VALIDITY OF AUTOMATED THRESHOLD AUDIOMETRY: A SYSTEMATIC REVIEW AND META-ANALYSIS

by

Faheema Mahomed
(27038158)

A dissertation submitted in fulfilment of the requirements for the degree M.Communication Pathology in the Department of Communication Pathology at the University of Pretoria, Faculty of Humanities.

SUPERVISOR: Prof. De Wet Swanepoel
CO-SUPERVISOR: Dr. Maggi Soer

January 2013

© University of Pretoria

ACKNOWLEDGEMENTS

- A special thank you to Professor De Wet Swanepoel for your constant guidance and encouragement throughout this process. Regardless of your extremely busy schedule you always made the time to respond to my long emails and many questions. Not only was I assisted with my research, but you exposed me to writing for journals and presenting at conferences. You made research an enjoyable process and I look forward to working with you in the future.

- Professor Robert Eikelboom, if it wasn't for Prof I'd still be trying to obtain the articles I needed. Thank-you for your continued assistance, guidance and support.

- Doctor Maggi Soer, your attention to detail assisted in ensuring that I referenced correctly and always made correct use of tenses, thank-you.

- To my family, I thank-you for your support and understanding during this process.

© University of Pretoria

Table of contents

© University of Pretoria

© University of Pretoria

© University of Pretoria

FIGURES

TABLES

© University of Pretoria

LIST OF ABBREVIATIONS

AC: Air Conduction

BC: Bone Conduction

dB: Decibel

Hz: Hertz

SD: Standard Deviation

FORMATING

APA referencing style was utilized in this dissertation.

© University of Pretoria

ABSTRACT

The need for hearing health care services across the world far outweighs the capacity to deliver these services with the present shortage of hearing health care personnel. Automated test procedures coupled with telemedicine may assist in extending services. Automated threshold audiometry has existed for many decades; however, there has been a lack of systematic evidence supporting its clinical use. The aim of this study was to systematically review the current body of peer-reviewed publications on the validity (test-retest reliability and accuracy) of automated threshold audiometry. A meta-analysis was thereafter conducted to combine and quantify the results of individual reports so that an overall assessment of validity based on existing evidence could be made for automated threshold audiometry.

A systematic literature review and meta-analysis was conducted using peer-reviewed publications. A multifaceted approach, covering several databases and employing different search strategies, was utilized to ensure comprehensive coverage and crosschecking of search findings. Publications were obtained using the following three databases: Medline, SCOPUS and PubMed, and by inspecting the reference list of relevant reports. Reports were selected based according to inclusion and an exclusion criterion, thereafter data extraction was conducted. Subsequently, the meta-analysis combined and quantified data to determine the validity of automated threshold audiometry.

In total, 29 articles met the inclusion criteria. The outcomes from these studies indicated that two types of automated threshold testing procedures have been utilized, the 'method of limits' and 'method of adjustments'. Reported findings suggest accurate and reliable thresholds when utilizing automated audiometry. Most of the reports included data on adult populations using air conduction testing, limited data on children, bone conduction testing and the effects of hearing status on automated threshold testing were however reported. The meta-analysis revealed that test-retest reliability for automated threshold audiometry was within typical test-retest reliability for manual audiometry. Furthermore, the meta-analysis showed comparable overall average differences between manual and automated air

© University of Pretoria

conduction audiometry (0.4 dB, 6.1 SD) compared to test-retest differences for manual (1.3 dB, 6.1 SD) and automated (0.3 dB, 6.9 SD) air conduction audiometry. Overall, no significant differences (p>0.01; Summarized Data ANOVA) were obtained in any of the comparisons between test-retest reliability (manual and automated) and accuracy.

Current evidence demonstrates that automated threshold audiometry can produce an accurate measure of hearing threshold. The differences between automated and manual audiometry fall within typical test-retest and inter-tester variability. Despite its long history however, validation is still limited for (i) automated bone conduction audiometry; (ii) automated audiometry in children and difficult-to-test populations and; (iii) automated audiometry with different types and degrees of hearing loss.

Keywords: automated threshold audiometry, method of limits, method of adjustments, air conduction, bone conduction, validation, test-retest reliability, accuracy, meta-analysis, literature review.

© University of Pretoria

1. INTRODUCTION

"Automation of healthcare services is becoming increasingly important in light of the global shortage of specialized healthcare personnel."
(Swanepoel, Mngemane, Molemong, Mkwanazi & Tutshini, 2010).

Automated or asynchronous healthcare services refer to a procedure or intervention conducted without requiring the necessary healthcare professional to be present. In situations where specialist healthcare personnel are limited or unavailable, this approach may ensure that services and healthcare resources are optimized (Swanepoel et al., 2010).

With the continually transforming nature of the audiological field, tests that were once widely used have been replaced by subsequently discovered objective automated measures of auditory functioning and integrity with greater diagnostic accuracy. Among these procedures are the Auditory Brainstem Response (ABR), Otoacoustic Emissions (OAE) and Magnetic Resonance Imaging (MRI). While some tests have evolved over time, a few fundamental aspects of audiological testing has not changed significantly. The common standard of determining hearing sensitivity continues to consist of utilizing manual pure tone audiometry testing.

1.1. Background

Automated threshold audiometry has existed for many years, however, it has not been widely used in clinical practice (Margolis & Morgan, 2008). The earliest record of automated threshold audiometry in clinical use was reported in the seminal report of Georg von Békésy (Békésy, 1947). This type of automated threshold audiometry was described by Margolis and Morgan (2008) as a computer-based program that adjusts the recording parameters automatically according to the responses of the patient during the test. This self-recording threshold audiometer automatically increases and decreases the sound intensity while sweeping through the test frequency range. This method of testing is known as 'sweep frequency Békésy audiometry'. The patient is required to press a response button when the test signal is heard and release it when they lose perception of the signal. This method of

© University of Pretoria

setting the threshold can be seen as a form of 'method of adjustment'. Subsequent systems utilized derivations of this technique with fixed frequency threshold seeking algorithms, referred to as 'fixed or discreet frequency Békésy audiometry'. A sweep in intensity occurs within a fixed frequency based on the patient's behavioural response relayed through a response switch (Frank & Ragland, 1987; Meyer-Bisch, 1996). This method produces a response pattern which not only provides threshold information but also indicates the 'site of lesion', thus conductive, cochlear and retrocochlear pathologies can be distinguished based on the response pattern.

In later years automated audiometry systems have been programmed according to the 'method of limits', the method upon which conventional manual audiometry procedural steps (Sparks, 1972) are based, typically utilizing versions of the Hughson and Westlake threshold-seeking method and adaptive psychophysical procedures (Hughson & Westlake, 1944, Fagan, 2009). The audiometer automatically makes adjustments to the intensity of the presented signal, up or downwards depending on the response or lack of response. This method has been modified in some cases to include forced-choice responses from the patient. Here the listener is forced to listen and make a response that indicates that a sound was either heard or not heard by touching a 'button' for example on a touch-screen monitor (Frank & Ragland, 1987; Margolis & Morgan, 2010).

Automated audiometry is a viable alternative, as the measures of manual threshold audiometry are especially suited to automation because they are based on predetermined sequenced steps (Margolis & Morgan, 2008). These steps are based on reliable algorithms and can be transferred to a computer capable of reproducing them. In addition, when using automated procedures, results are recorded electronically thus enabling all the advantages of electronic record keeping, such as reduced paperwork, transfer to other clinicians and tracking change in hearing status over time. Additionally, automated audiometry can incorporate quality-monitoring mechanisms to ensure consistent and reliable results (Margolis & Moore, 2011; Margolis, Saly, Le & Laurence, 2007). Automation may also potentially improve standardization of tests protocols and procedures across clinics and even within clinics.

© University of Pretoria

Although automated audiometry has existed for many decades, it has been used almost exclusively in industry as part of mass hearing screening and baseline monitoring or for research purposes. Clinical audiological practices, in contrast, have almost exclusively relied on conventional manual audiometry. The preference for manual audiometry may be attributed to concerns regarding the accuracy when compared to the customary manual air and bone conduction audiometry. According to Margolis and Morgan (2008) inaccurate audiograms will create the need for repeat testing, which contradicts the rationale of utilizing automated audiometry for its cost and time-efficiency. In addition, it would undermine the confidence that colleagues and other health professionals have in audiometric results. However, manual audiometry, as in automated audiometry, also presents with variability due to subject factors such as fatigue and concentration as well as due to different transducers utilized (ANSI, 1996; Margolis et al., 2007). Normal variability has typically been quantified by test-retest reliability and occasionally by inter-tester reliability (Margolis et al., 2007; Swanpoel & Biago, 2011) of manual audiometry. Normal variability of automated audiometry can be looked at in the same manner.

1.2. Rationale

The World Health Organization (WHO, 2010) has indicated that 278 million people worldwide experienced bilateral moderate-to-profound hearing impairments in 2005 (WHO, 2010). An estimated 80 percent of these individuals with hearing impairment live in developing countries (WHO, 2010) without access to hearing health care services predisposing them to poverty and limited educational and vocational opportunities (Swanepoel, Olusanya & Mars, 2010). Hearing loss may also have disabling socio-emotional effects such as depression, loneliness, anxiety and work-related fatigue (Nachtegaal, Smit, Smits, Bezemer & Van Beek, 2009) which results in poor quality of life.

