processes. Hearing loss usually disrupts the process of speech perception and thus delays language acquisition. Therefore, Cramer & Erber (1974) emphasized the need for an accurate and valid measure of speech perception for hearing impaired children. Arlinger (2001) emphasized the importance of using speech recognition to measure hearing aid benefit in children, and considered periodic evaluation of hearing aid benefit of great importance for children to monitor their language and speech development. Blamey, Sarant, Paatsch, Barry, Bow, Wales, Wright, Psarros, Rattigan, & Tooher (2001) used a measure of speech perception to evaluate benefit of amplification, and the relationship between speech perception and speech production, language, hearing loss and age in 87 children age 4 to 12 years. According to this study, the authors expected the language delay to be 4 to 5 years by the time children are 12 years old. Speech perception scores are expected to improve significantly with the improvement of language; specifically they expect children to score 90% on the open set Bench-Kowal-Bamford (BKB) sentence test when they reach the level of language proficiency of a 7 year old. Öster (2002) examined the relation between audiological measures and speech intelligibility for eleven profoundly deaf Swedish teenagers (age 15 to 17 years). Their pure-tone averages (PTA) ranged from 90 to 108 dB HL. Correlation analysis was assessed for the intelligibility of the children's speech and their pure-tone average, the shape of the audiogram, and residual hearing use. Results showed that there was great variation in their speech intelligibility in spite of the narrow range of PTA, indicating that the speech intelligibility cannot be estimated based on PTA. The correlation between speech recognition scores and speech intelligibility scores resulted in a positive correlation of 0.73, confirming a high correlation between residual hearing use and speech intelligibility. Öster (2002) concluded that a simple speech test can be used as a predictor of prelingually deaf children's ability of developing intelligible speech. Laukil & Fjermedal (1990) researched the reproducibility of bone conduction thresholds and the speech recognition thresholds. The results show low variability and no significant difference between the two measures making the speech recognition threshold measure a reliable one for Norwegian Spondees. These studies indicate that speech perception measures are very important and reliable clinical tools.

Speech Audiometric Tools

Speech audiometry is an important part of an audiologic evaluation battery (Arlinger 2001; Carhart 1951; Cramer & Erber, 1974; Curry 1949; Davis 1948; Erber, 1974). There are two essential measures of speech audiometry: the measure of speech recognition threshold (SRT) and the measure of speech recognition at suprathreshold levels (Carhart 1951; Davis, 1948). Davis (1948) has identified these measures as two dimensions of hearing loss; one is hearing loss (dB level) and the second is discrimination loss (word recognition score). Discrimination loss relates to the loss of the ability to recognize words even when they are made audible. Several tests have been developed to measure these two dimensions of hearing loss. The most frequently used ones for speech perception are speech recognition threshold (SRT) test and word recognition scores (WRS). An example of the SRT test is the Central Institute for the Deaf spondee word test (CID W-2) by Hirsh et al. (1952). Many word recognition materials lists have been developed over the years including the Phonetically Balanced 50 word lists (PB-50) described by Egan (1948), the CID W-22 (Hirsh et al., 1952), the Northwestern University Test No.6 (NU-6) by Wilson and Oyler (1997), the PB Kindergarten word test (PBK) by Haskins (1949), the Word Intelligibility for Picture Identification Test (WIPI) by Ross & Lerman (1970), and the Northwestern University Children's Perception of Speech (NU-CHIPS) by Elliot & Katz (1980). Egan (1948) specified the criteria of selecting word lists for the word recognition tests as follows:

1. Monosyllabic structured words
2. Equal average of difficulty between lists
3. Equal range of difficulty within lists

4. Equal phonetic composition between lists
5. A composition representative of spoken English
6. Commonly used words

The above listed speech perception tests were developed following Egan's guidelines. Speech audiometric tests have been developed in many different languages, each to fit the requirements of measuring speech perception based on the specific features of that language.

Measurement of speech perception tests included the articulation function that defines word recognition scores at different intensity levels. The measurement starts at a very low intensity level where the material is unintelligible. As the signal intensity level is increased, the listener is able to identify correctly more of the stimuli up to a point at which the intensity is high enough for the listener listening to his native language to identify all the material without error. Fletcher (1929) demonstrated that the shape of the curve changes from one material to another and, using the same material, measurement of speech recognition with different speakers results in different shaped curves. With speech discrimination loss, the articulation curve not only shifts at the dB axis but it also changes in shape. The curve reaches a plateau at percentage correct levels below 100% (Davis, 1948). Davis concludes that this effect results from the loss of sensitivity, especially in the high frequencies, which is important for consonant recognition and clarity of speech. The shape of the curve is again different for each hearing impaired individual. It might be of normal shape and shifted to the right indicating higher intensity levels, or it might have a different shape and not reach the 100% correct identification.

Pediatric Speech Perception Materials

Children’s speech perception tests have been developed with different materials and different tasks appropriate for different age and ability levels. Some tests are “open set” (i.e. the listener has no knowledge of the category of word or any contextual cues) and require verbal response (e.g. Phonetically Balanced Kindergarten 50 Word Test, Haskins 1949), whereas others are “closed set” (i.e. the listener is provided with a set of 3-6 options to chose from or the category of words is specified to provide some cues) requiring picture pointing (e.g., Word Intelligibility by Picture Identification, Ross & Lerman, 1970). Following is a detailed description of the most commonly used children speech perception tests (see table 2-1 for list).

Haskins (1949) developed The Phonetically Balanced Kindergarten 50 Word Test. Though the Haskins lists appeared on in the author’s masters’ thesis at Northwestern University, they have been widely used.The four lists were developed based on the phonetically balanced word lists (PB-50) used by Egan (1948) for assessment of speech perception in adults . Haskins selected words that were among the 2500 words most used by kindergarten children (The International Kindergarten Union, 1928). Measurement of the psychometric function and the equivalence of wordlists were completed with adult normal hearing participants with one randomization of lists presented at 5 intensity levels. As a result, lists 1, 3, and 4 were found to be equivalent while list 2 was easier. The slope of the psychometric function between 20 and 80% word correct was 4%/dB, as reported in Mayer & Pisoni (1999). In spite of the wide use of these lists, no formal data collection and analysis was completed with pediatric populations.

Other speech perception audiometric tests for children were mostly developed for closed sets of words (Elliot & Katz, 1980; Erber, 1974; Erber, 1980; Goldman et al., 1970; Jerger et al., 1980; Ross & Lerman, 1970), and sometimes with groups of hard of hearing children as the only participants (Erber & Alencewicz, 1976; Ross & Lerman, 1970). Several closed set picture presentation tests are used regularly in audiology clinics. These tests include the Word Intelligibility by Picture Identification (WIPI) by Ross and Lerman (1970), the Northwestern University Children's Perception of Speech (NU-CHIPS) by Elliott and Katz (1980), Goldman-Fristoe-Woodcock Test of Auditory Discrimination (Goldman et al., 1970), and The Pediatric Speech Intelligibility test (PSI) by Jerger et al. (1980). Erber & Alencewicz (1976) suggested a picture pointing closed test to evaluate the word recognition ability of children with hearing loss that provides a distinction of word recognition and word pattern recognition.

Adult Speech Perception Materials

Beattie & Warren (1983) described for adult word recognition tests in English, an increase in intelligibility with the increase of intensity equivalent to 4.5%/dB in the range of 20 to 80% scores, with an approximation of maximum intelligibility at level s of 25 dB SL. Wilson, Zizz, Shanks, & Causey (1990) reported the NU-6 word recognition threshold when spoken by a female speaker to be 4.5%, similar to other studies, while the intensity level of 50% correct recognition was shifted 5dB to the right (higher than previous studies). Wilson & Oyler (1997) compared the psychometric function of the CID W-22 word lists and the Northwestern University No.6 (NU-6) as spoken by the same talker and found the 50% score level to be at 15.6dB HL for the W-22 and 13.4 dB HL for the NU-6. The slopes between 20% and 80% points were 4.8%/dB for the W-22 and 4.4%/dB for the NU-6. These results are comparable to speech audiometry data for other languages For example, Harris et al. (2003a) studied the psychometric function of wordlists spoken by males and females and found the mean 50% level in Korean to be at

11.4 dB HL for male speakers and 10.7dB HL for female speakers with mean slopes between 20% and 80% points of 4.4%/dB for male and female speakers. Niessen et al. (2005a) found a threshold level for Chinese Mandarin speech materials of 5.4 dB HL for male speakers and 2.3dB HL for female speakers, and mean slopes between 20% and 80% points were 6.3%/dB for male speakers and 7.1% for female speakers. The difference in the Chinese word lists might have been because disyllabic words were used in the composition of the lists. Alusi et al. (1974) developed equivalent lists for speech recognition in Arabic with a threshold level of 22.5 dB HL and a slope of 5%/dB.

In summary, studies of monosyllabic lists in three languages (English, Korean, and Arabic) show similarity in psychometric function slopes (ranging from 4.4% to 5.1%) whereas the study of disyllabic Chinese word lists showed steeper slopes (6.3% and 7.1%). Table 2-2 lists the thresholds and slopes of word lists in the different languages listed above. Review of data displayed in Table 2-2 shows that there is a difference in threshold among languages, with the Chinese disyllabic words having the lowest threshold (2.3 dB HL) and the highest for Arabic monosyllabic word lists (22.5 dB HL).

Speech Reception Threshold Material

Studies of speech recognition thresholds (SRT) lists in different languages revealed steeper slopes than those reported for word recognition lists. SRT measures in English are composed of disyllabic spondees. Hirsh et al. (1952) reported a psychometric function slope between 20% and 80% of 8%/dB. Young, Dudley, & Gunter (1982) reported a slope of 10%/dB, and Wilson & Strouse (1999) reported a slope of 7.4%/dB. In studies of languages other than English, such as the study of trisyllabic Chinese Mandarin materials, Nissen et al. (2005b) found a slope between 20 and 80% of 9.7%/dB for a male speaker and 10.5%/dB for a female speaker. According to Nissen (2005b), slopes of SRT tests in other languages are at similar levels, i.e., Polish = 10.1%, Spanish = 11.1%, and Italian = 7.3%. Harris et al. (2003b) in their study of Korean disyllabic words found a slope of 10.3% for male speakers and 9% for female speakers. Ashoor & Prochazka (1982, 1985) reported a slope of SRT word lists of 5% for both adults and children's wordlists. Slopes of Ashoor's lists are less steep than those reported in other languages, but similar to slopes of word lists used in Alusi's word recognition scores for adults. Siegenthaler, Pearson, & Lezak (1954) investigated the speech reception threshold for children using monosyllabic words and found the slope to be 8.6%/dB between 20% and 80% correct word recognition. Ashoor & Prochazka found the threshold of word recognition to be at 2.2dB HL for adults (1982) and at 0 dB HL for children (1985). Table 2-3 lists speech reception threshold wordlists in different languages.

Full List and Half List Use

The use of full lists or half lists depends on the patients' performance on the test. Studies by Beattie and Warren (1983), Dubno , Lee, Klein, Matthews and Lam (1995) , and Thornton and Raffin (1978) investigating the confidence intervals of using full lists of 50 words and half lists of 25 words in initial testing and retesting of patients' performance on speech reception. Thornton and Raffin (1978) described the variability in speech discrimination scores based on the CID W-22 test, and highlighted the differences in variability between using full lists and half lists. Their results show the fact that the closer the scores are to either end of the spectrum of scores (0 or 100) and the more words are included in the list the less variability there is in scores and the smaller the confidence interval. For example if a patient scored 96% on a 50 word list the confidence interval is between 86-100. In other words if the patient scored 96% the first time and was retested, a score between 86-100 will be considered not different from the first score. While if the patient scored 96% on a half list of 25 words, the confidence interval is between 80 and 100. With lower scores, closer to 50%, the confidence interval grows larger more so for half lists than for full lists. This can be used as an indication for the need to use a full list versus a half list. That is if the score is closer to 50%, the use of 50 word lists would provide a more accurate measure. In comparing the test retest results from hearing impaired participants, Beattie and Warren (1983) found the standard deviation in test retest results using 25 word lists to be 10%, which was reduced to 8% using 50 word lists and to 6% when using 100 word lists. In Beattie et al.'s judgment, this difference was not significant to increase the size of test material and was satisfied with the 25-word list size.

Dubno et al. (1995) studied the correlation between the degree of hearing loss and word recognition scores. They studied the word recognition scores from 407 ears with normal hearing and mild to severe hearing loss, with the goal of providing data for confidence limits of scores on 25 and 50 NU-6 word lists in relation to the PTA. The authors provided tables of scores for 25 and 50 word lists corresponding to the 95% confidence limit of best performance (PBmax). They found a correlation between word recognition score and PTA, where a lower PTA resulted in higher scores. These findings are intended to help in diagnosis decisions on whether the score is considered within expected range for the degree of loss or whether it is poorer than expected and thus requiring additional testing. Dubno et al. (1995) cautioned about the use of these tables to generalize to other lists since different material would have different results. Still this gives an indication for clinicians and researchers to be cautious when using word lists at one presentation level and to keep this data in mind when making clinical decisions.