Hearing loss has a high prevalence and may have a significant consequence on quality of life as it is one of only four nonfatal conditions among the 20 leading causes of the global burden of disease. The need for hearing health care services across the world far outweighs the capacity to deliver these services with the present shortage of hearing health care personnel (Swanepoel et al., 2010). Goulios and

© University of Pretoria

Patuzzi (2008) stated that developing nations rely greatly on external technical and financial support. In addition, according to a survey conducted by Fagon and Jacobs (2009), these nations have a lack of training centres available, as well as a lack of resources and infrastructure. Automated test procedures coupled with telemedicine may address the lack of professionals in these regions to a large degree. Recent research has suggested that Telemedicine may provide a starting point in the provision of services to populations without access to certain medical services (Swanepoel et al., 2010). With regards to the global shortage of specialized hearing health care personnel, automation of audiological services combined with telemedicine is becoming an increasingly important area of development and validation.

1.3. Problem statement

In the light of the potential benefits of automation in threshold audiometry, its long history and the apparent lack of summative evidence supporting its utilization, the present study aimed to systematically review the current body of peer-reviewed publications on the validity (test-retest reliability and accuracy) of automated threshold audiometry. Additionally the study included a meta-analysis, using results from published reports, to quantify the test-retest reliability and accuracy of automated threshold audiometry.

© University of Pretoria

2. METHODOLOGY

2.1. Research aims

The research methodology describes the process that was followed in order to determine the validity of automated threshold audiometry. The aims of the research project were as follows:

Main aim

The aim of this study was to systematically review and quantify the validity, as measured by reliability (test-retest) and accuracy, of automated threshold audiometry compared to manual threshold audiometry in published literature.

Sub-aims

The following sub-aims were formulated in order to achieve the main aim:

- *Sub-aim 1:* to determine test-retest reliability of automated threshold audiometry compared to the gold standard of manual threshold audiometry for air and bone conduction testing.
- *Sub-aim 2:* to determine the accuracy of automated threshold audiometry compared to the gold standard of manual threshold audiometry for air and bone conduction testing.
- *Sub-aim 3:* to conduct a meta-analysis in order to combine and quantify results obtained for accuracy and test-retest reliability of automated threshold audiometry compared to manual threshold audiometry for air and bone conduction testing.

2.2. Research design

Over the past six decades there have been various studies conducted regarding the validity of automated threshold audiometry (Burns & Hinchcliffe, 1957; Fautsi, Frey, Henry, Knutsen & Olson, 1990; Fromby, Sherlock & Green, 1996; Gosztonyi, Vassailo & Sataloff, 1971; Ho, Hildreth & Lindsey, 2009; Ishak, Zhao, Stephans Culling, Bai, Meyer-Bisch, 2011; Swanepoel et al. 2011), however, no definitive recommendations regarding its use and application in the clinical diagnostic setting

© University of Pretoria

has emerged. Additionally, a systematic review had not been carried out on the evidence available regarding the validity of automated threshold audiometry. This study utilized a systematic literature review to determine the validity of automated threshold audiometry and identify any gaps regarding automated threshold audiometry in current research, in order to suggest areas for further investigation and provide a framework or background to appropriately position new research activities.

In addition, a meta-analysis was conducted allowing the researcher to make the best use of all the information gathered during the systematic review by increasing the value of the analyses. According to Cooper (2010) by statistically combining the results of similar studies it will improve the precision of our estimates, and assess whether effects are similar in similar situations. In this particular study the meta-analysis allowed the researcher to determine the validity (accuracy and test-retest reliability) of automated compared to manual threshold audiometry across currently published reports.

2.3. Ethical considerations

Ethical guidelines are set in research as a standard within the framework of good clinical practice. This study involved the review of published documents and did not involve human subjects. Thus ethical principles related to human research subjects were irrelevant. The following ethical aspects were taken into account during the research study:

2.3.1. Plagiarism

Plagiarism refers to the unauthorized use or close imitation of the language and thoughts of another author and the representation of them as one's own original work, as by not crediting the author (Leedy & Ormrod, 2001). To avoid any form of plagiarism all sources that were used to contribute to this study were acknowledged in the study.

2.3.2. Publication bias

Publication bias refers to the problem that positive results are more likely to be published than negative results. Publication bias can lead to systematic bias in

© University of Pretoria

systematic reviews unless special efforts are made to address this problem (Kitchenham, 2007). A multifaceted approach, covering several databases and employing different search strategies, was utilized to ensure comprehensive coverage and cross checking of search findings (White & Schmidt, 2005).

2.3.3. Reliability and validity of research

Bryman and Bell (2007) stated that validity is concerned with the trustworthiness of the conclusions that are generated from a piece of research whereas reliability is defined as the replication and consistency of measures. The integrity of the quantitative data collected was ensured by the following:

- The use of varied search strategies employed across several electronic databases to identify relevant research reports (excluding editorials, notes and short surveys) from peer-reviewed literature.
- Reviewing of reports to determine if the report met the inclusion criteria was conducted by the researcher and if any queries arose the supervisor reviewed the reports as well, ensuring reliability and validity of data obtained.

2.3.4. Compliance with the institutional review board protocol

The proposal was submitted to the Research Ethics Committee of the Faculty of Humanities of the University of Pretoria for approval. No data collection commenced prior to approval of the proposed study which was granted on the 26th of April 2012 (Appendix A).

2.4. Selection criteria for research material

Selection criteria are intended to identify those primary studies that provide direct evidence about the research question (Kitchenham, 2007). In order to reduce the likelihood of bias and to ensure reliability and validity of data, selection criteria needed to be specified. The following inclusion and exclusion criteria were utilized:

2.4.1. Exclusion criteria

- – reports published after 20 July 2012 were not included,
- – all duplicates and unrelated papers were excluded,

© University of Pretoria

- reports regarding other threshold seeking methods not utilising pure tones were not included,
- exclusion of reviews, editorials notes, letters and short surveys.

2.4.2. Inclusion criteria

- utilization of English reports only,
- peer reviewed reports of a comparative nature between automated and manual threshold audiometry were included,
- all reports published before 20 July 2012 were included.

2.5. Data collection procedures

The data collection procedures for the systematic literature review and meta-analysis are discussed independently.

2.5.1. Data collection procedures: systematic literature review

The aim of a systematic review is to find as many primary studies relating to the research question as possible using an unbiased search strategy. In order to achieve this, the research question was broken down into multiple facets i.e. validity, efficiency, automated, pure tone audiometry. A list of synonyms, abbreviations, and alternative spellings were determined for each facet. Once a comprehensive list was created, varied search strategies were employed, searching several electronic databases, to identify relevant research articles and conference proceedings (excluding reviews, editorials, notes, letters and short surveys) from the peer-reviewed literature. All relevant reports and conference proceedings that fit the inclusion criteria were used.

Medline database, Scopus and PubMed, using Medical Subject Heading (MeSH) were the three databases utilized during this study. The Medline database search utilized a strategy of relevant keywords to determine all records relating to the study aim. The second database, PubMed, was searched using the same search terms as for the Medline database, except that it was restricted to the available Medical Subject Heading (MeSH) terms. SCOPUS, the third database included in the search

© University of Pretoria

strategy, is the world's largest abstract and citation database of peer-reviewed literature also indexing Medline. This served as a cross-check for reports from PubMed and Medline databases.

The researcher reviewed the abstracts of all reports resulting from the searches to determine if the report complied with the inclusion criteria. If any queries arose the supervisor also reviewed the abstracts. Where an abstract was unavailable, the full paper was reviewed. After all duplicates and unrelated reports had been excluded, the remaining reports were reviewed in full to determine if they meet the inclusion criteria (Figure 1). A secondary search was employed to supplement the findings of the primary search. The secondary search involved reviewing the reference lists of all reports already identified for inclusion during the primary search strategy.

2.5.2. Data collection procedures: Meta-analysis

A meta-analysis was conducted to combine and quantify the results of individual reports so that an overall assessment of test-retest reliability and accuracy based on existing evidence could be made for automated audiometry. To be included in the meta-analysis, reports had to meet the following criteria: (i) the report had to include data comparing manual and automated audiometry in terms of accuracy; (ii) data had to be reported in the form of mean differences (real or absolute) and standard deviations with the number of observations reported.

© University of Pretoria

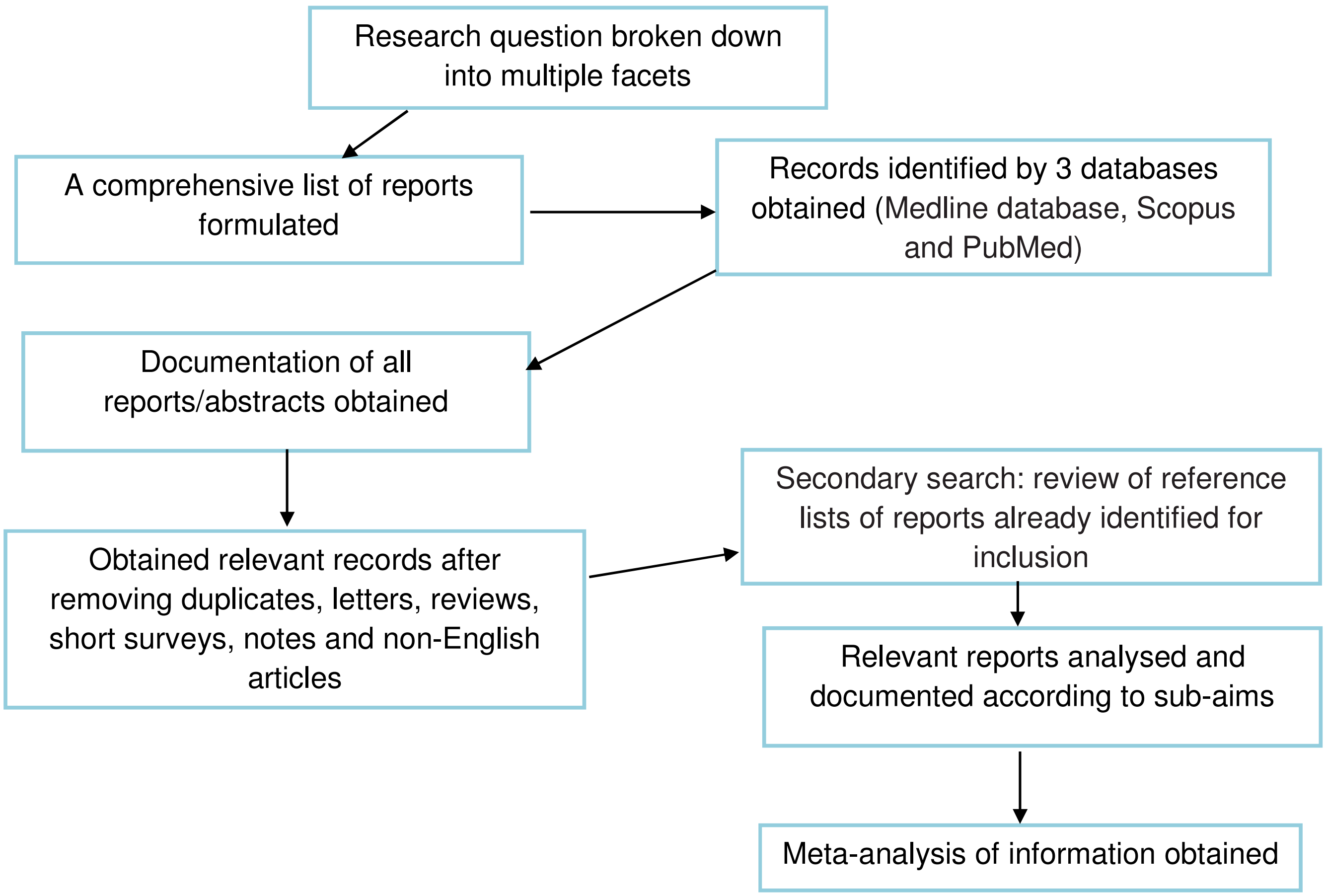

Research question broken down into multiple facets
A comprehensive list of reports formulated
Records identified by 3 databases obtained (Medline database, Scopus and PubMed)
Documentation of all reports/abstracts obtained
Obtained relevant records after removing duplicates, letters, reviews, short surveys, notes and non-English articles
Secondary search: review of reference lists of reports already identified for inclusion
Relevant reports analysed and documented according to sub-aims
Meta-analysis of information obtained

Figure 1. Data collection procedure

2.6. Data analysis procedures

2.6.1. Data analysis procedures: systematic literature review

Line of argument synthesis was utilized in analyzing the data obtained from the systematic review. This approach is used when researchers are concerned about what they can infer about a topic as a whole from a set of selective studies that look at a part of the issue. This analysis consists of two phases. First the individual studies are analyzed, then an attempt is made to analyze the set of studies as a whole (Kitchenham, 2007). Tables were structured in such a manner as to highlight similarities and differences between each studies outcome. The reports selected for review were carefully scrutinized and categorized according to the audiological threshold seeking method utilized (method of adjustment or method of limits), type of evaluation (diagnostic or screening), AC and/or BC thresholds, type of transducers and audiometer utilized, age, as well as hearing status of participants, type of statistical analysis for accuracy and test-retest reliability as well as the conclusions

© University of Pretoria

drawn by the paper (Appendix B). Quantitative information extracted from the studies regarding test-retest reliability and accuracy of each study was tabulated appropriately according to the research question (Appendix C and Appendix D)

2.6.2. Data analysis procedures: Meta-analysis

A meta-analysis was conducted, once all the data was systematically reviewed and tabulated to determine the accuracy and test-retest reliability of automated compared to manual threshold audiometry. In order to carry out the meta-analysis the average differences (real and absolute) and standard deviations were documented. An average difference (real) shows a systematic effect but negative and positive difference may cancel each other out even when large differences in either direction exist, whereas absolute average differences provides an indicator of the expected spread in variability. With this in mind, the meta-analysis made use of average differences (real and absolute) and standard deviations in order to draw conclusions regarding the validity of automated audiometry when compared to manual audiometry. Thereafter, weighted averages utilizing reported real and absolute average differences and standard deviations were determined for validation (test-retest reliability and accuracy) across studies, taking into account the number of observations reported. Bland and Altman (1986) recommend the use of absolute average differences and standard deviation as a more appropriate measure of correspondence since it provides an indicator of the expected spread in variability.

Furthermore a comparison of test-retest threshold differences for manual and automated threshold audiometry, indicative of variability, was made with the difference between automated and manual threshold audiometry (accuracy) using an analysis of variance test (http://statpages.org/anovalsm.html). A significant difference in variability was noted by a p-value level of <0.01.

© University of Pretoria

3. VALIDITY OF AUTOMATED THRESHOLD AUDIOMETRY: A SYSTEMATIC REVIEW AND META-ANALYSIS

Authors: Faheema Mahomed, De Wet Swanepoel, Robert H Eikelboom, Maggi Soer.

Journal: Ear and Hearing

Accepted: Accepted for publication pending minor revisions (13 March 2013)

Note: This manuscript was edited in accordance with editorial specifications of the journal and may differ from the editorial style of the rest of the dissertation. The term Supplemental Digital Content as utilized in the Ear and Hearing manuscript was replaced by appendices in the dissertation. Three of the Supplemental Digital Content items in the Ear and Hearing manuscript have been included as tables in this chapter of the dissertation.

3.1. Abstract

Objectives: A systematic literature review and meta-analysis on the validity (test-retest reliability and accuracy) of automated threshold audiometry compared to the gold standard of manual threshold audiometry was conducted.

Design: A systematic literature review was completed in peer-reviewed databases on automated compared to manual threshold audiometry. Subsequently a meta-analysis was conducted on the validity of automated audiometry. A multifaceted approach, covering several databases and employing different search strategies was utilized to ensure comprehensive coverage and to cross check search findings. Databases included: Medline, SCOPUS and PubMed with a secondary search strategy reviewing references from identified reports. Reports including within-subject comparisons of manual and automated threshold audiometry were selected according to inclusion/exclusion criteria before data was extracted. For the meta-analysis weighted mean differences (and SD's) on test-retest reliability for automated compared to manual threshold audiometry were determined to assess the validity of automated threshold audiometry.

© University of Pretoria

Results: In total, 29 reports on automated threshold audiometry (method of limits and the method of adjustment techniques) met the inclusion criteria and were included in this review. Most reports included data on adult populations using AC testing with limited data on children, BC testing and the effects of hearing status on automated threshold audiometry. Meta-analysis test-retest reliability for automated threshold audiometry was within typical test-retest variability for manual threshold audiometry. Accuracy results on the meta-analysis indicated overall average differences between manual and automated AC threshold audiometry (0.4 dB, 6.1 SD) to be comparable to test-retest differences for manual (1.3 dB, 6.1 SD) and automated (0.3 dB, 6.9 SD) threshold audiometry. No significant differences (p>0.01; Summarized Data ANOVA) were seen in any of the comparisons between test-retest reliability of manual and automated threshold audiometry compared to differences between manual and automated threshold audiometry.

Conclusion: Automated threshold audiometry provides an accurate measure of hearing threshold, but validation data is still limited for (i) automated BC audiometry; (ii) automated audiometry in children and difficult-to-test populations and; (iii) different types and degrees of hearing loss.

3.2. Introduction

Automated healthcare services may include screening, diagnostic and intervention procedures that can be conducted without the necessary healthcare professional's direct involvement. In situations where specialist healthcare personnel are limited or unavailable, this approach may ensure that services and healthcare resources are optimized (Margolis & Morgan, 2008; Swanepoel et al., 2010). Automated threshold audiometry has existed for many years; however it has not been used widely in clinical practice apart from occupational health care settings (Margolis & Morgan, 2008).

The earliest record of automated threshold audiometry was in the seminal report of Georg von Békésy (Békésy, 1947). This self-recording threshold audiometer automatically increased and decreased the sound intensity while sweeping through the test frequency range and became known as 'sweep frequency Békésy

© University of Pretoria

audiometry'. The patient is required to press a response button when the test signal is heard and release it when they lose perception of the signal. This method of determining the threshold is commonly known as the 'method of adjustment'. Subsequent systems utilized derivations of this technique with fixed frequency threshold seeking algorithms, referred to as fixed or discreet frequency Békésy audiometry, where a sweep in intensity occurs within a fixed frequency based on the patient's behavioural response relayed through a response switch (Frank, 2001; Meyer-Bisch, 1996).

In later years automated audiometry systems were programmed according to conventional manual audiometry procedural steps (Sparks, 1972), typically utilizing versions of the Hughson and Westlake threshold-seeking method (Hughson & Westlake, 1944). The audiometer automatically makes adjustments to the intensity of the presented signal, up or downwards depending on the response or lack of response. This method is known as the 'method of limits'. This method has also been modified in some cases to include forced-choice responses from the patient. Here the listener is required to listen and make a response that either indicates that a sound was heard or not. This can be done, for example, by pressing the appropriate 'button' on a touchscreen monitor after a signal was presented (Frank, 2001; Margolis & Morgan, 2008).