Arabic Speech Audiometry

Several speech perception tests are available in the Arab countries. These include speech recognition tests for adults in Moroccan (Messouak, 1956), Iraqi (Alusi et al., 1974), Egyptian (Soliman, 1976) dialects, and an SRT test (Ashoor & Prochazka, 1982) in Saudi dialect. Pediatric speech audiometry tests have been developed including an SRT test in Saudi (Ashoor & Prochazka, 1985) and Egyptian (Soliman et al., 1987b) dialect, an Arabic word intelligibility (recognition) by picture identification in Egyptian dialect (Soliman et al., 1987a), and an Arabic speech pattern contrast (ArSPAC) test developed in Israel (Kishon-Rabin & Rosenhouse, 2000). Some of these publications are inaccessible (Messouak, 1956; Soliman, 1976; Soliman et al., 1987a; Soliman et al., 1987b). Insufficient details are found in the literature to allow for in depth discussions of these tests. Recordings of the test materials are not widely available, even though the word lists are printed in the publications (Alusi et al., 1974; Ashoor & Prochazka, 1985; Kishon-Rabin & Rosenhouse, 2000). Allusi et al.'s (1974) and Ashoor & Prochazka's (1982 and 1985) studies address word recognition comparable to the present study; therefore these studies are discussed in detail.

Comparison of Arabic Speech Reception Tests

Alusi et al. (1974) and Ashoor & Prochazka (1982) used monosyllabic words in their word lists. The structure of monosyllabic words was CVC, CVCC, CVVC, and CVVCC, with all 28 consonants and 6 vowels of standard Arabic represented. Alusi et al. (1974) divided the 150 monosyllabic words into 6 phonetically balanced lists of 25 words. Ashoor & Prochazka's adult lists (1982) contained 120 words divided into 6 phonetically balanced lists of 20 words, and their children's lists (1985) included 80 words divided into 8 lists of 10 words. Both Alusi and Ashoor based their phonetic balance of lists on the frequency of consonant and vowel occurrence compared to continuous text, based on counts conducted at the time of study. Ashoor & Prochazka (1985) in his children's lists focused on equal distribution of syllable structure more than the phonetic representation, while he kept the overall phonetic balance in all 80 words.

To ensure word familiarity, both researchers used standard Arabic, the main teaching language in schools and universities as well as the language of mass media (i.e., newspaper, radio and TV broadcast). The word sources of choice were elementary school books, children's stories, and daily newspapers. Absurd words and technical vocabulary were excluded. In addition both researchers chose words that are similar in standard Arabic and colloquial. Ashoor & Prochazka (1982 and 1985) ensured the word familiarity by collecting ratings from a large number of participants coming from 14 different regions of Saudi Arabia.

Lists were recorded at voice intensity of 70 to 75dB SPL, fluctuation was limited to ±5 dB on volume meter in sound treated booths, and ambient noise did not exceed 30 dBA SPL. Neither researcher used carrier phrases. Alusi et al.,'s recording (1974) was in standard Baghdad dialect while Ashoor and Prochazka's recordings (1982 and 1985) were in standard Saudi dialect. The rates of recording varied. Alusi et al. (1974) recorded 8 words per minute, whereas Ashoor and Prochazka recorded 12 words per minute for adults and 6 words per minute for children (Ashoor & Prochazka, 1982, and 1985).

The intelligibility tests were performed with somewhat different criterion. Alusi et al. (1974) included 17 participants age 20 to 38 years representing several Arab countries, since his goal was to develop a test that could be used in different Arab countries. Ashoor & Prochazka (1982) enrolled 74 adult participants age 20 to 35 years representing 14 rural and urban areas in Saudi Arabia, and 100 children age 4 to 9 years representing most Saudi dialect areas. Both studies (Alusi et al., 1974; and Ashoor & Prochazka, 1982) presented the stimulus words via headphones to their adult participants, while Ashoor and Prochazka (1985) presented the stimulus words via sound field to their pediatric participants. In all studies the authors attempted to measure the difference in inter-list difficulty and also the difference associated with presentation method. The word lists were presented at different intensity levels ranging from speech detection thresholds, increasing by 5dB steps, up to the point where 100% of the words were identified correctly. There was no difference between lists in terms of difficulty or the order of stimuli presentation in ascending or descending dB level, within each study. Resulting articulation function curves from all three studies were similar to those of other languages, including English, in shape and slope. The findings suggest the lists are suitable as speech recognition measure.

Ashoor & Prochazka (1982) found a slight difference between the adult groups of students and non-students, i.e., a curve shift further to the right for the latter group indicating higher threshold levels (in dB). In addition, he found a difference between two age groups (4 to 5 years and 5 to 9 years). The younger group needed higher intensity levels than the older group to reach threshold level of 50% word recognition, a finding that may be related to maturation and knowledge of the language.

Use of Arabic Speech Reception Tests

Though Alusi et al. (1974) and Ashoor & Prochazka (1982 and 1985) recorded their material, these recordings were not marketed for wide distribution. The limited distribution may have several explanations. One is the small number of audiologists and audiological services in Arab countries. For example, Saudi Arabia has one of the most advanced services for communication disorders among the Arab countries. With a population of 25 million, there are 14 registered audiologists at the Saudi Speech Pathology and Audiology Association and five facilities that provide audiological services (SSPAA, 2004). The second possible explanation for the limited distribution of speech materials is the difference in dialects between Arab countries. Although Arab countries share the standard written Arabic language, there is a wide range of dialects (Fatihi, 2001). Published speech recognition tests are in Moroccan, Baghdadi, Egyptian or Saudi standard dialect. The possibility of using one test across the Arab countries has not been investigated. Alusi et al. (1974) has suggested the possibility of using the word lists he developed in all Arab dialects since the words were taken from standard Arabic. However, Alusi's speech materials were recorded in a Baghdad standard dialect. In developing the speech test, Alusi had a limited number of participants (17) representing "several" Arab countries (the author did not specify which countries), who were young educated adults. The sample did not necessarily represent the large Arabic speaking population). However, Alusi et al. did attempt to meet the criterion of word familiarity by choosing words from children's books and newspapers in order to include educated and un-educated populations. He did not describe a specific comparison between participants from different countries to support his argument.

Dialectal Differences in Arabic

One of the goals in the present study is to determine whether there is a difference in word recognition scores for normal hearing and hearing impaired children listening to Jordanian and Saudi dialects. The issue of Arabic language diversity and its dialects is important in the selection of speech material and speaker. Although a single standard Arabic language is used in all Arab countries, dialectal differences do appear. The difference between standard and dialect with the Arabic language, similar to other languages, such as Greek, Swiss German, and Haitian Creole, is referred to as diglossia. Furgeson (1959) defined diglossia as the presence of a stable situation of a language in which there is a dialect of a language primarily used in daily communication and a very different superimposed variety of the language that is part of a highly respected large body of written literature or a previous period. The written literature is learned by formal education and is used in formal speech (such as news and political speeches), but not in every day conversation. All written materials (e.g., school books, news paper, commercial material, official documents, and instructional materials) are written in standard Arabic and in social settings people use the colloquial dialect. Thus children learn the colloquial dialect first and the standard language is taught formally in educational settings (Al-Kahtani, 1997; Altoma, 1969; Fatihi, 2001; Ferguson, 1956). With the increase in mass media and early education, children nowadays are more exposed to standard Arabic through television programs, radio and early reading experience.

The differences between the two versions of Arabic are grammatical, phonetic, and lexical in nature. Standard Arabic is considered syntactically more complex and richer in lexicon. In spite of these differences, there are many similarities. Altoma (1969) found that 83.5% of words in different colloquial dialects are shared with standard Arabic, as well as the syllabic structures of words. In general standard Arabic is considered a more prestigious language but there is no competition between the two versions of language since each serves a different purpose and they are not totally interchangeable in use by situation. In other words there are situations where standard Arabic is inappropriate and visa versa (Al-Kahtani, 1997; Ferguson, 1959; Ibrahim, 2000).

Abd-el-Jawad (1987) and Al-Kahtani (1997) reported that educated Arabic speakers frequently switch between standard and colloquial Arabic in a very natural manner. Differences are to some extent between social groups, but these differences have no effect on communication. Linguists have determined the status of a dialect based on the characteristic of mutual intelligibility among dialects. Mutual intelligibility is based on a scale of the physical proximity of the regions, that is, the closer the region the more mutually intelligible, and the further apart the regions, the less mutually intelligible (Fatihi, 2001). The differences and similarities between dialects and between standard Arabic and dialects must be considered when developing speech perception test materials and when choosing speakers. Several groups of researchers (Beattie, et al. 1975; Beattie, et al. 1977; Carhart, 1965; Doyne & Steer, 1951; Hirsh, et al. 1954; Hood & Poole 1980; Kruel, et al. 1969; Palmer, 1955; Wilson & Carter, 2001; Wilson & Oyler, 1997) demonstrated significant differences in speech perception scores with different speakers, regardless of gender or age, assuming the speech perception test material were recorded by native speakers of English in standard dialects. In the case of Arabic, it can be argued that there is no difference among the different Arab countries when using standard Arabic. The question whether standard Arabic is a representative sample of the spoken dialects remains unanswered.

Special Considerations for the Present Study

Participants

Speech perception differs significantly between age groups. Elliot (1979) found that scores of children age 9 were poorer than older children on speech perception in noise test (SPIN), while the older group (15 to 17 year olds) scores were comparable to adult scores. Elliot, Connors, Kille, Levin, Ball & Katz (1979) found no significant difference in scores for children age 5 to 8 years while 10 year olds performed at adult levels (no 9 year old participants were included in this study). Schwartz and Goldman (1974) assessed the performance of young children in nursery, kindergarten and first grade and found significant differences between the three groups. In another study Sanderson-Leepa & Rintelmann (1976) compared the speech performance of children ages 3.5 to 11.5 on different speech perception tests and found no significant difference in the 7.5 and 9.5 age groups on the WIPI and PBK-50 tests compared to the younger and older groups. Ashoor & Prochazka (1985) found similar age differences for the Saudi Arabic test when comparing scores for 4 to 5 year old children to scores for 6 to 9 year old children'. Based on these results, participants aged 6 to 9 years were recruited in the present study.

As mentioned above, studies by Elliot (1979), Elliot et al. (1979), Sanderson-Leepa & Rintelmann (1976), Schwartz & Goldman (1974), and Goldman et al. (1970), show that significant differences in word recognition scores are found for children younger than 6 years and older than 9 years in comparison to children age 6 to 9 years. These others reported no significant difference in word recognition scores between children aged 6 through 9 years.

Speakers

Individual differences between speakers can affect speech perception scores (Hood & Poole 1980). Different speakers produce different articulation curves when using the same words. Most words maintain their order of difficulty across speakers. Palmer (1955) investigated the effect of gender on speech perception scores. He based his question on the notion that hard of hearing individuals have an easier time hearing men's voice than women's voice. In his study he used nine speakers; three male adults, three female adults, and three female children. When he compared scores from each group for hard of hearing and normal hearing participants at a fixed intensity level, no significant difference was found across speakers. One goal in the present study is to investigate the possibility of a difference in scores with different dialects. Following Palmer's methodology, in the present study three Jordanian and three Saudi male speakers were selected to complete the word recording.

Stimuli for Dialectal Differences

In the present study the researcher's goal was to investigate the possibility of difference in scores with different dialects since. In this study, a +9 dB signal to noise ratio was used to avoid ceiling effect in normal hearing children's performance and to reduce the variability in scores. The choice of +9 dB signal to noise difference was based on the Goldman et al. (1970) study of speech perception of children in quiet and in noise. Goldman et al. (1970) observed a reduction in scores compared to the quiet condition that started at -9 dB noise level. Resulting word recognition scores are expected to be less than 100% correct. Schwartz & Goldman (1974) used the same level of signal to noise ratio (+9 dB) to investigate the effect of different contexts and listening environments (quiet and noise). They observed a significant increase in number of errors for all contexts when noise was introduced. The effect of the smallest amount of noise was clearly demonstrated in a study by Larson, Petersen, & Jacquot (1974) when they tested the use of NU-6 word lists with children age 5.5 to 6.5 years of age. The presence of noise at +20 dB S/N ratio had a significant effect on the children's performance compared to adult performance under the same conditions. Keep in mind that for their study Larson et al. have used adult material to test very young children. Based on the above listed studies, a signal to noise level of + 9 dB will be used for the present study.

Table 2-1 Children's speech reception tests in English.

Test	Investigator	Stimulus	Response format	Response task	Target population	Published
PBK-50	Haskins 1949	Monosyll. Words	Open set	Verbal	6-9 years	No
GFW	Goldman, Fristoe, & Woodcock 1970	Monosyll. Words	Closed set	Picture pointing	≥ 4 years	Yes
Spondee recognition test	Erber 1974	Spondee words	Closed set	Writing	8-16 years	No
WIPI	Ross & Lerman 1970	Monosyll. Words	Closed set	Picture pointing	3-6 years	Yes
BKB	Bench, Koval, & Bamford 1979	Sentences	Open set	Verbal	8-15 years	No
PSI	Jerger & Jerger 1980	Monosyll words and sentences	Closed set	Picture ointing & verbal	3-10 years	Yes
NU-CHIPS	Elliott & Katz 1980	Monosyll. Words	Closed set	Picture pointing	≥2.5 years	Yes
ANT	Erber 1980	Numbers	Closed set	Picture pointing	3-8 years	No

Table 2-2 Speech recognition word lists' mean dB HL levels at 50% and slopes.

Investigator	Language	Target population	Test	dB @ 50%	Slope %/dB
Wilson & Oyler 1997	English	Adults	CID W-22	15.6	4.8
Wilson & Oyler 1997	English	Adults	NU-6	13.4	4.4
Harris et al. 2003a	Korean	Adults	Male speaker	11.4	5
Harris et al. 2003a	Korean	Adults	Female speaker	10.7	5.1
Nissen et al. 2005a	Chinese Mandarin	Adults	Male speaker	5.4	7.3
Nissen et al. 2005a	Chinese Mandarin	Adults	Female speaker	2.3	8.2
Alusi et al. 1974	Arabic	Adults		22.5	5
Haskins 1949	English	Children	PBK-50	NA	4

Table 2-3 Speech reception threshold wordlists' mean threshold dB HL levels and slopes between 20 and 80%.

Investigator	Language	Target population	Test	dB @ 50%	Slope %/dB
Hirsh et al. 1952	English	Adults	CIDW-1	0.5	8
Yourng et al. 1982	English	Adults	CIDW-2	-0.3	10
Ashoor & Prochazka 1982	Arabic	Adults	Mono-syllabic	2.2	5
Harris et al. 2003b	Korean	Adults	Male speaker	6.8	10.3
Harris et al. 2003b	Korean	Adults	Female speaker	5.6	9
Nissen et al. 2005b	Chinese Manderin	Adults	Male speaker	-0.2	9.7
Nissen et al. 2005b	Chinese Manderin	Adults	Female speaker	-0.7	10.5
Siegenthaler et al. 1954	English	Children	Mono-syllabic	NA	8.6
Ashoor & Prochazka 1985	Arabic	Children	Mono-syllabic	0	5

CHAPTER 3
METHODOLOGY

This dissertation consists of two studies. In the first study, monosyllabic word intelligibility was investigated as a function of presentation level in Jordanian dialect. The following questions were addressed in the first study: Is there a significant difference in word recognition ability of Jordanian children, given increases in intensity presentation levels? Can four parallel word lists be developed (i.e. their psychometric qualities do not differ)? The results of this study were used in the development of four parallel lists of fifty monosyllabic words to be used as a speech perception test in Jordan.

The second study compared word recognition scores of Jordanian children listening to words spoken in Jordanian and Saudi dialects. Specifically, the study was designed to answer the following questions: Do word recognition abilities of Jordanian children differ when listening to words presented in a Jordanian dialect compared to words presented in a Saudi dialect? Do word recognition abilities of Jordanian children who display normal hearing and hearing disabilities differ when hearing words presented in a Jordanian dialect compared to words presented in a Saudi dialect? Does word recognition ability differ when hearing speakers using the same dialect?

Participants

Inclusion criteria included age range and hearing sensitivity. Age range was limited to children age six to nine years. Children displaying normal hearing sensitivity and mild to moderately severe hearing loss were recruited. Hearing status was determined using pure tone average thresholds of 500, 1000 and 2000 Hz. A pure tone average threshold of 15 dB HL or less were considered normal hearing and thresholds greater than 26 dB HL and less than 69 dB HL were considered within the range of mild to moderately severe hearing loss. Tympanometry and pure tone audiometry were used to determine hearing sensitivity fitting inclusion criteria. If the tympanogram showed normal middle-ear function, the researcher proceeded to pure tone audiometry using supra aural headphones (TDH-39). The participant was excluded if middle-ear function was abnormal. If the pure tone average did not meet criteria, the testing was terminated. In both cases results were explained to the parent if present. In case of hearing loss detection or abnormal middle-ear function, the caregiver or adult accompanying the participant was informed and briefly counseled on intervention.

Participants were recruited by word of mouth from the larger Amman area. Participating children came to the clinic in Amman accompanied by their caregivers or an adult with the parent's permission. Twenty normal hearing participants, ten males and ten females, were included in the first study. A total of ten normal hearing, four males and eight female, and ten hard of hearing, six male and four female, participated in the second study. For detailed description of participant recruitment see Appendix A.

Speakers

Speech materials were recorded by native Arabic speakers who were recruited from the area of Provo, Utah, the location of Brigham Young University where the recording took place. Potential speakers were initially personally contacted by phone or by word of mouth. The purpose of the study and the speaker's role in the study were explained. Five adult Jordanian males agreed to participate in the recording. Preliminary 5-minute recordings of continuous speech were made for the purpose of judgment of dialect and clarity of speech. Five different adult Jordanian natives (three females and two males) were asked to judge the speakers using two criterion, i.e., the dialect was a general Jordanian dialect, and the ease of understanding the speaker's speech as rated on a 10-point scale. The top ranked speaker completed the recording of the 304 monosyllabic words for study one. For the completion of recordings for study two the 3 top ranked Jordanian speakers were selected and agreed to participate in the recording. Only three Saudi speakers agreed to participate in the recording for study two.

For the purpose of creating the 8 multi-speaker babble noise, four additional female Arabic speakers were recruited. The female speakers included the primary investigator and three females recruited form the Provo area by word of mouth. The recordings from the 2 male Jordanian speakers who were not included in the final recording and 2 of the speakers used for word recording randomly selected as well as the 4 female speakers were used to compose the noise track.

All speakers have acquired Arabic in their home country. They were all students or spouses of students who have come to the USA to peruse higher education degrees. All speakers have been in the USA for no longer than 5 years.

Recording and Editing

The recording took place in an anechoic chamber at Brigham Young University campus in Provo, Utah, USA. The chamber has a 0 dB SPL noise floor allowing for a recording of 60 dB signal to noise ratio (S/N). The equipment used in the recording was the same as reported by Harris et al. (2003a) and Nissen et al. (2005a). The recording microphone used was a Larson-Davis model 2541, which has a flat frequency response up to 20 kHz. The microphone was positioned approximately 6 inches from the speaker at a 0° azimuth and covered by a 3 inch windscreen. The microphone signal was amplified by a Larson-Davis model 900B microphone preamp, coupled to a Larson-Davis model 2200C preamp power supply. The signal was digitized at a 44.1 KHz sampling rate with 24-bit quantization by an Apogee AD-8000 analog-to-digital converter, with preamp gain set to utilize the full range of the 24-bit analog-to-digital converter. The digitized signal was stored on a hard drive for later editing. The speech material was printed in large font on A4 paper and clipped to a clipboard that was suspended in front of the speaker at a comfortable height to allow him to read the words facing the microphone at 0° azimuth. The speaker was asked to say each word a minimum of four times. A native judge (the primary investigator) rated each word for clarity and the best production was selected for inclusion in the word recognition test. The intensity of each selected word was edited using Sadie Disk Editor software (Studio Audio & Video Ltd. 2004) to produce a final recording with the same average root mean square (RMS) power as the 1 kHz calibration tone in an attempt to equate the test word threshold audibility (Harris et al. 2003a; Nissen et al. 2005a; and Wilson and Strouse, 1999). The recording was converted from 24- to 16-bit quantization using the NS high dither option in the Sadie Disk Editor software. The final word recording was digitally saved as wav files. Each word was saved as a separate file. A total of 250 words were selected to be included in the word recognition test; study one. For study two 33 monosyllabic words were selected from Ashoor and Prochazka's (1985) speech reception threshold list for children. The words were edited digitally to ensure same level of RMS as the 1 kHz tone.

The multi speaker babble material was completed using 4 of the male Jordanian speakers and 4 Jordanian female speakers. The speakers were asked to read an article taken form a Jordanian daily newspaper. The recording was completed following the same procedure as for the word list. Three minutes of continuous speech was selected for each of the 8 speakers. All silence between words and sentences was removed for each speaker. Then the RMS level for each speaker was adjusted to have all 8 speakers at the same level. Next the recordings were digitally mixed and run through a compression/limiter to limit the range of peak variability. The resulting recording average RMS was measured and adjusted to equal the level of the 1 kHz calibration tone. The 8 multi-speaker babble noise was saved in a separate wav file that would allow the play of noise independently from the word lists.

Instrumentation for Data Collection

Data collection took place at the Middle East Hearing Association clinic in Amman, Jordan in a sound treated booth designed locally for hearing testing purposes. Pure tone and speech audiometry were conducted using an Interacoustics AC40 audiometer, connected to TDH-39 headphones. The audiometer was calibrated using a Larson Davis System 824 sound level meter and a 6cc coupler. Calibration was based on ANSI standards 2004. Measurements of sound levels at octave and half octave frequencies met the ANSI standard 2004 with a deviation range of -0.6 to +0.3 dB. The sound levels for speech through external input A and external input B were consistent with ANSI standard 2004 with a deviation of -0.5 to +0.1 dB. Repeated measurements of sound pressure level produced by the audiometer were within permissible ANSI tolerance level of ± 3dB for frequencies of 500 to 4000Hz and ±5dB for 6000 to 8000 Hz.

Special software was developed by Dr. Richard Harris at Brigham Young University (2005) to control the playback of 1 kHz tone, noise and word lists from wav files. The software also provides the documentation of data in an excel file spread sheet with the following details: the date and time of presentation, participant assigned number, participant gender, test ear, speaker gender, intensity level, signal to noise ratio, list name, time of recording per list, wav file, word (in this case in Arabic), and the score.

Prior to data collection, the VU meter was adjusted to 0 VU using 1 KHz tone. For the purpose of consistency in speech audiometry data collection, the left headphone was randomly selected to always be used over the test ear. This step was taken to reduce variability in the sound level presented.

Study One

Speech Material

Speech material was selected following Egan's (1948) criterion of monosyllabic words, representative of spoken Arabic and commonly used words. A total of 304 monosyllabic words were selected for initial recording. Words that have similar forms in colloquial and standard Arabic were selected. Children and teachers of first through third grade rated these words as familiar. Rating took place at four elementary schools in Amman Jordan. The words were read aloud to the children, they were asked to raise their hand if they knew the meaning of the word and were able to use it in a meaningful sentence. To ensure the accuracy of their response, the primary investigator (Nadia Abdulhaq) started with three trial words that were familiar to children (window, door, and table), randomly asked children to put words in a sentence, in addition unfamiliar words (from old classical text) and nonsense words were included. For more detailed description of word selection see Appendix B.

The primary investigator judged the recorded 304 words for clarity of recording and in conjunction with familiarity ratings selected 250 words to be used in data collection. The 250 words were divided randomly into two sets of ten lists of 25 words. To allow the presentation using the software developed by Dr. Harris, a text file was created for each list A1 to A10 and B1 to B10 for each set of randomization respectively. Lists were presented in counterbalanced order. Lists A1 through A10 were presented to participants 1 through 10 and lists B1 through B10 were presented to participants 11 through 20. See Table 3-1 for order of presentation and level of presentation. The signal was routed to the subjects through the Interacoustics AC40 audiometer, via the TDH-39 headphones. Speech material was presented at 10 different intensity levels starting at 0 dB HL and increasing by 5 dB increments up to 45 dB HL; to include a range of low to high word recognition scores.

Procedure for Data Collection

Each participant was given the following general instructions:

"You will hear some words through the headphones, your will repeat the words you hear. For example, if you hear the word /kitab/ (book), you would say?" The researcher would wait for the participant to repeat the word. "That is good. And if you hear the word /daftar/ (copy book) you would say? and the word /madrasa/ (school) you would say?" Once the participant has followed instructions further details were provided:

"The words will be spoken very softly; you might or might not hear them at first. That is OK. If you hear a word, repeat it. If you don't, just wait to hear the next one. The words will get louder and louder."

The earphone was placed over the test ear. Two lapel microphones were attached to the participant's shirt, one for talk back and the second for audio input of the audio-visual recording. A video camera was set in front of the participant to obtain video recording. The video camera was adjusted in height and angle to provide the best view of the participant's face.

After the participant repeated a word, the investigator entered 1 for a correct repetition and 0 for an incorrect repetition. If the participants did not repeat the word, the investigator periodically used the talk forward microphone to ask the participant if he/she had heard the word, and to encourage him/her to repeat whatever they heard. The investigator waited 3to 5 seconds before recording an incorrect score in the case of no repetition. After each list presentation, the presentation level was increased by 5dB. Participants were encouraged between lists using verbal praise ("good job", or "you are doing really well"). Most participants did not need a break or rest period throughout the data collection procedure. All data for each participant, including presentation order and scoring details, were saved in excel spreadsheets.

Statistical Analysis

The dependent variable, defined as correct word recognition, was saved in binary format (correct versus incorrect). The independent variables were different presentation intensity levels and different word lists. The raw data were used in logistic regression analysis. The logistic regression analysis provides derived variables that are key parameters for the regression line that may be tested and compared between conditions (in this study, the word lists). The raw scores of each word in the four lists were used in the logistic regression analysis to calculate the logistic parameters for each list and half list. See Appendix C for details on logistic regression analysis.

Recall that this study was designed to test the following hypotheses: 1. Increased presentation sound intensity level will improve word recognition ability. 2. It is possible to create four word lists that are parallel in function and can be used interchangeably to test Jordanian children's word recognition ability.

Study Two

Speech Material

The speech material consists of 33 monosyllabic words from Ashoor's (1985) children's list. Ashoor's list was used as source of material because it was developed for the purpose of Arabic speech perception specifically for children. The selection of the 33 words was based on two criteria: monosyllabic words and familiar to Jordanian children. Recall that Ashoor's list consisted of a mixture of mono- and di-syllabic words, familiar to Jordanian Arabic speaking children. The familiarity of words to Jordanian Arabic speaking children was established through ratings provided by children in 1st through 3rd grades (Abdulhaq, unpublished). See Appendix D for the list of 33 words and their meaning in English.

A total of six recordings, each recorded by a different speaker (three Jordanian and three Saudi speakers), were presented to all participants in counterbalanced order at constant intensity levels (see table 3-2 for details). The speech material was presented by playing wav files from a standard desktop PC connected to external inputs of the audiometer using the same software as described in study one.

The words were presented to the hard of hearing participants at an audible level of 40 dB SL (re: the participant's pure tone average). For normal hearing participants, the multi-speaker babble track was presented then the word list was presented. Words were presented at 50 dB HL and the multi-speaker babble noise at 41 dB HL, i.e., with a signal to noise ratio (SNR) of +9 dB.