Pure tone threshold audiometry measures are especially suited to automation because they are based on predetermined sequenced steps (Margolis & Morgan, 2008). In addition, when using a computer, results can be recorded automatically enabling all the advantages of electronic record keeping, such as reduced paperwork, transfer to other clinicians and tracking change over time. Additionally, automated testing can incorporate quality-monitoring mechanisms to ensure consistent and reliable results as has recently been demonstrated (Margolis et al., 2007; Margolis et al., 2011). Automation may also potentially improve standardization of tests protocols and procedures across clinics and even within clinics.

At present the need for hearing health care services globally far outweighs the current capacity to deliver the services (Fagan & Jacobs, 2009; Goulios & Patuzzi,

© University of Pretoria

2008; Swanepoel, 2010; Margolis et al., 2010; 2011). Automated audiometry has been proposed as a way to increase the reach of audiometry in underserved areas especially when conducted within asynchronous telehealth framework (Swanepoel et al., 2010; Swanepoel & Hall, 2010). An automated audiometer cannot replace an audiologist, but a system that can determine pure-tone hearing thresholds with similar accuracy to that of manual audiometry may be beneficial in addressing the demand for hearing health services. Optimizing limited professional resources by incorporating automation may improve the reach of current audiological services and can improve the efficiency of current hearing health care resources (Margolis & Morgan, 2008; Swanepoel et al. 2010).

Although automated threshold audiometry has existed for many decades, it has been used almost exclusively in industry as part of mass hearing screening and baseline monitoring and for research purposes. Clinical audiological practices, in contrast, have almost exclusively relied on conventional manual audiometry. This may partly be attributed to perceived concerns regarding the accuracy and reliability of automated air conduction (AC) and bone conduction (BC) audiometry and the availability of validation studies (Margolis & Morgan, 2008; Sparks, 1972). However, being a behavioural test procedure manual audiometry also presents with variability in threshold determination (test-retest or inter-tester differences) due to subject factors such as fatigue and concentration as well as due to different transducers and test environments employed (ANSI, 1996; Margolis et al., 2007). Normal variability in audiometry has typically been quantified by test-retest reliability and occasionally by inter-tester reliability (Ishak et al., 2011; Margolis et al., 2007).

In the light of the potential benefits of automation in threshold audiometry, its long history and the apparent lack of summative evidence supporting its utilization, the present study aimed to systematically review the current body of peer-reviewed publications on the validity (test-retest reliability and accuracy) of automated threshold audiometry. Additionally the study included a meta-analysis, using results from published reports, to quantify the test-retest reliability and accuracy of automated threshold audiometry.

© University of Pretoria

3.3. Materials and methods

3.3.1. Systematic Review

A systematic review of peer-reviewed literature was conducted to determine the validity, as measured by the accuracy and reliability, of automated threshold audiometry compared to manual threshold audiometry. Accuracy is defined as the indirect method of measurement between two different techniques measuring the same variable of which one is the gold standard (Bland & Altman, 1999). Manual audiometry served as the gold standard and automated audiometry as the comparison method for determining auditory thresholds. Test-retest reliability refers to the ability of a test to give similar results when applied more than once on the same subjects under the same conditions (Dobie, 1983).

A varied search strategy was employed across several electronic databases to identify relevant research reports (excluding editorials, notes and short surveys) from peer-reviewed literature. For inclusion reports were required to include some within-subject comparison of automated threshold audiometry to manual threshold audiometry (accuracy). Test-retest reliability information was also captured from the identified reports.

A multifaceted approach, covering several databases and employing different search strategies, was utilized to ensure comprehensive coverage and crosschecking of search findings (White & Schmidt, 2005). An initial search strategy was undertaken using the following databases and search engines: Medline, SCOPUS and PubMed. Searches were conducted on 20 July 2012, including all relevant reports published until this date. Table 1 indicates the databases, search strategy and search terms employed.

The Medline database search utilized a strategy of relevant keywords to determine all records relating to the study aim (Table 1). The second database, PubMed, was searched using available Medical Subject Heading (MeSH) terms. SCOPUS, the third database included in the search strategy, is the world’s largest abstract and citation database of peer-reviewed literature also indexing Medline. This served as a cross-check for reports from PubMed and Medline databases.

© University of Pretoria

Table 1. Databases and search strategy details

	Search strategy	Identifiers	Results	Limiters
Medline	Articles reporting findings of automated audiological testing. Terms occurring in the title, abstract, or keywords of articles.	“Automatic” OR “computerized” OR “computer-based” OR "pc-based" OR "automation" OR "automated" OR “audioscan” AND "audiometry" OR "hearing measurement" OR "hearing thresholds" OR "auditory thresholds" OR "hearing assessment" OR "hearing evaluation"	463	Articles published prior to 1946 not included
PubMed	MeSH terms related to automated audiological testing, occurring in the title and abstract.	“automatic” OR “computerized” OR “computer-based” OR “pc-based” OR “automation” OR “automated” OR “audioscan” AND “audiometry”	195	MeSH terms utilized only
Scopus	Articles reporting findings of automated audiological testing. Terms occurring in all fields.	“automatic” OR “computerized” OR “computer-based” OR “pc-based” OR “automation” OR “audioscan” OR “automation” “automated”, “self-recording”, “self-recorded” OR “Békésy” AND “audiometry”, “hearing measurement”, “hearing thresholds”, “auditory thresholds”, “hearing assessment” and “hearing evaluation”.	1274	None

Inclusion and exclusion criteria: only reports of a comparative nature between automated and manual threshold audiometry, written in English were included. Descriptions of automated audiometry without these comparisons, reviews, articles, notes and short surveys were not included.

The first author reviewed the abstracts of all reports resulting from the searches to determine if the report complied with the inclusion criteria. If any queries arose the second author also reviewed the abstracts. Where an abstract was unavailable, the full paper was reviewed (Table 2). After all duplicates and unrelated reports had been excluded, the remaining reports were reviewed in full to determine if they meet the inclusion criteria. A secondary search was employed to supplement the findings of the primary search. This involved reviewing the reference lists of all reports already identified for inclusion during the primary search strategy for additional reports not identified with the primary search.

© University of Pretoria

Table 2. Results from the applied search strategies

Procedural steps	Number of reports	Description
1. Database search results	1932	3 Databases (Medline, PubMed, Scopus).
2. Database results excluding duplicates	1311	621 duplicates omitted.
3. Database results excluding non-English reports	1072	223 reports omitted.
4. Database results excluding reviews, short surveys and notes omitted	971	101 reports omitted.
5. Database results related to scope of review based on abstract and title	63	971 titles and abstracts reviewed for relevance, 908 records omitted, 63 complete articles reviewed.
6. Database results within scope of review based on full article	26	37 reports omitted based on inclusion/exclusion criteria. One could not be tracked due to incorrect indexing on the journal archive.
7. Additional reports within scope of review	3	3 reports identified from secondary search strategy surveying reference lists of 26 identified reports.
8. Final reports	29	Reports utilized in systematic review.
9. Reports utilized in meta-analysis	12	Reports with data appropriate to meta-analysis aims

The reports selected for review were carefully scrutinized and categorized according to the audiological threshold seeking method utilized (method of adjustment or method of limits), type of evaluation (diagnostic or screening), AC and/or BC thresholds, type of transducers and audiometer utilized, age and hearing status of participants, type of statistical analysis for accuracy, test-retest reliability and the conclusions drawn by the paper.

3.3.2. Meta-analysis

A meta-analysis was conducted to combine and quantify the results of individual reports so that an overall assessment of test-retest reliability and accuracy based on existing evidence could be made for automated audiometry. To be included in the meta-analysis, reports had to meet the following criteria: (i) The report had to include data comparing manual and automated audiometry in terms of accuracy; (ii) Data had to be reported in the form of mean differences (real or absolute) and standard deviations with the number of observations reported.

© University of Pretoria

Mean differences and standard deviations were documented. Weighted averages, utilizing reported real and absolute average differences and standard deviations, were determined for validation (test-retest reliability and accuracy) across studies, taking into account the number of observations reported. Furthermore a comparison of test-retest threshold differences for manual and automated threshold audiometry, indicative of normal variability, was made with the difference between automated and manual audiometry (accuracy) using an analysis of variance test (http://statpages.org/anovalsm.html). A significant difference in variability was noted by a p <0.01.

3.4. Results

3.4.1. Systematic review

The systematic review procedural outcomes are summarized in Table 2. After excluding duplicates, reviews, short surveys, notes and non-English language records, 971 reports remained. Sixty-three reports were identified and subsequently the full-text was reviewed. One report (Raza, 2008) could not be traced since its indexing on all databases did not correspond to the actual journal listing. Despite efforts to contact the authors and journal the report could not be sourced. A total of 26 full reports were identified that met the inclusion/exclusion criteria.

The second stage search strategy, involving a review of the reference lists of identified reports, revealed three additional reports, bringing the total number to 29 reports.

The final list of reports included in the systematic review date from 1956 to 2011 (Figure 2). Appendix B provides a summary of all reports included according to authors, year of publication, subject descriptions, test parameters, automated threshold seeking method (method of limits/method of adjustment), research findings (accuracy and/or test-retest reliability) and conclusions.