Statistical Analysis

For the purpose of statistical analysis of the data in these studies, the dependent variable was defined as the correct word recognition score, and the independent variables were defined as dialect, speakers, and hearing status. Raw scores were used in mixed ANOVA to determine difference between speakers and dialects. To determine the effect of dialect on normal hearing and hard of hearing children's scores a paired t-test of dialects based on scores from each group separately was used. To determine the significance of differences between speakers within a dialect paired t-tests of speakers were used based on scores from all participants.

Table 3-1 Order of list presentation by participant and level of presentation

Participant	**Presentation dB HL level**									
	0	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**
1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
2	A2	A3	A4	A5	A6	A7	A8	A9	A10	A1
3	A3	A4	A5	A6	A7	A8	A9	A10	A1	A2
4	A4	A5	A6	A7	A8	A9	A10	A1	A2	A3
5	A5	A6	A7	A8	A9	A10	A1	A2	A3	A4
6	A6	A7	A8	A9	A10	A1	A2	A3	A4	A5
7	A7	A8	A9	A10	A1	A2	A3	A4	A5	A6
8	A8	A9	A10	A1	A2	A3	A4	A5	A6	A7
9	A9	A10	A1	A2	A3	A4	A5	A6	A7	A8
10	A10	A1	A2	A3	A4	A5	A6	A7	A8	A9
11	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
12	B2	B3	B4	B5	B6	B7	B8	B9	B10	B1
13	B3	B4	B5	B6	B7	B8	B9	B10	B1	B2
14	B4	B5	B6	B7	B8	B9	B10	B1	B2	B3
15	B5	B6	B7	B8	B9	B10	B1	B2	B3	B4
16	B6	B7	B8	B9	B10	B1	B2	B3	B4	B5
17	B7	B8	B9	B10	B1	B2	B3	B4	B5	B6
18	B8	B9	B10	B1	B2	B3	B4	B5	B6	B7
19	B9	B10	B1	B2	B3	B4	B5	B6	B7	B8
20	B10	B1	B2	B3	B4	B5	B6	B7	B8	B9

Table 3-2 Word list and dialect presentation order.

Participants	**Present. order**					
	1st	2nd	3rd	4th	5th	6th
1	J1	S1	J2	S2	J3	S3
2	S1	J1	J3	S3	S2	J2
3	S2	J2	J1	S1	S3	J3
4	J2	S2	S3	J3	J1	S1
5	S3	J3	S1	J1	J2	S2
6	J3	S3	S2	J2	S1	J1
7	J1	S1	J2	S2	J3	S3
8	S1	J1	J3	S3	S2	J2
9	S2	J2	J1	S1	S3	J3
10	J2	S2	S3	J3	J1	S1

J = Jordanian speaker, S = Saudi speaker. The same order of presentation was used for the normal hearing and hard of hearing group.

CHAPTER 4
RESULTS

Study One

Twenty individuals participated in this study. Pure tone thresholds average (PTA) indicated that all participants fit the criteria of a PTA of 15dB HL or better with a mean PTA of 7.8 dB HL (standard deviation = 2.8, range = 3.3 to 12.5 dB HL). The better ear was selected based on PTA or at random when both ears had the same PTA. Ten participants were tested on the right ear and 10 were tested on the left ear. Average pure tone thresholds for both ears of all participants are shown in Figure 4-1. For detailed description of participants see Table 4-1.

Most participants completed the 250 words in a single session of 20 and 29 minutes without a break other than the time it took to load each new list and adjust the intensity of presentation. The pacing of the test depended on how quickly participants responded; the program required the examiner to record the response before the next word could be presented.

The raw score data of all participants at the 10 intensity levels were compiled for each of the 250 words. The total of responses resulted in a score of correct recognition out of 20 possible correct score per word. Words then were reordered by difficulty from easiest to most difficult based on the total score (with lower scores indicating higher difficulty). Four lists of 50 words each were created using the top 200 words ranked by difficulty, as follows: Words that were tied in difficulty ranking were randomized prior to sorting into lists. To help ensure range of difficulty between lists, the four most difficult words were assigned randomly to four lists. The next four most difficult words again were assigned randomly to the four lists and so on until a total of 50 words were included in each list. The lists were labeled 1, 2, 3, or 4.

Within each list, the word's level of difficulty was used to assign it in order to create two half lists of 25 words from each original 50-item list with the goal of having an equal range of difficulty on both half lists. The half lists were labeled 1A and 1B (the two lists taken from list 1) through 4A and 4B (the two lists taken from list 4). This procedure used, to distribute words into lists, has been successful in developing difficulty equivalent word lists in studies by Harris et al. (2003a and 2003b) and Nissen et al. (2005a and 2005b). The eight 25-Arabic monosyllabic word half lists are presented in Tables 4-2. The words are written in Arabic with their phonetic transcription, please see appendix E for the English translations of the words.

Word recognition scores increased with the increase of intensity. A scatter plot of the percent correct of responses, for the half lists (25 words) as a function of stimulus level, is presented in Figure 4-1. Scores at the lowest intensity level of 0 dB HL ranged from 0 to 2% while scores at the highest intensity level of 45dB HL ranged from 94 to 96%. Chi-square analysis (X^2) of the half lists showed a highly significant effect of intensity (X^2 = 1,134.18, p <0.0001). Statistical analysis of data confirmed, as expected, that scores at different intensity levels are not equal.

Differences between lists (X^2 (3) = 1.67, p =0.64) and between half lists (X^2 (7) = 8.25, p = 0.31) were not significant. The mean slope of full lists at the 50% point was 4.44%/dB and ranged from 4.21%/dB to 4.63%/dB (range = 0.42%/dB). The mean slope at the 20-80% range was 3.84%/dB ranging from 3.65 to 4.01%/dB (range = 0.36%/dB). The mean threshold was 21.25 dB, and ranged from 21.21 dB to 21.32 dB (range = 0.107 dB). For the half lists, the mean slope at 50% was 4.47%/dB and ranged from 3.86 to 5.36%/dB (range = 1.5%/dB). The mean slope at the 20-80% range was 3.87%/dB with a range from 3.34 to 4.64%/dB (range =1.30%/dB). Mean threshold was 21.25 dB, and ranged from 21.04 dB to 21.38 dB (range = 0.34 dB). Table 4-3 and Table 4-4 include the detailed results of the logistic regression parameters, the slopes at the 50% and 20-80% levels, threshold intensity level, as well as threshold deviation from mean threshold, for the lists and half lists respectively. Figure 4-3 shows the psychometric function for all four lists and Figure 4-4 shows the psychometric function for the eight half lists. By visual inspection, the curves are almost identical.

The intensity levels at the threshold point had a narrow range around the mean value, from -0.03 to 0.11 dB for the full lists and from -0.21 to 0.13 dB for the half lists. The difference between the calculated threshold and the mean threshold for each list and half list was used to adjust the intensity of each list and half list digitally to achieve a better equivalence between lists and half lists. The percent correct was recalculated based on these adjustments for all lists and half lists and an adjusted logistic regression curve was created. Figure 4-5 shows the difference between the unadjusted and adjusted curves. The difference between the unadjusted and adjusted curves was small and barely detectable.

Study Two

Ten normal hearing and 10 hard of hearing individuals participated in this study. Pure tone threshold average measures indicated that all participants fit the criteria of a PTA of 15dB HL or better with mean PTA of 6.3 dB HL (SD = 2.8, minimum = 1.7dB

HL and maximum = 10dB HL). The better ear was selected based on PTA or at random when both ears had the same PTA; as a result 6 participants were tested on the right ear and 4 were tested on the left ear. Figure 4-6 shows the average of pure tone thresholds of normal hearing participants, for detailed description see Table 4-5.

Pure tone threshold average measures indicated that all ten hard of hearing participants had a mild to moderately severe hearing loss with a mean PTA of 50.2 dB HL (SD = 11.8, minimum = 31.7dB HL and a maximum = 63.3dB HL). All but two participants had sensorineural hearing loss (SNHL). One participant (#2 in table 4-6) had chronic conductive hearing loss and one participant (#7 in table 4-6) had a mixed loss. Again the better ear was selected based on PTA or at random when both ears had the same PTA; resulting in 5 right and 5 left ears being used. Figure 4-7 shows the average of pure tone thresholds of hard of hearing participants, for detailed description see Table 4-6.

Most participants completed all six lists of 33 words in a range of 13 to18 minutes, and did not require a break in data collection. The pacing of data collected depended on how quickly participants responded; the program required the examiner to record the response before the next word could be presented. After data collection was completed the score for each participant per speaker was calculated into percent correct. See Table 4-7 for detailed results. These data were analyzed to evaluate the hypotheses.

Mixed analysis of variance (ANOVA) indicated a significant difference between dialects and speakers. Scores of all participants for all speakers were included in a mixed ANOVA of speaker and dialect. In summary, analysis showed a significant difference between dialects ($F= 8.865$, $p = 0.008$), significant differences among speakers (F =

6.181; p = 0.005), and a significant difference between speakers within each dialect (F = 5.153; p = 0.011). For detailed results see Table 4-8.

Paired sample t-test analysis indicated significant difference between dialects for the normal hearing group, but no significant difference for the hard of hearing group. The scores for each participant were averaged by dialect, and the averages were used in paired sample t-test for each group separately (hard of hearing and normal hearing). Difference between dialects for the normal hearing participant group were significant (t-test = -2.923 p = 0.017). Differences between dialects for the hard of hearing participant group were not significant (t-test = -1.327 p = 0.217). Thus, hearing loss seemed to outweigh dialectal difference. See Table 4-9 for detailed results.

Within each dialect one speaker was significantly different than the other two speakers. A paired sample t-test was performed comparing results of speakers within each dialect for all participants. Results indicated that scores associated with speaker J1 (in the Jordanian dialect) were higher than those for J2 and J3 (t = 4.203 and 3.802, respectively; p = 0.0001 and 0.001). Paired scores associated with speakers J2 and J3 were not different (t-test -0.496, p= 0.625). For the Saudi dialect, scores associated with speaker S2 were significantly lower than those for S1 and S3 (t-test = 2.658 and -2.183 respectively and p = 0.016 and 0.042). Differences between speakers S1 and S3 were not significant (t-test = 0.047 and p = 0.962). See Table 4-10 for detailed results.

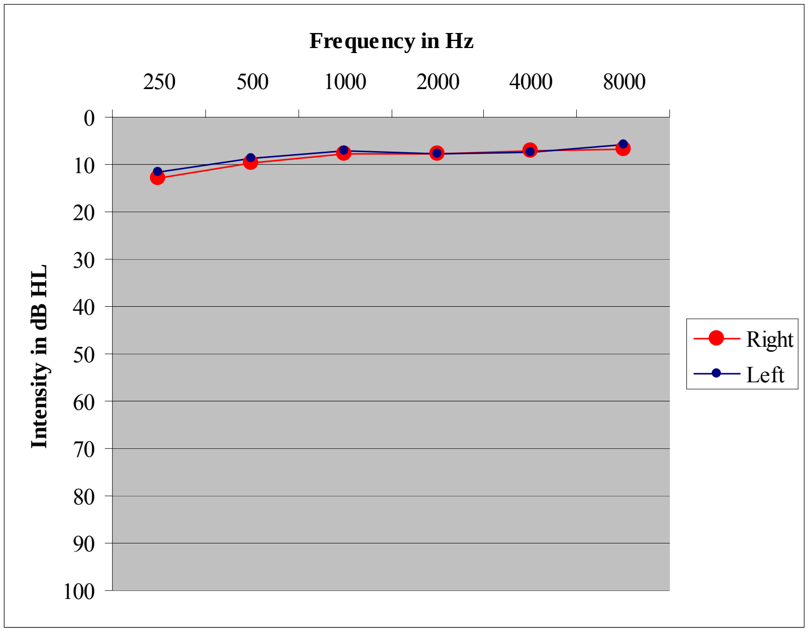

Figure 4-1 Average pure tone thresholds at all frequencies for all 20 normal hearing participants at 250, 500, 1000, 2000, 4000, and 8000 Hz.