© University of Pretoria

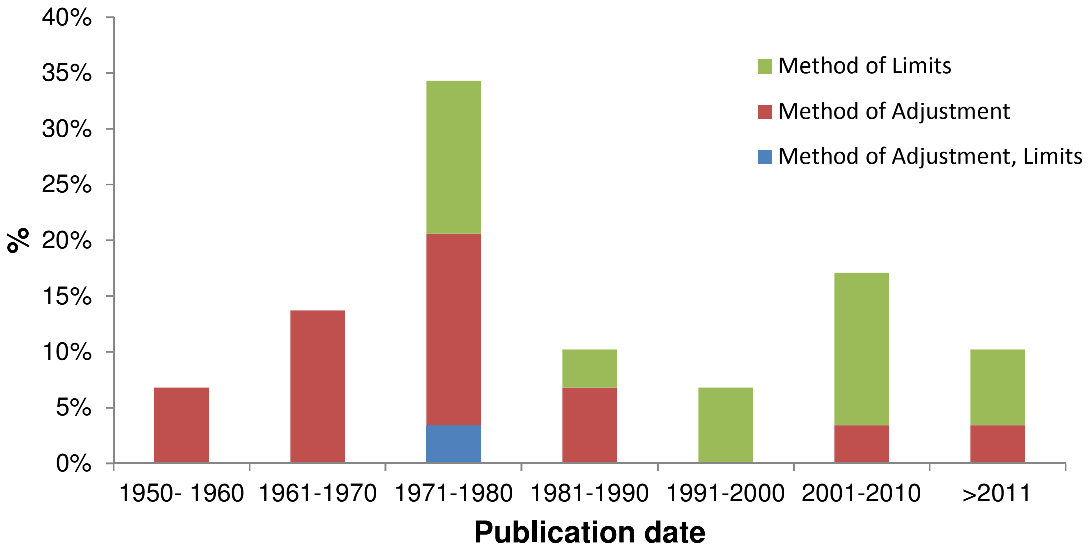


Figure 2. Distribution of reports included in systematic review (n=29) date of publication and type of automated audiometry (method of limits; method of adjustment, method of limits and adjustment).

Of the 29 reports, 15 utilized the method of adjustment and 13 the method of limits whilst one report utilized both methods (Harris, 1979). The majority of reports covered diagnostic audiometry whilst four reports included screening applications of automated audiometry (1 for method of limits, 3 for method of adjustment).

Table 3 provides a description of data on accuracy and test-retest reliability included in the systematic review records. Test-retest reliability was included by 11 reports (seven for method of adjustment and four for method of limits). Ten of these included only AC audiometry, whilst one included both AC and BC audiometry. Of these ten, three included participants with a hearing loss, whilst four did not indicate the hearing status of participants (Table 3).

© University of Pretoria

Table 3. Distribution of air and bone conduction data for adults and children reported across studies identified in the systematic review (n=29)

Type of hearing	Accuracy				Test-retest reliability			
	Normal hearing	**Hearing loss**	**∗Both**	**Not indicated**	**Normal hearing**	**Hearing loss**	**∗Both**	**Not indicated**
Adults								
AC testing	5	3	3	8	2	3	1	4
AC and BC testing	-	3	3	1	-	-	-	1
Subtotal	**5**	**6**	**6**	**9**	**2**	**3**	**1**	**5**
Children								
AC testing	1	-	1	1	-	-	-	-
AC and BC testing	-	-	2	-	-	-	-	-
Subtotal	**1**	**-**	**3**	**1**	**0**	**0**	**0**	**0**
TOTAL	**6**	**6**	**9**	**10**	**2**	**3**	**1**	**5**

∗*Indicating that both hearing and hearing loss subjects were included in the study.*

Records obtained reported data using a variety of statistical analyses (Table 4). The most common presentation of test-retest data was presented in terms of average differences and standard deviations (n=4) and average thresholds and standard deviations (n=3).

All 29 reports provided information on the accuracy of automated threshold audiometry. Twenty-six records reported results for adult populations, 19 of these included AC audiometry only, whilst, seven included AC and BC audiometry. Six of the 26 adult reports included persons with hearing loss only, five included persons with normal hearing, whilst six included persons with normal hearing or a hearing loss, and nine did not indicate the hearing status of their samples. Furthermore, only five of the studies reported results on children, two of which included AC and BC results.

Various techniques were utilized to document the accuracy, referred to as validity in records, of automated audiometry (Table 4). The most commonly used measures of accuracy were average differences between automated and manual audiometry with accompanying standard deviations (n=11) and average thresholds and standard deviations (n=11). Less commonly used techniques included absolute average differences and standard deviations (n=6), t-Test (n=4) and ANOVA analysis (n=2).

© University of Pretoria

Table 4. Statistical measures of accuracy and test-retest reliability employed in systematic review reports (n=29)

Type of analysis	Number of studies
Accuracy (threshold comparison with manual audiometry)	
Average differences and standard deviation	11
Average thresholds and standard deviation	11
Absolute average differences and standard deviation	6
t-Test	4
Linear regression and correlation coefficients	4
Pearsons product	3
Standard deviations only	3
ANOVA analysis	2
Average deviation	1
Error analysis	1
Contrast analysis	1
X^2 Test	1
Sensitivity and specificity analysis	1
Comparison of Kappa values of agreement	1
Standard error bars	1
Test of significance	1
Within subject variability test	1
F-ratio	1
Two way analysis of variance	1
Reliability coefficients- Hoyts solution	1
Sheffe’s test of statistical significance	1
Dispersion relationships	1
K-coefficients	1
Confidence intervals	1
Estimation of asymptomatic data	1
Test-retest reliability	
Average differences and standard deviation	4
Average thresholds and standard deviation	3
Absolute average differences and standard deviation	2
t-test	2
Pearson Product moment correlation coefficients	2
Standard deviation	1
Standard of variance	1
Standard error bars	1
k-coefficients	1
Repeated ANOVA	1
Variance of hearing threshold (σ^2)	1

3.4.2. Meta-analysis

The meta-analysis utilized mean differences (real and absolute) and standard deviations at each frequency extracted from the reports, if available. In some reports the mean differences and standard deviations across all frequencies were not determined and thus were calculated when possible (i.e. if the number of

© University of Pretoria

observations were included). Appendix C and Appendix D indicate summaries of the data obtained for test-retest reliability and accuracy across individual studies utilized in the meta-analysis. Weighted average calculations were subsequently obtained across these studies (Table 5 and Table 6).

Only five reports provided data on test-retest reliability in the form of mean differences (real and absolute) and standard deviations for automated testing and manual testing. Test-retest variability for automated threshold audiometry indicated average differences that ranged between -1.1 dB to 2.2 dB with the standard deviation ranging between 6.2 dB to 10.4 dB for individual test frequencies, whereas the absolute average differences ranged between 2.0 dB and 4.9 dB with a standard deviation of 3.0 dB to 4.8 dB (Table 5).

Table 6 provides a summary of weighted average differences between manual and automated audiometry, in the adult population. Results indicate that the overall (n=10) average differences between automated and manual audiometry ranged between -5.0 dB and 2.1 dB across the frequency spectrum with the standard deviations ranging from 5.3 dB to 8.7 dB. Furthermore, the average differences obtained between the automated method of limits and manual audiometry ranged between -1.7 dB and 3.8 dB with standard deviations between 4.4 dB to 7 dB. Additionally, method of adjustment audiometry yielded lower results at 0.125, 0.25, 6 and 8 kHz (-0.1 dB to -2.3 dB) whilst manual audiometry yielded higher results at the remaining frequencies, with the standard deviations ranging from 5.3 dB to 9 dB. The combined absolute differences ranged from 2.9 dB to 4.2 dB with standard deviations ranging from 3.2 dB to 4.5 dB.

Lastly, it should be noted that data from the two studies on children (four to ten years old) were excluded from the meta-analysis as only one study utilizing the method of adjustment (Békésy fixed frequency testing) reported results in the form of average differences. These ranged between 3.6 dB to 20.3 dB with standard deviations ranging from 2.6 dB to 7.2 dB for 0.25, 1 and 4 kHz (Hartley & Siengenthalar, 1964). Another study reported results in terms of absolute differences across all frequencies (4.1 dB), with a standard deviation of 1.7 dB (Margolis et al., 2011), when utilizing an automated method of limits technique.

© University of Pretoria

Analysis of variance comparisons of the meta-analysis weighted averages were conducted between the test-retest differences for manual and automated audiometry and the average difference between manual and automated thresholds (accuracy comparison) for the real and absolute differences. This was done for the combined category (method of limits and method of adjustment) and between method of adjustment and method of limits average differences. No statistically significant differences ($p>0.01$; Summarized Data ANOVA) were obtained between any of the comparisons of test-retest (manual and automated) threshold differences and automated compared to manual threshold differences.