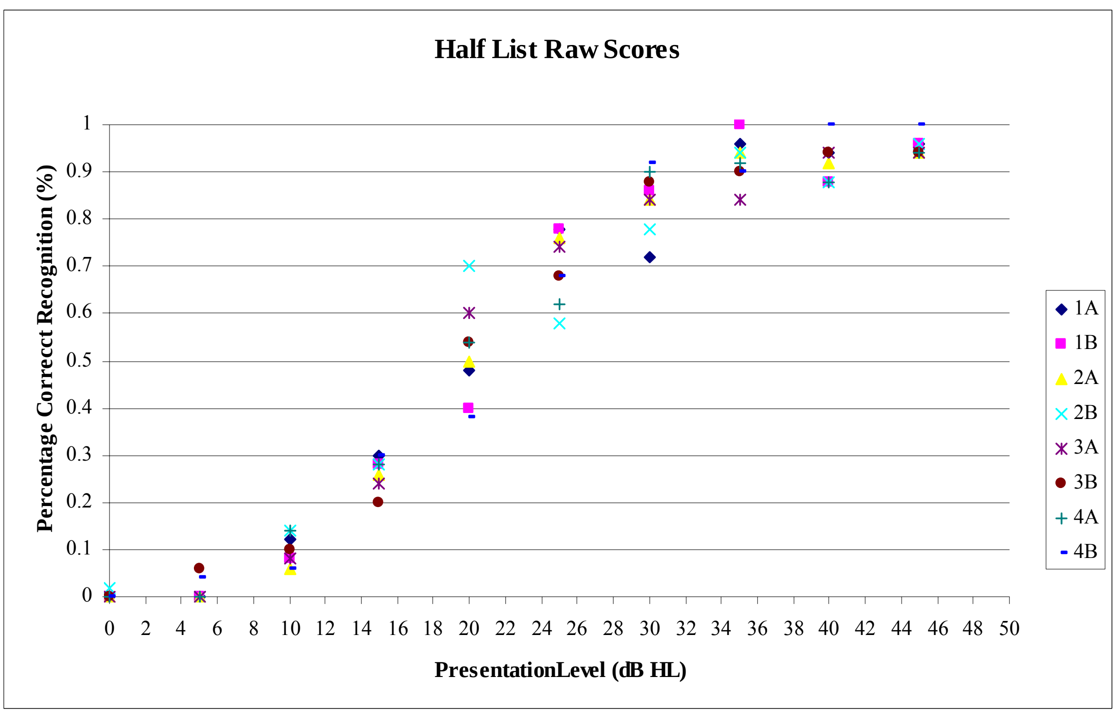

Figure 4-2 Half lists raw data scatter plot

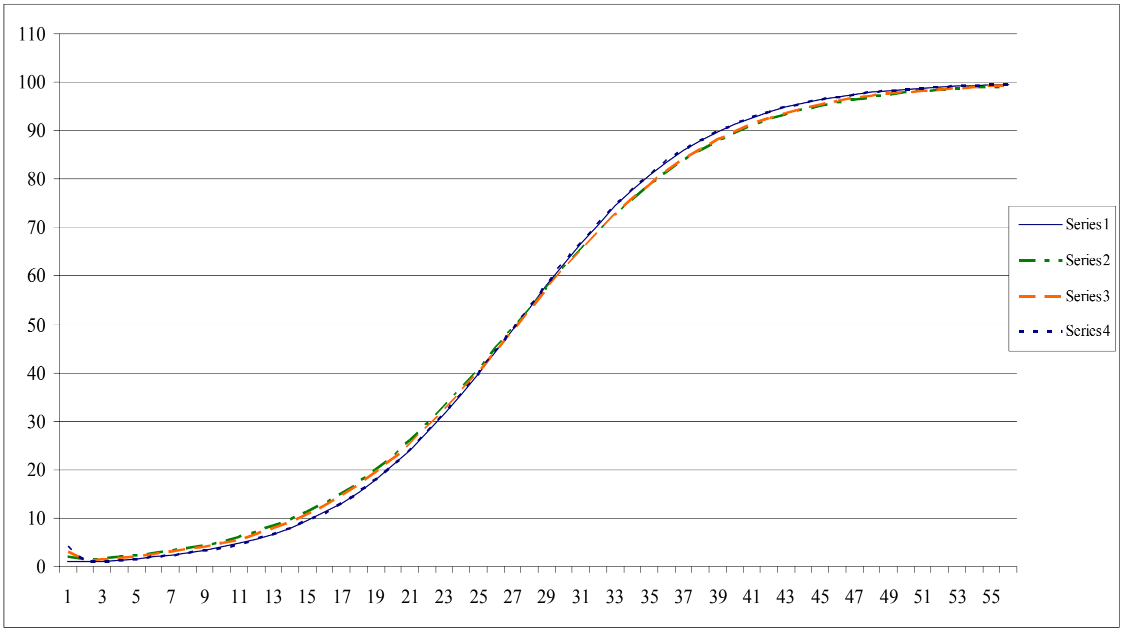

Figure 4-3 Psychometric function of 4 lists of 50 words based on calculated percent correct.

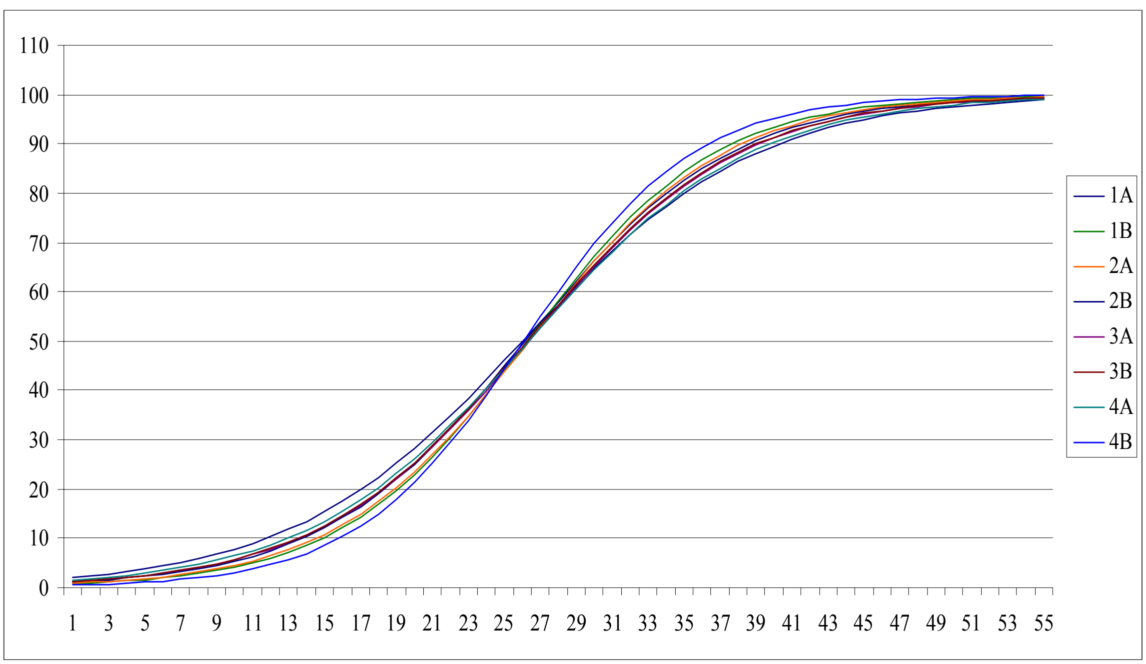

Figure 4-4 Psychometric function of 8 half lists of 25 words based on calculated percent correct.

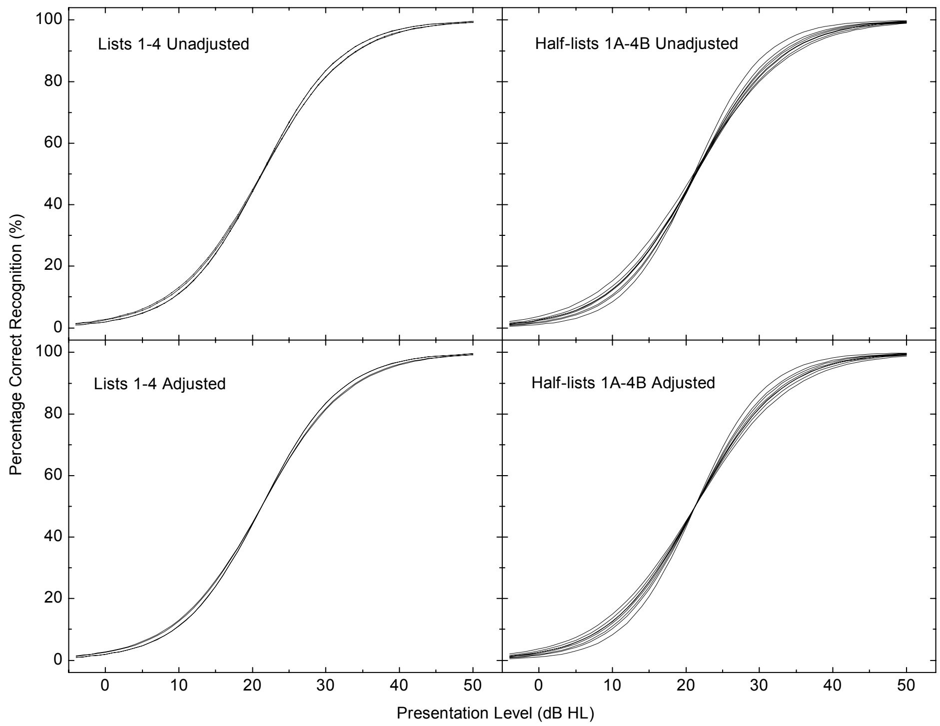

Figure 4-5 Arabic monosyllabic psychometric functions for lists 1-4 (left) and half-lists 1A-4B (right). The top two figures depict unadjusted psychometric functions and the bottom two figures depict psychometric functions adjusted for equal performance at 50% correct recognition.

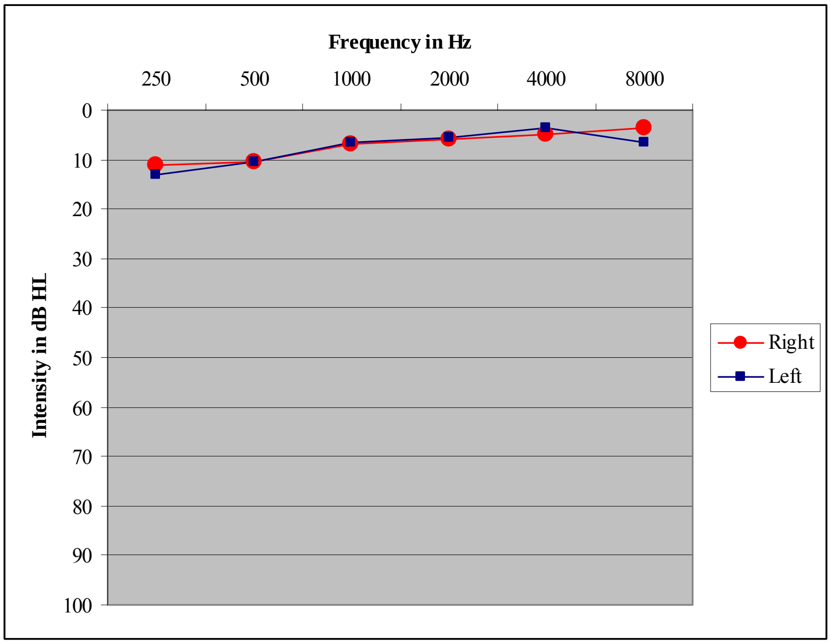

Figure 4-6 Average of pure tone thresholds of normal hearing participants at 250, 500, 1000, 2000, 4000, and 8000 Hz.

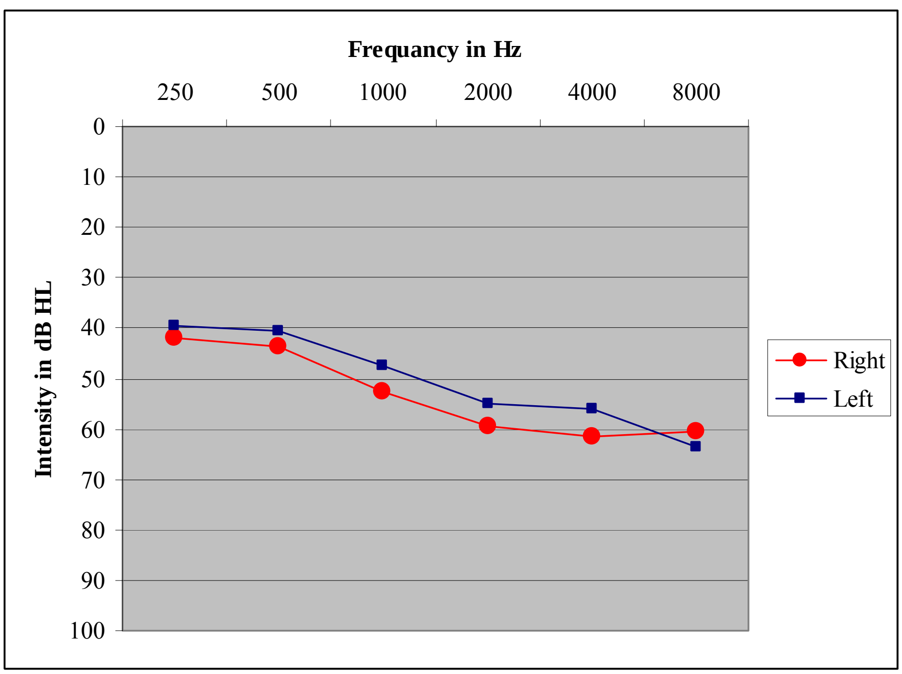

Figure 4-7 Average of pure tone thresholds of hard of hearing participants at 250, 500, 1000, 2000, 4000, and 8000 Hz.

Table 4-1 Normal Hearing participants' age, gender, test ear and PTA

Participant	Gender	Age	Test ear	PTA
4	F	6	Left	6.7
8	F	6	Right	10.8
10	F	6	Right	5.0
5	F	7	Left	9.2
7	F	7	Left	9.2
12	M	7	Right	3.3
15	M	7	Right	7.5
16	M	7	Right	5.8
19	F	7	Right	5.8
20	M	7	Right	6.7
6	F	8	Right	12.5
11	M	8	Right	4.2
1	M	9	Left	6.7
2	M	9	Left	4.2
3	F	9	Right	9.2
9	F	9	Left	9.2
13	M	9	Left	10.8
14	M	9	Left	6.7
17	M	9	Left	11.7
18	F	9	Left	11.7
Mean		7.8		7.8
Standard Deviation		1.2		2.8
Range		3.0		9.2
Minimum		6.0		3.3
Maximum		9.0		12.5

Table 4-2 Half word lists and transcription in IPA.

1A		2A		3A		4A	
خَير	khair	رَب	rab	مَرج	marʒ	شَمس	ʃams
كوع	ku:ʕ	حاج	ħa:ʒ	خَيل	χeɪl	تاج	ta:ʒ
بيت	bɛ:t	زر	zir	جِن	ʒin	تين	ti:n
رأس	rɑ:s	لَف	laf	حال	ħa:l	بئر	bi:r
سوق	su:ʔ	نار	nɑ:r	رَد	rad	طوب	t̠u:b
لوز	lo:z	شَد	ʃad	طول	t̠u:l	تحت	taħt
خال	χa:l	غاز	ɤa:z	بَنك	bank	عَكس	ʕaks
كأس	ka:s	طير	t̠ɛ:r	حوت	ħu:t	سام	sa:m
بَرد	bard	كُم	kum	ريح	ri:ħ	جيب	ʒeib
أهل	ʔahl	حُر	ħur	صوص	s̠u:s̠	كوخ	ku:χ
باص	bɑ:s̠	سَد	sad	عَين	ʕein	حَب	ħɑb
صَيد	s̠ɑid	خوف	χo:f	جاف	ʒa:f	كَلب	kalb
مَوت	mo:t	عود	ʕu:d	سَطح	sɑt̠ħ	سيخ	si:χ
زَوج	zo:ʒ	عِز	ʕiz	روح	ru:ħ	عَد	ʕad
سود	su:d	صب	s̠ab	رَف	raf	حَوض	ħo:d̠
غَرب	ɤarb	مَجد	maʒd	شَخص	ʃaχs̠	زير	zi:r
ثوب	θo:b	شاي	ʃa:y	بُرج	burʒ	رَمز	ramz
واد	wa:d	شَيخ	ʃɛ:χ	بَر	bar	صعب	s̠ɑʕb
وز	waz	صين	s̠i:n	بِنت	bint	بُعد	buʕd
ثَلج	θalʒ	دار	dɑ:r	طَقس	t̠aqs	درس	dars
فيل	fi:l	دود	du:d	أرض	ʔɑrd̠	مُر	mur
شيء	ʃeɪʔ	شَعر	ʃɑʕr	وَرد	ward	ضَيف	d̠eif
زيد	zeɪd	قَلب	qɑlb	أنف	ʔanf	ألف	ʔalf
شَط	ʃɑt̠	ريش	ri:ʃ	حَق	ħɑq	سَبت	sabt
صوت	s̠o:t	دين	di:n	حُب	ħub	بطّ	bɑt̠

Table 4-2 continued

1B		2B		3B		4B	
خس	χas	فرض	fard̠	بيض	beid̠	جِد	ʒid
جَد	ʒad	ثوم	θo:m	ثور	θo:r	خيط	χeit̠
كَف	kaf	بَيع	bɛ:ʕ	موز	mo:z	ظِل	ð̠il
نِصف	nus̠	خَلف	χalf	تخت	taχt	حَل	ħal
لمس	lams	جَيش	ʒɛɪʃ	عيد	ʕi:d	إسم	ʔism
نَفس	nafs	عام	ʕa:m	حَرب	ħɑrb	عَرش	ʕarʃ
أم	ʔum	زَي	zay	قلب	qɑlb	سيف	seif
روس	rʊ:s	حَجّ	ħaʒ	زاد	za:d	فول	fu:l
بوق	bu:q	سَير	sɛɪr	شَر	ʃɑr	جار	ʒɑ:r
لوح	lo:ħ	سِر	sɛ:r	لِص	lɪs̠	فأس	fa:s
عَرض	ʕɑrd̠	جاد	ʒa:d	أب	ʔab	شاش	ʃa:ʃ
شوك	ʃo:k	يوم	jwm	حار	ħɑ:r	دَرب	darb
حَر	ħɑr	رُز	ruz	غَوص	ɣo:s̠	كرش	karʃ
توت	tu:t	جَو	ʒaw	خط	χat̠	سور	su:r
شَك	ʃak	جوع	ʒu:ʕ	نور	nu:r	ثُلث	θuleθ
زيت	zɛ:t	ريف	ri:ħ	قوس	qo:s	فار	fɑ:r
ديك	di:k	مال	ma:l	ضوء	d̠aw	دَم	dam
دور	do:r	شب	ʃab	ناس	na:s	أخت	ʔuχt
حِزب	ħɪzb	أخ	ʔaχ	دب	dub	حَد	ħad
بوت	bo:t	ريم	ri:m	لُب	lub	نوم	no:m
أنس	ʔuns	خاص	χɑ:s̠	خوخ	χo:χ	حين	ħi:n
حَي	ħaj	صَف	s̠af	هَم	ham	لون	lo:n
قِرد	qɪrd	عُش	ʕuʃ	شَرق	ʃarq	فَرد	fard
كنز	kanz	صوف	s̠u:f	دُش	duʃ	كيس	ki:s
شيب	ʃɛ:b	مَد	mad	عَم	ʕam	جِلد	ʒild

Table 4-3 Logistic regression results for the full lists, the calculated slopes at the 50% and the 20-80% levels, the threshold dB level, and difference of threshold levels from the mean threshold

List	a	b	Slope at 50%	Slope from 20-80%	Threshold dB	dB difference
1	3.918	-0.185	4.614	3.994	21.225	-0.021
2	3.574	-0.169	4.212	3.646	21.214	-0.033
3	3.663	-0.172	4.295	3.718	21.321	0.074
4	3.930	-0.185	4.629	4.007	21.226	-0.020
M	3.771	-0.178	4.438	3.841	21.246	0.000
Minimum	3.574	-0.185	4.212	3.646	21.214	-0.033
Maximum	3.930	-0.169	4.629	4.007	21.321	0.074
Range	0.356	0.017	0.416	0.361	0.107	0.107
SD	0.180	0.009	0.215	0.186	0.050	0.050

Table 4-4 Logistic regression results for the half lists, the calculated slopes at the 50% and the 20-80% levels, the threshold dB level, and difference of threshold levels from the mean threshold

Half Lists	a	b	Slope at 50%	Slope from 20-80%	Threshold dB	dB difference
1A	3.748	-0.177	4.426	3.832	21.170	-0.077
1B	4.104	-0.193	4.821	4.173	21.281	0.034
2A	3.969	-0.186	4.642	4.018	21.379	0.132
2B	3.249	-0.154	3.860	3.341	21.040	-0.206
3A	3.663	-0.171	4.285	3.709	21.370	0.123
3B	3.663	-0.172	4.305	3.726	21.272	0.025
4A	3.498	-0.164	4.093	3.543	21.363	0.117
4B	4.522	-0.214	5.361	4.640	21.087	-0.159
M	3.802	-0.1790	4.47	3.87	21.25	0.00
Minimum	3.249	-0.2144	3.86	3.34	21.04	-0.21
Maximum	4.522	-0.1544	5.36	4.64	21.38	0.13
Range	1.273	0.0600	1.50	1.30	0.34	0.34
SD	0.392	0.0187	0.47	0.40	0.13	0.13

Table 4-5 Selected characteristics of normal hearing participants

Participant	Gender	Age	Test Ear	PTA
14	M	6	Right	5.0
22	M	6	Left	10.0
10	F	7	Right	8.3
11	M	7	Right	8.3
3	F	8	Left	3.3
4	F	8	Left	1.7
5	F	8	Left	10.0
6	M	8	Right	5.0
20	F	8	Right	6.7
21	F	9	Right	5.0
Mean		7.5		6.3
Standard Deviation		1.0		2.8
Range		3.0		8.3
Minimum		6.0		1.7
Maximum		9.0		10.0

Table 4-6 Selected characteristics of hard of hearing participants

Participant	Gender	Age	Test Ear	PTA	Hearing Loss Type
8	F	6	Right	63	SNHL
16	M	6	Left	48	SNHL
18	M	6	Left	62	SNHL
2	F	8	Right	32	Conductive
7	M	8	Right	52	Mixed
1	F	9	Right	32	SNHL
15	M	9	Left	58	SNHL
17	M	9	Left	60	SNHL
19	F	9	Right	53	SNHL
23	M	9	Left	42	SNHL
Mean		7.9		50.2	
Standard Deviation		1.4		11.8	
Range		3.0		31.7	
Minimum		6.0		31.7	
Maximum		9.0		63.3	

Table 4-7 Word recognition scores in percent correct for each subject per talker-talker and the average of scores by dialect.

Group	Parti-cipant	J1	J2	J3	S1	S2	S3	Average Jordanian	Average Saudi
1	1	0.94	0.91	0.97	0.94	1.00	0.97	0.94	0.97
1	2	0.94	0.94	0.97	0.85	0.88	0.97	0.95	0.90
1	8	0.58	0.30	0.42	0.45	0.24	0.42	0.43	0.37
1	15	0.82	0.76	0.67	0.82	0.73	0.61	0.75	0.72
1	16	0.70	0.70	0.45	0.61	0.61	0.55	0.62	0.59
1	17	0.70	0.52	0.67	0.70	0.61	0.73	0.63	0.68
1	18	0.58	0.64	0.52	0.79	0.79	0.64	0.58	0.74
1	19	0.70	0.48	0.82	0.76	0.61	0.82	0.67	0.73
1	23	0.97	0.64	0.85	0.85	0.94	0.94	0.82	0.91
1	7	0.67	0.39	0.67	0.70	0.52	0.76	0.58	0.66
2	6	0.94	0.55	0.76	0.82	0.73	0.85	0.75	0.80
2	5	0.82	0.73	0.82	0.79	0.73	0.79	0.79	0.77
2	4	0.85	0.61	0.79	0.67	0.79	0.97	0.75	0.81
2	3	0.70	0.64	0.64	0.79	0.82	0.88	0.66	0.83
2	10	0.88	0.79	0.58	0.91	0.67	0.52	0.75	0.70
2	11	0.79	0.82	0.42	0.85	0.67	0.82	0.68	0.78
2	22	0.82	0.70	0.55	0.88	0.85	0.76	0.69	0.83
2	14	0.94	0.97	0.88	0.97	0.91	0.97	0.93	0.95
2	20	0.82	0.58	0.73	0.88	0.64	0.91	0.71	0.81
2	21	0.85	0.82	0.73	0.82	0.82	0.94	0.80	0.86

Group1 = hard of hearing, Group 2= normal hearing, J1= Jordanian talker 1, J2= Jordanian talker 2, J3= Jordanian talker 3, S1= Saudi talker 1, S2= Saudi talker 2, S3 = Saudi talker3

Table 4-8 Mixed ANOVA results comparing dialects for all participants' scores

	df	Mean Square	F-test	Sig.
Dialect	1	0.066	8.865	0.008
Talkers	2	0.091	6.181	0.005
Talker within dialect	2	0.027	5.153	0.011

Table 4-9 Paired Sample T-tests comparing dialectal difference in the hard of hearing group and normal hearing group.

Group	df	T-test	Sig. (2-tailed)
Hard of Hearing	9	-1.327	0.217
Normal Hearing	9	-2.923	0.017

Table 4-10 Paired Sample T-test results including all participants' scores

Pairs	**df**	**T-test**	**Sig. (2-tailed)**
J1-J2	19	4.203	0.0001
J1-J3	19	3.802	0.001
J2-J3	19	-0.496	0.625
S1-S2	19	2.658	0.016
S1-S3	19	0.047	0.963
S2-S3	19	-2.183	0.042

CHAPTER 5
DISCUSSION

The primary goal of the present study was to create four equivalent lists of words for clinical assessment of speech perception in Jordanian Arabic speaking children. There are no published word lists developed for the specific purpose of clinical speech audiometry (word recognition) for Jordanian Arabic speaking populations. Although other speech recognition materials have been developed and published over the past 3 decades, no study has evaluated the feasibility of applying Arabic speech audiometry materials among different countries or different Arabic language dialects (e.g., the use of Egyptian Arabic word lists in Jordan). In addition, recorded materials are not widely distributed and, thus, not commonly used in audiology clinics within Arabic countries.

Four lists of 50 words were developed in the present study. The lists were not significantly different in their psychometric functions. All had the same shape of curve and very similar slopes (mean 4.4 %/dB, S.D. 0.215) and threshold points (mean 21.25 dB, S.D. 0.05), properties that are appropriate for clinical use as speech recognition measures. The word lists developed in the present study were comparable in slope and curve shape to word lists published in English (Haskins, 1949; Wilson & Oyler, 1997), Korean (Harris et al. 2003a), and other words lists published in Arabic (Alusi et al. 1974). The point of threshold (50% correct word recognition) was comparable to the Arabic adult lists reported by Alusi et al. (1974). The point of threshold for the Arabic word lists in the present study differed from other languages, suggesting the likelihood of a language specific feature. The range of thresholds for the word recognition materials in the present study (from 2.3 to 22.5 dB HL), confirm the appropriateness of the materials as measures of word recognition ability.

There was no attempt in the present study to develop words lists that were phonetically balanced. Carhart (1965) stated that the phonetically balanced CNC lists and W-22 lists approximated the phonetic balance of every day spoken English but cannot be true representation of everyday phonetic balance. The effect of familiarity of words on speech perception performance is greater than the phonetic balance of the word lists. Owens (1961) studied the effect of word familiarity on word recognition. He found that listeners were more likely to make errors on less familiar words and, when they made the errors, the listener's response was more likely to be a familiar word. Martin (2000) stated that phonetic balance is not the only, or the main, factor in word list equivalence. Therefore, an objective of the present study was to utilize familiar words with a balanced range of difficulty between lists. In the present study, the psychometric curves of the full lists and half lists were very similar. Statistical analyses of the full lists ($X^2(3) = 1.67$, $p = 0.64$) and half lists ($X^2(7) = 8.25$, $p = 0.31$) showed no significant difference between word lists and half lists despite the fact that no effort was made to account for phonetic balance.