© University of Pretoria

Table 5. Meta-analysis weighted average test-retest reliability differences for manual and automated audiometry

Frequencies	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz	All
MANUAL THRESHOLD AUDIOMETRY										
Average differences and standard deviations (3 reports)										
Average difference	-	-	2.3	2.1	1.5	2.0	-0.4	-1.7	-	1.3
n	-	-	500	500	500	40	500	40	-	532
Standard deviation	-	-	6.7	4.8	5.0	4.7	6.9	7.6	-	6.1
n	-	-	500	500	500	40	500	40	-	532
Absolute average differences and standard deviations (2 reports)										
Absolute average difference	4.8	3.4	2.9	3.2	2.7	-	2.8	-	3.0	3.2
n	60	80	80	80	80	-	80	-	80	80
Standard deviation	5	3.7	3.7	3.4	3.6	-	3.5	-	4.3	3.9
n	60	60	60	60	60	-	60	-	60	60
AUTOMATED THRESHOLD AUDIOMETRY										
Average differences and standard deviations (3 reports)										
Average difference	-	-	0.3	-1.1	0.0	2.1	0.7	1.7	-	0.3
n	-	-	500	500	500	40	500	40	-	532
Standard deviation	-	-	7.1	6.8	6.4	6.2	7.1	10.4	-	6.9
n	-	-	500	500	500	40	500	40	-	532
Absolute average differences and standard deviations (2 reports)										
Absolute average difference	4.9	3.4	2.9	2.6	2.6	-	2.3	-	2.0	2.9
n	60	80	80	80	80	-	80	-	80	80
Standard deviation	4.8	3.5	3.6	3.2	4.1	-	3.0	-	3.2	3.8
n	60	60	60	60	60	-	60	-	60	60

© University of Pretoria

Table 6. Weighted average differences and standard deviations between manual and automated threshold audiometry (manual minus automated)

Frequencies	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz	All
AVERAGE DIFFERENCES AND STANDARD DEVIATIONS										
Combined (10 reports)										
Average difference	-2.5	-3.5	-1.5	-1.2	-0.1	2.1	-3.6	-2.1	-5.0	0.4
n	232	360	796	796	796	428	796	556	384	820
Standard deviation	8.6	6.7	5.4	5.3	5.5	6.1	5.7	7.7	8.7	6.1
n	232	420	766	766	526	578	526	466	420	798
Method of limits (3 reports)										
Average difference	-	-0.4	-0.7	0.4	-1.3	-0.8	3.8	-1.3	-1.7	0.3
n	-	60	84	84	24	24	84	24	84	116
Standard deviation	-	5.1	4.4	5.3	5.8	-	4.9	-	7.0	5.5
n	-	60	60	60	60	-	60	-	60	92
Method of adjustment (7 reports)										
Average difference	-2.0	-2.3	0.5	0.3	2.1	1.1	0.1	-1.0	-3.1	0.8
n	232	360	796	796	796	428	796	556	384	796
Standard deviation	8.6	6.9	5.4	5.3	5.5	6.1	5.8	7.7	9.0	6.2
n	232	360	706	706	466	578	466	466	360	706
ABSOLUTE AVERAGE DIFFERENCES AND STANDARD DEVIATIONS										
Combined (4 reports)										
Absolute Average Difference	4.2	3.6	3.4	3.5	3.4	-	2.9	-	3.1	4.2
n	136	196	196	196	196	-	196	-	196	360
Standard deviation	4.0	3.5	3.9	3.6	3.8	-	3.2	-	4.5	5.0
n	136	196	196	196	196	-	196	-	196	345

© University of Pretoria

3.5. Discussion

Comparing of two audiometric threshold techniques, such as automated and manual audiometry, has been performed using a variety of statistical analyses (Table 4). Measures of agreement determined by the two threshold seeking methods most commonly included the average difference (with SD), average thresholds (with SD) and average absolute differences (with SD). The average difference is valuable to show a systematic effect but negative and positive difference may cancel each other out even when large differences in either direction exist. Bland and Altman (1986) recommend the use of absolute average differences and standard deviation as a more appropriate measure of correspondence since it provides an indicator of the expected spread in variability. With this in mind, the meta-analysis the meta-analysis was conducted using average differences (real and absolute) and standard deviations to draw conclusions regarding the validity of automated audiometry when compared to manual audiometry.

3.5.1. Automated audiometry test-retest reliability

Test-retest reliability is defined as the repeatability of a technique and allows comparison of techniques to determine which is more precise (Bland & Altman, 1986). Eleven reports in this systematic review included results on test-retest reliability, of which four used the ‘method of limits’ and seven the ‘method of adjustment’ for threshold audiometry. In each case, reported test-retest reliability for automated audiometry was indicated to be within typical variability when compared to the test-retest reliability of manual audiometry (Burns & Hinchcliffe, 1957; Erlandsson et al., 1979a; 1979b; Fautsi et al., 1990; Formby et al., 1996; Gosztonyi et al., 1971; Ho et al., 2009; Ishak et al., 2011; Lutman et al., 1989; Robinson & Whittle, 1973; Swanepoel et al., 2011). Only Ishak et al. (2011) reported higher test-retest variability with Bèkèsy sweep- frequency audiometry, but reported that employing a slower sweep rate of 20 sec per octave would improve the acquired test-retest reliability.

Several reports indicated that the second test session produced slightly lower (i.e. better) thresholds than the first session when utilizing manual and automated audiometry (Burns & Hinchcliffe, 1957; Erlandsson et al., 1979a; 1979b; Fautsi et al.,

© University of Pretoria

1990; Fromby et al., 1996; Gosztonyi et al., 1971; Ho et al., 2009; Ishak et al., 2011; Luteman et al., 1989; Robinson & Whittle, 1973; Swanepoel et al., 2011). Several of the reports attributed the lower thresholds during the second session to the learning effect (Erlandsson et al., 1979a; 1979b; Ishak et al., 2011; Luteman et al., 1989). This suggests that subsequent studies should consider randomizing the order of testing techniques and control the previous experiences participants had with audiometric testing.

The meta-analysis showed overall test-retest variability for automated (5 reports) and manual AC audiometry (5 reports) to be similar. Average differences obtained for manual and automated test-retest audiometry respectively were, 1.3 dB (6.1 SD) and 0.3 dB (6.9 SD) and absolute differences of 3.2 dB (3.9 SD) and 2.9 dB (3.8 SD). The meta-analysis test-retest difference for automated compared to manual audiometry (Table 5), demonstrated no statistically significant difference (ANOVA; p>0.01). Higher variability was noted at 6 kHz for both automated and manual AC audiometry, but this was because only one paper reported data at 6 kHz (Burns & Hinchcliffe, 1957). Burns and Hinchcliffe (1957) reported a high variability for 6 kHz, with standard deviations of 3 dB to 4 dB, higher than those obtained at the other tested frequencies in the study (Appendix C).

Meta-analysis test-retest results are consistent with previously reported standard deviations of average test-retest differences for manual audiometry, ranging between 4.4 dB and 6.2 dB for a group of adults and children (Stuart et al., 1991). A recent report (Swanepoel & Biagio, 2011) on manual audiometry obtained absolute average test-retests differences (3.6 dB, 3.9 SD) that were in line with the meta-analysis results (2.9 dB, 3.8 SD). The AC test-retest threshold differences for automated audiometry fall well within current test-retest limits.

Ho et al. (2009) was the only study to report on automated BC test-retest reliability. Results were reported in terms of paired thresholds, the study concluded that test-retest reliability of automated BC audiometry was appropriate (Ho et al., 2009) and within typical manual BC test-retest reliability (Laukli & Fjermedal, 1990; Margolis et al., 2010; Swanepoel & Biagio, 2011).

© University of Pretoria

3.5.2. Automated audiometry accuracy

Over the six decades since the first description of automated audiometry, only 29 reports (15 on method of adjustment, 13 on method of limits, and one utilizing both method of limits and adjustment), have reported on the validation of automated audiometry by comparing results to the gold standard of manual audiometry.

The meta-analysis showed that overall average differences between manual and automated AC audiometry (0.4 dB, 6.1 SD) correspond to test-retest difference for manual (1.3 dB, 6.1 SD) and automated (0.3 dB, 6.9 SD) audiometry. No statistically significant difference (ANOVA; p>0.01) was evident between overall absolute differences for manual and automated audiometry (4.2 dB, 5.0 SD) and the test-retest absolute differences for manual (3.2 dB, 3.9 SD) and automated (2.9 dB, 3.8 SD) audiometry (Table 5).

Average differences for manual and automated BC audiometry were only reported by nine studies. These studies utilised varied forms of analyses in term of agreement (Table 4) and as a result weighted averages for bone conduction threshold audiometry could not be determined across studies.

3.5.2.1. Method of adjustment

As demonstrated in Figure 2 the method of adjustment was the first type of automated threshold audiometry. Overall 16 reports were identified including comparisons of manual and method of adjustment automated threshold audiometry. The manual audiometry threshold determination techniques in these reports included the modified Hughson-Westlake method and some variations thereof (Burns et al., 1957; Corso, 1956; Erlandsson et al., 1979a; 1979b; Hartley et al., 1964; Ishak et al., 2011; Jokinen, 1969; Knight, 1965; Robinson & Whittle, 1973) as indicated in Appendix B.

Several reports included in the systematic review indicated that automated audiometry using the method of adjustment (Békésy sweep or Békésy fixed frequency method) generally yields lower (i.e. better) thresholds compared to manual audiometry (Burns et al., 1957; Corso, 1956; Erlandsson et al., 1979a; 1979b;

© University of Pretoria

Frampton & Courter, 1989; Harris, 1979; Hartley et al., 1964; Ishak et al., 2011; Jokinen, 1969; Knight, 1965; Maiya & Kacker, 1973; Robinson & Whittle, 1973). A single report showed manual audiometry having lower thresholds than the method of adjustment technique at certain frequencies (0.25, 6 and 8 kHz). The authors reported that the reason for this phenomenon was probably the threshold seeking method utilized (Ishak et al., 2011).