In comparison to the speech materials reported previously (Alusi et al., 1974; Ashoor & Prochazka, 1982, and 1985) the recordings developed in the present study can easily be applied clinically by audiologists in Arabic-speaking patient populations with a CD player and distributed to audiologists throughout the Arab countries. However, the issue of dialectal differences requires further investigation. Specifically, the possible effect of dialect on word recognition performance should be assesses for children in different Arabic speaking countries.

It was hypothesized that participants in this study would perform better when listening to words spoken in Jordanian versus Saudi dialect and, in addition, that the effect would be similar for both normal hearing and hard of hearing children. Dialect appeared to exert an effect on word recognition ability of normal hearing Jordanian Arabic speaking children. Results reported herein confirmed that word recognition scores were higher when normal hearing children listened to words spoken in Saudi dialect. In contrast, no significant effect for Jordanian versus Saudi dialect was found in word recognition ability of children with mild to moderately severe hearing loss . In general, word recognition performance was lower and more variable among hard of hearing children than for normal hearing children, independent of dialect. However, the higher word recognition scores for the Saudi dialect were persistent in both hard of hearing and normal hearing groups. Perhaps because each list was limited to 33 words that were highly familiar to the target population (i.e., Jordanian Arabic speaking children age 6 to 9 years). In addition, the words were spoken similarly in standard Arabic and in both Jordanian and Saudi dialect, with the speaker remaining as the only potential factor affecting performance. Another consideration affecting data reported in the present study is the small sample size. Only 10 participants were included in each group, with three speakers in each dialect. Although the power analysis requirement was met by statistical criteria (power .95), a larger sample size would probably yield more representative results.

The results from the present study are not consistent with Alusi's (1974) contention that his word lists could be used in all Arab countries since the word lists were derived from standard Arabic. There are several clear differences between the present study and the study reported by Alusi. For example, the subjects in Alusi's study were educated adults, whereas subjects in the present study were children. Also, the sample size (N = 17) in Alusi's study was inadequate for comparison of performance from different dialectal backgrounds. Furthermore, the speaker in Alusi's study used Baghdad general dialect, whereas in the present study the speaker used general Jordanian and Saudi dialects. These differences among studies highlight the need for further investigation to compare word recognition performance with the speech materials from the present study with children from other Arabic speaking countries, particularly in relation to psychometric functions and the familiarity of words.

Further clinical investigations should be conducted utilizing the word lists developed in this study. A first step would be to collect data for standardization of the word lists. Several variables must be considered in the collection of standardization data. Data should be collected from a large sample of the target population that is representative of rural as well as urban Jordan. In addition, the sample should include participants of all socioeconomic groups, as well as children who display normal and impaired hearing abilities. Another step is to establish test reliability. Factors to be considered are lists versus half list reliability, children who display normal and impaired hearing, as well as gender (males and females).

Although the present study focused on word recognition, there is a need to develop other speech diagnostic materials that for evaluation of speech perception abilities of younger and older pediatric populations, as well as populations with different auditory problems, such as auditory processing disorder. Diagnostic speech audiometry materials that are needed clinically include:

- A “word recognition by picture identification test” for children younger than 6 years. The words used in the present study could be assessed for familiarity with younger children and on the basis of the availability of picture representations of the words. A picture-pointing word recognition measure would be very helpful in diagnosing and monitoring benefit of hearing aids and intervention in younger children.
- Material for central auditory processing diagnosis, such as filtered words, figure ground tests, and dichotic tests could be developed based on the present word lists. The fact that the words are digitally recorded facilitates the process of developing additional tests.
- Speech reception tests for estimation of threshold are lacking for Arabic-speaking children and adults. The procedures used to develop the speech materials in the present study have shown to be effective as in other studies by Harris et al. (2003b) and Nissen et al. (2005b), and can be duplicated.

Results from the present study will play a significant role in the development of additional Arabic speech audiometry materials and in improving audiological services provided in Jordan. The technique for recording the speech materials and the data collection conditions in the present study render these word lists to be of high quality. The speaker was carefully chosen to have clear speech and representative of general Jordanian dialect. The words were selected based on familiarity to children who are native speakers of Arabic in Jordan. The data were collected using headphones, the most common method for presenting speech signals in this age group. And, finally, the words were digitally recorded under excellent recording sound conditions providing a very clear signal, and materials that can be easily duplicated for the production of multiple audio copies of the word lists and half lists The word recognition materials developed in the present study are likely to be of considerable value in providing diagnostic services and intervention to the pediatric hard of hearing population in Jordan. In addition, these word recognition materials will contribute to the rapid development of diagnostic speech audiometry materials that are currently lacking in Arabic countries.

Table 5-1 Descriptive statistics for Hard of Hearing and Normal Hearing based on dialect

	Hard of Hearing		Normal Hearing	
	Jordanian	Saudi	Jordanian	Saudi
Mean	0.69	0.72	0.74	0.81
Standard Deviation	0.18	0.18	0.13	0.10
Range	0.66	0.76	0.55	0.46
Min	0.30	0.24	0.42	0.52
Max	0.97	1.00	0.97	0.97

APPENDIX A
PARTICIPANT RECRUITMENT

The initial pool of potential participants with normal hearing included 18 children; five of them were excluded after the initial pure tone audiometric testing due to their elevated thresholds that have exceeded the inclusion criteria of 15dB HL at least at one frequency. Further testing using bone conduction audiometry revealed mild conductive hearing loss. The children's caregivers were counseled and retest was recommended. In one case mild sensorineural hearing loss was diagnosed. This participant was included in the hard of hearing group. Data from 2 participants was excluded due to technical problems during data collection, which rendered results unreliable. Among the 20 potential participants in the hard of hearing group seven participants were excluded for having a severe hearing loss, based on the screening, and some had developmental delays that they were not able to understand the task.

APPENDIX B
SPEECH MATERIAL SELECTION

The words in this study were chosen from three different sources: a word frequency lists provided by Landau (1959) which was derived from daily newspapers from Lebanon, Egypt, Iraq, and Palestine, and modern prose; lists of Arabic key words for learners of everyday Arabic as a second language (Quitregard, 1994); and from children's story books.

To achieve Egen's (1948) criteria for word selection, the researcher selected words that fit one of the following possible monosyllabic word structures in Arabic: CV, CVC, CVCC, CVVC, and CVV. According to Altoma (1969) these monosyllabic word structures are shared by colloquial and standard Arabic. As mentioned above, words were selected from a range of written material. To ensure that selected words are representative of spoken Arabic, the procedure employed by Alusi (1974) and Ashoor and Prochazka (1982 and 1985) were followed. Words that have similar forms in colloquial and standard Arabic were selected. This step can be supported by Altoma's (1969) finding that 83.5% of the words in a colloquial form have shared origins with words in standard Arabic. Words from Landau (1959) and Quitregard (1994) and a list of (135 words) that are familiar to first, second and third grade Jordanian Arabic speaking students (Abdulhaq, unpublished) were used. Abdulhaq's familiar word list is composed of monosyllabic words chosen from children's story books and Jordanian daily newspaper articles, and rated by 320 first, second, and third grade students as familiar or unfamiliar. The words were read aloud to the children, they were asked to raise their hand if they knew the meaning of the word and were able to put it in a meaningful sentence. To ensure the accuracy of their response, Abdulhaq started with three trial words that were familiar to children (window, door, and table), randomly asked children to put words in a sentence, in addition unfamiliar words (from old classical text) and nonsense words were included. The initial list included 210 words. One hundred and thirty five words were rated as familiar at least by 80% of the children. Additional words from Landau and Quitregard then were added. The researcher used three different procedures to ensure the familiarity of words chosen for the word lists. Monosyllabic words were matched between the word lists from Landau (1959) and Quitregard (1994). Words that occurred in both lists were considered for further familiarity testing. Second the matching words and Abdulhaq's list were rated as familiar or unfamiliar by four Jordanian teachers of first, second and third grade. Finally teachers' ratings were compared to the children's ratings for consistency. Out of these lists, the top 304 words rated as highly familiar by students and teachers were selected for recording.

APPENDIX C
LOGISTIC REGRESSION ANALYSIS

Logistic regression analysis calculates the probability of a response based on the independent variable, in this study, the stimulus intensity level. The logistic regression analysis provides derived variables that are key parameters for the regression line that may be tested and compared between conditions (in this study, the word lists). The results from the logistic regression analysis, *a*, the logistic regression intercept and *b*, the logistic regression slope, were used to calculate the probability of correct identification of the words at each intensity level. Based on the model described by Nissen et al. (2005a) in the description of performance on the word list in terms of logistic regression, the following modified equation was used to calculate the percent correct:

Equation 1 $$P = \left(1 - \frac{\exp(a + b \times i)}{1 + \exp(a + b \times i)}\right) \times 100$$

Where P is the percent correct at an intensity level, a is the regression intercept, b is the regression slope, and i is the intensity level in dB HL. Using the regression intercept, slope and intensity levels in equation 1 made it possible to predict the percent correct at any given intensity level.

APPENDIX D
LIST OF WORDS FOR DIALECT COMPARISON

	Word	Meaning		Word	Meaning
1	أخ	Brother	18	صَف	Class
2	باب	Door	19	طير	Bird
3	بيض	Egg	20	عُش	Nest
4	ثوب	Dress	21	عود	Stick
5	ثور	Bull	22	عَين	Eye
6	جَد	Grandparent	23	عَم	Uncle (paternal)
7	حَجّ	Pilgirim	24	فم	Mouth
8	خيط	Thread	25	فيل	Elephant
9	دب	Bare	26	كوخ	Cottage
10	دَم	Blood	27	لِص	Thief
11	ديك	Rooster	28	لَوح	Board
12	رُز	Rice	29	لوز	Almonds
13	ريش	Feather	30	موز	Banana
14	سوق	Market	31	نار	Fier
15	سيف	Sword	32	نوم	Sleep
16	شاي	Tea	33	يَد	Hand
17	شوك	Thornes			

APPENDIX E
HALF WORD LISTS AND MEANING

1A		2A		3A		4A	
خَير	Good	رَب	God	مَرج	Meadow	شَمس	Sun
كوع	Elbow	حاج	Pilgrim	خَيل	Horses	تاج	Crown
بيت	Home	زر	Button	جن	Fairies	تين	Figs
رأس	Head	لف	Wrap	حال	Situation	بئر	Well
سوق	Market	نار	Fire	رَد	Answer	طوب	Brick
لوز	Almonds	شَد	Tight	طول	Length	تحت	Under
خال	Uncle (maternal)	غاز	Gas	بَنك	Bank	عَكس	Opposite
كأس	Cup	طير	Bird	حوت	Whale	سام	Poisonous
بَرد	Cold	كُم	Sleeve	ريح	Wind	جيب	Pocket
أهل	Family	حُر	Free	صوص	Chick	كوخ	Cottage
باص	Bus	سَد	Dam	عَين	Eye	حَب	Seeds
صَيد	Hunt	خوف	Fear	جاف	Dry	كَلب	Dog
مَوت	Death	عود	Stick	سَطح	Roof	سيخ	Skewer
زَوج	Husband	عِز	Prosperity	روح	Spirit	عَد	Count
سود	Blacks	صب	Poured (adj)	رَف	Shelf	حَوض	Tub
غَرب	West	مَجد	Glamour	شَخص	Person	زير	Jug
ثوب	Dress	شاي	Tea	بُرج	Tower	رَمز	Symbol
واد	Valley	شَيخ	Shaikh	بَر	Wilderness	صعب	Difficult
وز	Geese	صين	China	بنت	Girl	بُعد	Distance
ثَلج	Snow	دار	House	طَقس	weather	درس	Lesson
فيل	Elephant	دود	Turn	أرض	Earth	مُر	Bitter
شيء	Something	شَعر	Hair	وَرد	Roses	ضَيف	Guest
زيد	Zaid (name)	قلب	Heart	أنف	Nose	ألف	Thousand
شَط	Beach	ريش	Feather	حَق	Right	سَبت	Saturday
صوت	Voice	دين	Religion	حُب	Love	بطّ	Ducks

1B		2B		3B		4B	
خس	Lettuce	فَرض	Requirement	بيض	Egg	جِد	Grandfather
جَد	Serious	ثوم	Garlic	ثور	Bull	خيط	String
كَف	Palm	بَيع	Sell	موز	Banana	ظِل	Shadow
نِصف	Half	خَلف	Behind	تخت	Bed	حَل	Solution
لمس	Touch	جَيش	Armey	عيد	Holiday	إسم	Name
نَفس	Same	عام	Year	حَرب	War	عَرش	Throne
أم	Mother	زَي	Outfit	قلب	Turn over	سيف	Sword
روس	Heads	حَجّ	Pilgrimage	زاد	Increase	فول	Lima bean
بوق	Trumpet	سَير	Traffic	شَر	Evil	جار	Neighbor
لوح	Board	سِر	Secret	لِص	Thief	فأس	Axe
عَرض	Width	جاد	Serious	أب	Father	شاش	Gauze
شوك	Thorns	يوم	Day	حار	Spicy	دَرب	Way
حَر	Hot	رُز	Rice	غَوص	Dive	كرش	Belly
توت	Berry	جَو	Space	خط	Line	سور	Fence
شَك	Check	جوع	Hunger	نور	Light	ثُلث	Third
زيت	Oil	ريف	Rural	قوس	Arch	فار	Mouse
ديك	Rooster	مال	Money	ضوء	Lamp	دَم	Blood
دور	Turn	شب	Youngman	ناس	People	أخت	Sister
حِزب	Party	أخ	Brother	دب	Bear	حَد	Limit
بوت	Boot	ريم	Reem (name)	لُب	Core	نوم	Sleep
أنس	friendliness	خاص	Private	خوخ	Plum	حين	When
حَي	Neighborhood	صَف	Class	هَم	Worry	لون	Color
قِرد	Monkey	عُش	Nest	شَرق	East	فَرد	Gun
كنز	Treasure	صوف	Wool	دُش	Shower	كيس	Bag
شيب	Gray hair	مَد	Tide	عَم	Uncle (paternal)	جِلد	Skin

LIST OF REFERENCES

Abd-el-Jawad, H. (1987). Cross-dialectal variation in Arabic: competing prestigious forms. *Language in Society, 16;* 359-368.

Al-Kahtani, A. (1997). The "Problem" of Diglossia in the Arab world: an attitudinal study of modern standard Arabic and the Arabic dialects. *Al-Arabiyya, 30;*1-30.

Al- Masri, M. (2003). *Early Detection and Habilitation of Hearing Loss, MEHA Project 4*. Amman, Jordan.

Altoma, S. (1969). The Problem of Diglossia in Arabic, a comparative study of classical and Iraqi Arabic. *Harvard Middle Eastern Monographs XXI*.

Alusi, H.A., Hinchcliffe, R., Ingham, B., Knight, J.J., & North C. (1974). Arabic speech audiometry. *Audiology, 13*; 212-130.

American National Standards Institute (2004). *Specification for Audiometers.* ANSI S3.6-2004. New York: ANSI.

Ashoor, A.A. & Prochazka, T. Jr. (1982). Saudi Arabic speech audiometry. *Audiology, 21*; 493-508.

Ashoor, A.A. & Prochazka, T. Jr. (1985). Saudi Arabic speech audiometry for children. *British Journal of Audiology, 19*; 229-238.

Arlinger, S. (2001). How to assess outcome of hearing aid fitting in children. *Scandinavian Audiology, 30 Suppl. 53;* 68-72.

Beattie, R.C., Edgerton, B.J., & Svihovec, D.V. (1977). A comparison of the Auditec of St. Louis Cassette recordings of NU-6 and CID W-22 on a normal-hearing population. *Journal of Speech and Hearing Disorders, 42(1).* 60-64.

Beattie, R.C., Svihovec, D.V., & Edgerton, B.J. (1975). Relative intelligibility of the CID spondees as presented via monitored live voice. *Journal of Speech and Hearing Disorders, 40(1)*; 84-91.

Beatti, R.C. & Warren, V. (1983). Slope characteristics of CID W-22 word functions in elderly hearing –impaired listeners. *Journal of Speech and Hearing Disorder, 4;*119-127.

Blamey, P., Sarant, J., Paatsch, L., Barry, J., Bow, C., Wales, R., Wright, M., Psarros, C., Rattigan, K. & Tooher, R. (2001). Relationship among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech, Language, and Hearing Research, 44(2);* 264-283.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*; 62-71

Carhart, R. (1965). Problems in the measurement of speech discrimination. *Archives of Otolaryngology, 82;* 253-260.

Cramer, K. & Erber, N.P. (1974). A spondee recognition test for young hearing-impaired children. *Journal of speech and hearing Disorders, 39;* 304-311.

Curry, E.T. (1949). A study of the relationship between speech thresholds and audiometric resluts in perception deafness. *Journal of Speech and Hearing Disorders, 14;* 104-110.

Davis, H. (1948). The Articulation area and the social adequacy index for hearing. *Laryngoscope, 58*; 761-778.

Doyne, M. P., & Steer M.D. (1951). Studies in speech reception testing. *Journal of Speech and Hearing Disorders, 16;* 132-139.

Dubno, J.R., Lee, F.S., Klein, A.J. & Matthews, L.J. (1995). Confidence limits for maximum word-recognition score. *Journal of Speech and Hearing Research, 38;* 490-502.

Egan, James (1948). Articulation testing methods. *Laryngoscope, 58*; 955-991.

Elliot, L. (1979). Performance of children age 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 66;* 651 -653.

Elliot, L, Connors S., Kille, E., Levin, S., Ball, K. & Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *Journal of the Acoustical Society of America, 66;* 12-21.

Elliot, L. & Katz, D.R. (1980). *Northwestern University Children's Perception of Speech (NU-CHIPS).* AUDITEC of St. Louis, St. Louis, Missouri.

Erber, N.P. (1974). Pure-tone thresholds and word-recognition abilities of hearing-impaired children. *Journal of Speech and Hearing Research, 17;* 194-202

Erber, N.P. (1980). Use of the auditory numbers test to evaluate speech perception abilities of hearing-impaired children. *Journal of Speech and Hearing Disorders, 45;* 527-532.

Erber, N.P. & Alencewicz,C. (1976). Audiologic evaluation of deaf children. *Journal of Speech and Hearing Disorders, 41;* 256-267.

Fatihi, A.R. (2001). *Aspects of Arabic Phonology*. Kalinga Publication, Delhi.

Ferguson, C. (1959). Diglossia. *Word, 15;* 325-340

Fletcher, H. (1958). *Speech and Hearing in Communication.* D. van Nostrand Company, Inc., Princeton, New Jersy.

Fletcher, H. (1995). The ASA edition of speech and hearing in communication. Ed. Allen, J.B. Acoustical Society of America, Woodbury, NY.

Fletcher, H. & Steinberg, J.C. (1929). Articulation testing methods. *Bell System Technical Journal, 8;* 806-854.

Goldman,R., Fristoe, M. & Woodcock, R. (1970). *Goldman-Fristoe-Woodcock Test of Auditory Discrimination.* American Guidance Service, Inc. USA.

Harris, R., Kim, E. & Eggett, D. (2003a). Psychometrically equivalent Korean monosyllabic speech discrimination materials spoken by male and female talkers. *Korean Journal of Communication Disorders, 8 (1);* 217-243.

Harris, R., Kim, E. & Eggett, D. (2003b). Psychometrically equivalent Korean bisyllabic words spoken by male and female talkers. *Korean Journal of Communication Disorders, 8 (1);* 244-270.

Haskins, H.A. (1949). *A phonetically Balanced Test of Speech Discrimination for Children.* Master's thesis. Northwestern University, Evanston , IL.

Hirsh, I.J., Davis, H., Silverman, R., Reynolds, E.G., Eldert,E., & Benson, R. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorder, 17;*321-337.

Hirsh, I. J., Reynolds, E.G., Joseph, M. (1954). Intelligibility of different speech materials. *The Journal of the Acoustical Society of America, 26(4);* 530-538.

Hood, J.D. & Poole, J.P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *Audiology, 19;* 434-455.

Ibrahim, Z. (2000). Myths about Arabic revisited. *Al-Arabiyya, 33;* 13-28.

The International Kindergarten Union, Child Study Committee (1928). *A Study of Vocabulary of Children Before Entering the First Grade.* Washington, D.C.

Jauhiainen, T. (2001). Hearing impaired children in developing countries. Paper presented at the 3rd International Congress of Pediatric Audiology, 9-12 May 2000, Copenhagen. *Scandinavian Audiology, 30 Suppl 53;* 83-84.

Jerger, S., Lewis, S., Hawkins, J., & Jerger, J. (1980). Pediatric speech intelligibility test. I. Generation of test materials. *International Journal of Pediatric Otorhynolaryngology, 2;* 217-230.

Kishon-Robin, L. & Rosenhouse, J. (2000). Speech perceptin test for Arabic-speaking children. *Audiology, 39;* 269-277.

Kreul, E.J., Bell, D.W. & Nixon, J.C. (1969). Factors affecting speech discrimination test difficulty. *Journal of Speech and Hearing Research, 12;* 281-287.

Landau, J. (1959). *A Word Count of Modern Arabic Prose.* New York: American Council of Learned Societies in cooperation with the school of education at the Hebrew University, Jerusalem.

Larson, G.W., Petersen, B., & Jacquot, W.S. (1974). Use of Northwestern University Test No.6 for speech discrimination testing with children. *The Journal of Auditory Research, 14;* 287-292.

Laukli, E. & Fjermedal, O. (1990). Reproducibility of Hearing Threshold Measurements, supplementary Data on Bone-conduction and speech audiometry. *Scandinavian Audiology, 19*; 187-190.

Martin, F.N., Champlin, C.A., & Perez, D.D. (2000). The question of phonetic balance in word recognition testing. *Journal of the American Academy of Audiology, 11;* 489-493.

Messouak, K (1956). Audiometrie vocale en Arabic maghrebin. *Les cahiers de la CFA, No. 4.* Compagnie Francaise d'Audiologie, Paris.

Meyer, T.A & Pisoni, D.B (1999). Some computational analyses of the PBK test: effect of frequency and lexical density on spoken word recognition. *Ear and Hearing, 20 (4);* 363-371.

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

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

Öster, A.M. (2002). The relationship between residual hearing and speech intelligibility- Is there a measure that could predict a prelingually profoundly deaf child's ability to develop intelligible speech? *Speech, Music and Hearing, Quarterly Progress and Status Report (TMH-QPSR), 43*; 5156.

Owens, E. (1961). Intelligibility of words varying in familiarity. *Journal of Speech and Hearing Research, 4(2);* 113-129.

Palmer, J. M. (1955). The effect of speaker differences on the intelligibility of phonetically balanced word lists. *Journal of Speech and Hearing Disorders, 20*; 192-195.

Quitregard, D. (1994). *Arabic Key Words, the basic 2000 word vocabulary arranged by frequency in a hundred units.* The Oleander Press, New York.

Ross, M. & Lerman, J. (1971). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13;* 44-53.

Sanderson-Leepa, M. & Rintelmann, W.F. (1976). Articulation functions and test-retest performance of normal-hearing children on three speech discrimination tests: WIPI, PBK-50, and NU Auditory test No. 6. *Journal of Speech and Hearing Disorders, 41(4);* 503-519.

Saudi Speech Pathology and Audiology Association, SSPAA (2004). http://www.sspaa.org , last accessed October 2004.

Schwartz, A & Goldman, R. (1974). Variables in fluencing performance on speech-sound discrimination tests. *Journal of Speech and Hearing Research, 17*; 25-32.

Siegenthaler, B.M, Pearson, J. & Lezak, R.J. (1954). A speech reception threshold test for children. *Journal of Speech and Hearing Disorder, 19;* 360-366.

Smith, A.W. (2001). WHO activities for prevention of deafness and hearing impairment in children. Paper presented at the 3rd International Congress of Pediatric Audiology, 9-12 May 2000, Copenhagen. *Scandinavian Audiology, 30 Suppl 53;* 93-100.

Soliman S (1976). Speech discrimination audiometry using Arabic phonetically balanced words. *Ain Shams Medical Journal, 27*; 27-30.

Soliman, S., Abd El-Hady, M., Saad, M., & Kolkaila, E. (1987a). Development and study of Arabic word intelligibility by picture identification (WIPI) in normal hearing Egyptian Children. *Scientific Journal of Al-Azhar University, Medical Faculty (girls), 6.*

Soliman, S., Fathalaa, A., & El-Mahalawi, T. (1987b). Simple speech test as a predictor of speech reception threshold (SRT) in preschool children. *Proceedings of Annals of the Xth Annual Medical Congress of Aid Shams University*, Cairo, Egypt.

Tillman, T.W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words (Northwestern University Auditory Test No.6). *Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine*, Brooks Air Force Base, Texas.

Thornton, A. & Raffin, M. (1978). Speech discrimination scores modeled as a binominal variable. *Journal of Speech and Hearing Research 21*; 507-518.

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

Wilson, R.H. & Oyler, A.L. (1997). Psychometric function for the CID-W22 and NU Auditory Test No.6. materials spoken by the same speaker. *Ear and Hearing, 18(5);* 430-433.

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

Wilson, R.H., Zizz, C., Shanks, J., & Causey, G. (1990). Normative data in quiet, narrowband noise, and competing message for Northwestern university auditory test No.6 by a female speaker. *Journal of Speech and Hearing Disorders, 55*; 771-778.

World Health Organization (1998). *Report of the International Workshop on Primary Care and Hearing Care.* Cape Town; South Africa March 12-14.

Young, J.R., Dudley, B., & Gunter, M.B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech and Hearing Research, 25;* 586-593.

Zakrzewski, A, Jassem, W., Pruszewicz, A., & Obrebowski, A. (1976). Speech audiometry for children and subjective probability of Polish words. *Audiology, 15;* 228-231.

BIOGRAPHICAL SKETCH

Nadia Abdulhaq was born and raised in Nablus, Palestine. The third of six children, she graduated from high school in 1990 and enrolled at Birzeit University in Birzeit, Palestine. She obtained her Bachelor of Arts degree in English language and literature with a minor in translation, in June, 1995. She then went on to receive her Master of Arts degree in speech-language pathology from the University of Jordan in Amman, Jordan, in June, 1997.

After graduating from the University of Jordan, she accepted the position of speech pathologist in the Department of Rehabilitation at the Palestine Red Crescent Society. Her job description included establishing three speech therapy units in three different cities across the West Bank of Palestine and training six rehabilitation workers as speech therapy assistants. Her work focused on children with hearing loss. After two years of work and achievements, she decided to return to the academic world and pursue further knowledge to be able to better serve the hard of hearing and deaf community of Palestine.

In 2000 she received the Fulbright scholarship to start her doctoral studies at the University of Florida in Gainesville, Florida. During her five year study she was an active member of the international student body initiating projects such as the International Student Speakers Bureau and the international student welcome information table. She received numerous scholarships and recognitions, O. Ruth McQown Scholarship, Grinter Fellowship, Gibbson Dissertation Fellowship, and the Outstanding Academic

Achievement award. She will graduate in December 2005 with a Ph.D. in communication sciences and disorders and Doctor of Audiology degrees. Upon her graduation, she will return to Palestine to start a speech and hearing program for the hard of hearing, new born hearing screening program, and an audiology clinic.