The meta-analysis showed an average differences of 0.8 dB (6.2 SD) between automated (method of adjustment) and manual AC audiometry. There was no statistically significant difference (ANOVA; p>0.01) when these results were compared to test-retest reliability of both manual (1.3 dB, 6.1 SD) and automated threshold audiometry (0.3 dB, 6.9 SD). The accuracy of automated (method of adjustment) threshold audiometry is therefore within the normal variability as defined by test-retest reliability. Margolis et al. (2010) compared automated and manual threshold differences between two audiologists using manual audiometry as opposed to test-retest reliability. The inter-tester differences (0.6 dB, 5.5 SD) for manual audiometry were similar to the average differences (0.8 dB, 6.2 SD) between manual and automated audiometry results obtained in the meta-analysis.

Four reports included screening audiometry, comparing manual and automated thresholds (method of adjustment). Three of these studies utilized children (Delany et al., 1966; Hartley & Siengenthalar, 1964; McPherson et al., 2011) and one utilized an adult population (Gosztonyi et al., 1971). Delany et al. (1966) indicated that automated audiometry for participants provided results substantially in agreement with manual audiometry, however, as observed with adults, automated audiometry tends to produce thresholds that are slightly lower (-0.8 to -3.3 dB) than manual testing. Additionally, the authors (Delany et al., 1966; Hartley & Siengenthalar, 1964; McPherson et al., 2010) indicated that automated audiometry can produce useful threshold data with children down to the age of six years. As age decreases however, a greater proportion of children are either unable to perform the test at all or frequently lose concentration so that portions of the test need to be repeated at a later stage to obtain a full audiogram.

© University of Pretoria

Gosztonyi et al (1971) reported on industrial screening conducted on salaried and hourly workers (n=38 ears). This study indicated that manual audiometry thresholds may be significantly lower than automated thresholds but the authors later discovered that the reason for this phenomenon was the fact that all participants involved in medicolegal cases. Thus the phenomenon of non-organic hearing loss significantly increased the threshold differences obtained between manual and automated audiometry.

Although findings on the application of automated audiometry using the method of adjustment are promising, limited data are available for paediatric populations and BC testing. An important reason for no BC data in the method of adjustment technique is attributed to the difficulty in utilizing masking with this method. It is challenging to use a masking noise on the contralateral ear as the narrow-band noise level should theoretically change with the tested frequency (Meyer-Bisch, 1996). Additional to the technical difficulties of such an operation, the test may become difficult to follow for the patient (Meyer-Bisch, 1996).

3.5.2.2. Method of limits

In the 1970's the focus of research on automated audiometry started to shift from method of adjustment techniques to the method of limits (Figure 2). Overall 13 reports utilized the method of limits for automated audiometry compared to manual audiometry. All the studies obtained in the systematic literature review reported no statistically significant difference for AC between manual and automated audiometry.

Meta-analysis weighted average difference (0.3 dB, 5.5 SD) obtained when comparing automated method of limits technique to manual audiometry was similar to the weighted average difference for the method of adjustment and manual audiometry (0.8 dB, 6.2 SD), no statistically significant difference was noted (ANOVA; p>0.01). These findings correspond to test-retest reliability results of automated (1.3 dB, 6.1 SD) and manual (0.3 dB, 6.9 SD) audiometry, indicating no statistically significant difference (ANOVA; p>0.01). The accuracy of method of limits automated audiometry is within normal variability as defined by test-retest reliability.

© University of Pretoria

Seven of the thirteen reports included findings on BC audiometry (Ho et al., 2009; Margolis et al., 2007; Margolis et al., 2010; Margolis & Moore, 2011; Picard et al., 1993; Sparks, 1972; Wood et al., 1973). No statistically significant difference between manual and automated BC audiometry was noted across these studies. Margolis and Moore (2011) indicated a statistically significant difference between AC thresholds for manual and automated audiometry. The difference was partly attributed to the different transducers used (manual- TDH 50, automated- Sennheiser HDA 200) and the differential effect of low and high frequencies being tested.

3.6. Conclusion

Automated threshold audiometry developed over six decades from method of adjustment (Békésy methods) procedures to automated audiometry incorporating conventional manual audiometry (method of limits) threshold seeking methods. Current evidence demonstrates similar test-retest reliability for automated compared to manual threshold audiometry and automated audiometry thresholds being within typical test-retest and inter-tester variability of manual thresholds. Despite its long history however, validation is still limited for (i) automated BC audiometry; (ii) automated audiometry in children and difficult-to-test populations and; (iii) different types and degrees of hearing loss.

© University of Pretoria

4. DISCUSSION AND CONCLUSION

4.1. Discussion of results

Automation of procedures to determine hearing thresholds is not a new phenomenon but dates back more than six decades (Békésy, 1947). Apart from the use of self-recording Békésy audiometry as a site-of-lesion test, automated threshold audiometry has not entered routine clinical audiological practice. However, new developments and an increasing appreciation for the global dearth of hearing health services have brought automated audiometry into the spotlight.

Several studies have been conducted independently on the validity of automated threshold audiometry; however, a consensus regarding its usage in the clinical setting has not been made (Burns & Hichcliffe, 1957; Erlandsson et al., 1979a; 1979b; Harris, 1979; Swanepoel et al., 2010). This precipitated the need for a systematic review and meta-analysis in order to combine and quantify the results of individual reports so that an overall assessment of test-retest reliability and accuracy based on existing evidence could be made for automated audiometry. The systematic literature review revealed a total of 29 reports, dating from 1956 to 2011 (Figure 2). Of the 29 reports, 15 reports utilized the method of adjustment and 13 the method of limits whilst one report utilized both methods (Harris, 1979). The majority of reports covered diagnostic audiometry whilst four reports included screening applications of automated threshold audiometry (one for method of limits, three for method of adjustment).

In order to reach a general consensus, amongst the systematic review articles, regarding the validity of automated threshold audiometry, normal variability had to be considered. Normal variability has typically been quantified by test-retest reliability and occasionally by inter-tester reliability (Margolis et al., 2007; Swanepoel & Biago, 2011) of manual threshold audiometry. With this in mind the study findings are discussed according to test-retest reliability thresholds obtained automatically compared to test-retest thresholds obtained manually. It should be noted that results could not be reported specifically for method of limits and method of adjustment due to the limited amount of data on each. Thus an overall conclusion of test-retest

© University of Pretoria

reliability of automated audiometry was made. Test-retest reliability results were then compared to accuracy results obtained automatically in terms of the different methods utilized (method of adjustment and method of limits).

4.1.1. Test-retest reliability of automated audiometry

Test-retest reliability is defined as the repeatability of a technique and allows comparison of techniques to determine which is more precise (Bland & Altman, 1986). Eleven reports included results on test-retest reliability, of which four used the method of limits and seven the method of adjustment for threshold audiometry. Only five reports provided data on test-retest reliability in the form of mean differences (real and absolute) and standard deviations for automated and manual threshold audiometry. Automated threshold audiometry (3.8 SD) demonstrated test-retest reliability similar to that of manual threshold audiometry (3.9 SD) and similar to previously reported test-retest ranges (4.4 SD to 6.2 SD) for audiometry with various types of transducers for a group of adults and children (Stuart, Stenstorm, Tompkins & Vandenhoff, 1991). Furthermore, a recent report (Swanepoel & Biagio, 2011) on manual audiometry obtained absolute average test-retests differences (3.6 dB, 3.9 SD) that were in line with the meta-analysis results.

Independently, most of the reports included in the systematic review reported that automated audiometry (3.8 SD) demonstrated test-retest reliability similar to that of manual audiometry (Burns & Hinchcliffe, 1957; Erlandsson et al., 1979a; 1979b; Fautsi et al., 1990; Fromby et al., 1996; Gosztonyi et al., 1971; Ho et al., 2009; Ishak et al., 2011; Luteman et al., 1989; Robinson & Whittle 1973; Swanepoel et al., 2011). Only Ishak et al. (2011) reported higher test-retest variability with Bèkèsy audiometry, however, the author indicated that employing a slower sweep rate of 20 seconds per octave would improve the acquired test-retest reliability. Higher variability was also noted at 6 kHz for both automated and manual AC audiometry, but this was because only one paper reported data at 6 kHz (Burns & Hinchcliffe, 1957).

Furthermore, a number of reports in the systematic literature review indicated that the second test session produced slightly lower (i.e. better) thresholds than the first

© University of Pretoria

session when utilizing manual and automated threshold audiometry (Burns & Hinchcliffe, 1957; Erlandsson et al., 1979a; 1979b; Fautsi et al., 1990; Fromby et al., 1996; Gosztonyi et al., 1971; Ho et al., 2009; Ishak et al., 2011; Luteman et al., 1989; Robinson & Whittle, 1973; Swanepoel et al., 2011). Several of the reports attributed the lower thresholds during the second session to be due to the learning effect (Erlandsson et al., 1979a, 1979b; Ishak et al., 2011; Luteman et al., 1989). This suggests that subsequent studies should consider randomizing the order of testing techniques and control the previous experiences participants had with audiometric testing. In addition, it was noted that test-retest correspondence for automated audiometry was slightly better than manual threshold audiometry as indicated in the meta-analysis, although not statistically significant. A similar pattern was previously reported by Jerlvall, Dryselius and Arlinger (1893) in their comparison of manual and automated threshold audiometry. The objective nature of the automated procedure means that threshold-seeking procedures and criteria for determining the threshold are consistently applied without bias. Although this should be the case in manual audiometry that follows a set threshold-seeking procedure and threshold criteria, there is still subjective decision making on the part of the clinician. Furthermore, manual audiometry is biased by the prior knowledge the tester acquires during the test. This phenomenon is noted when one threshold is found placing constraints on what other thresholds could be affecting the variability of manual thresholds in a way that automated thresholds are not affected. As a result, automated threshold audiometry may present with slightly better test-retest reliability based on group data and also avoid any possibility of variability between clinicians (Jerlvall et al.,1893).

Ho et al. (2009) was the only study to report on automated BC test-retest reliability. Results were reported in terms of paired thresholds. The study concluded that test-retest reliability of automated BC threshold audiometry was accurate (Ho et al., 2009) and were found to be in accordance with reported typical manual BC test-retest reliability (Laukli & Fjermedal, 1990; Margolis et al., 2010; Swanepoel & Biagio, 2011).

© University of Pretoria

It can be concluded that the AC test-retest threshold differences for automated audiometry fall well within current test-retest limits. Thus it can be noted that the validity of automated audiometry in terms of test-retest reliability is reliable and valid.

4.1.2. Accuracy of automated audiometry

Accuracy of automated audiometry is defined according to its correspondence to the gold standard of conventional manual audiometry. Thus accuracy can be seen as the difference obtained between thresholds obtained automatically and those obtained manually. Only 29 reports (15 on method of adjustment, 13 on method of limits, and one utilizing both method of limits and adjustment), over the six decades since the first description of automated audiometry, have reported on the accuracy of automated audiometry. The majority of these reports covered diagnostic AC audiometry on adults whilst limited data was reported for the paediatric populations and BC threshold testing.

Overall, the meta-analysis results between automated threshold audiometry (method of limits and method of adjustment) and manual threshold audiometry revealed no statistical significance (ANOVA; p>0.01). Furthermore, a comparison between absolute average differences for accuracy and test-retest for manual and automated threshold audiometry results was made. Again no statistically significant difference was noted (ANOVA; p>0.01). It can thus be seen that overall average differences between manual and automated AC threshold audiometry (0.4 dB, 6.1 SD) correspond to test-retest differences for manual (1.3 dB, 6.1 SD) and automated (0.3 dB, 6.9 SD) threshold audiometry.

In the past, automated threshold audiometry did not include BC audiometry (Margolis & Morgan, 2008). Advances in technology now make it feasible to include BC testing and contralateral masking when appropriate (Margolis et al., 2010). However, due to the limited and only recent interest in BC audiometry, studies utilised varied forms of analyses in terms of agreement and as a result weighted averages for BC threshold audiometry could not be determined across studies.

Lastly it should be noted that the time required to perform manual and automated threshold audiometry was similar (Harris, 1979; Jerlvall et al., 1983; Swanepoel et

© University of Pretoria

al., 2010). Furthermore, Swanepoel et al. (2010) reported that subjects preferred automated audiometry threshold seeking methods due to the predictable nature of the threshold presentation. Picard, Liecki and Baxter (1988) reported that changes to the presentation of stimuli in a more arrhythmic configuration may in fact be more appropriate to avoid false-positive responses due to anticipation of the listener.

4.1.2.1. Accuracy of method of adjustment

The systematic literature review revealed that the method of adjustment was the first type of automated threshold audiometry to be investigated from 1950's. Thereafter a reduction in interest occurred during 1980 to the 2000's (31%); however, a recent increase in interest has occurred since 2000. Overall 16 reports were identified including comparisons of manual audiometry to automated method of adjustment audiometry. The manual audiometry threshold determination techniques in these reports included the modified Hughson-Westlake method and some variations thereof (Burns & Hichcliffe, 1957; Corso, 1956; Erlandsson et al., 1979a; 1979b; Hartley & Siengenthalar, 1964; Ishak et al., 2011; Jokinen, 1969; Knight, 1965; Robinson & Whittle, 1973).

The meta-analysis included data obtained from seven reports obtained in the systematic review. Results indicated no statistically significant difference (ANOVA; p>0.01) between the average differences of automated (method of adjustment) and manual AC audiometry (0.8 dB, 6.2 SD) when compared to test-retest reliability of both manual (1.3 dB, 6.1 SD) and automated threshold audiometry (0.3 dB, 6.9 SD). The accuracy of automated (method of adjustment) threshold audiometry is therefore within the normal variability as defined by test-retest reliability. These findings are comparable to results reported by Margolis et al. (2010) when comparing automated thresholds to manual threshold differences obtained between two audiologists using manual audiometry as opposed to test-retest reliability. The inter-tester differences (0.6 dB, 5.5 SD) for manual threshold audiometry were similar to the average differences (0.8 dB, 6.2 SD) between manual and automated threshold audiometry results obtained in the meta-analysis.

© University of Pretoria

Findings reported by individual studies in the systematic review indicated that diagnostic automated threshold audiometry using the method of adjustment (Békésy sweep or Békésy fixed frequency method) generally yields lower (i.e. better) thresholds compared to manual audiometry (Burns & Hichcliffe, 1957; Erlandsson et al., 1979a; 1979b; Frampton & Courter, 1989; Harris, 1979; Ishak et al., 2011; Jokinen, 1969; Knight, 1965; Maiya & Kacker, 1973; Robinson & Whittle, 1973;). Delany, Whittle and Knox (1966) reported similar findings when testing children. A single report showed manual audiometry having lower thresholds than the method of adjustment technique at certain frequencies (0.25, 6 and 8 kHz). The authors reported that the reason for this phenomenon was probably due to the threshold seeking method utilized (Ishak et al., 2011). Furthermore, Maiya and Kacker (1973) advised that midpoint tracing be utilized when obtaining thresholds. The authors indicated that midpoint tracing produced results that were in close agreement with thresholds obtained manually at all frequencies (0.125 to 8 kHz), whereas upper and lower limit points did not agree with thresholds obtained manually. Whereas, Jokinen (1969) indicated that interrupted Békésy obtained the best thresholds; the reason for this phenomenon is that a pulsed tone is easier to listen to and that adaption can take place when presenting continues tones versus interrupted tones.

The study by Gosztonyi et al. (1971) was the only report that reported findings of industrial screening on adults (n=38 ears). This study indicated that manual audiometry thresholds may be significantly lower than automated thresholds but the authors later discovered that the reason for this phenomenon was due to the fact that all participants were involved in medico-legal cases. Thus the phenomenon of significantly increased the threshold differences obtained between manual and automated audiometry. However, recent research findings (Margolis et al., 2007; Swanepoel et al., 2010) have indicated that diagnostic audiometry software may include quality monitoring indicators. These include recording the patient response time for each response, and if inconsistent or long response times occur when compared with normative ranges, a malingerer or patient requiring reinstruction can be noted. A number of such indices, including false-positive and true-positive response rates, may objectively aid the interpretation of test findings and resolve the dilemma of utilizing automated threshold audiometry when testing malingerers or difficult to test patients.

© University of Pretoria

Three reports obtained in this review utilized children (Delany et al., 1966; Hartley & Siengenthalar, 1964; McPherson, Law & Wong, 2011). The authors (Delany et al., 1966; Hartley & Siengenthalar, 1964; McPherson et al., 2010) indicated that automated threshold audiometry can produce useful threshold data with children down to the age of six years. As age decreases however, a greater proportion of children are either unable to perform the test at all or frequently lose concentration so that portions of the test needed to be repeated at a later stage to obtain a full audiogram. However, limited data exists with regards to the paediatric population.

Findings on the application of automated threshold audiometry using the method of adjustment are promising. However, limited data is available for paediatric populations and BC testing. An important reason for no BC data in the method of adjustment technique is attributed to the difficulty in utilizing masking with this method. It is challenging to use a masking noise on the contralateral ear as the narrow-band noise level should theoretically change with the tested frequency (Meyer-Bisch, 1996). Additional to the technical difficulties of such an operation, the test may become difficult to follow for the patient (Meyer-Bisch, 1996).

4.1.2.2. Accuracy of method of limits

The method of adjustments was the primary interest of automated threshold audiometry until the 1970's when an interest in the method of limits occurred; thereafter the method of limits became the primary interest of automated threshold audiometry. Overall 13 reports utilized the method of limits for automated audiometry compared to manual audiometry. Only three reports presented data that could be utilized in the meta-analysis. No statistically significant difference was noted (ANOVA; p>0.01) when weighted average difference (0.3 dB, 5.5 SD) obtained using the automated method of limits technique to manual threshold audiometry were compared to weighted average difference for the method of adjustment and manual threshold audiometry (0.8 dB, 6.2 SD). Furthermore these findings correspond with test-retest reliability results of automated (1.3 dB, 6.1 SD) and manual (0.3 dB, 6.9 SD) audiometry, indicating no statistically significant difference (ANOVA; p>0.01). It can thus be seen that the accuracy of method of limits automated threshold audiometry is within normal variability as defined by test-retest reliability.

© University of Pretoria

As with the method of adjustments, results from the meta-analysis indicate findings that are comparable to results reported by Margolis et al. (2010) when comparing automated thresholds to manual threshold differences obtained between two audiologists using manual audiometry as opposed to test-retest reliability. The inter-tester differences (0.3 dB, 5.5 SD) for manual threshold audiometry were similar to the average differences (0.8 dB, 6.2 SD) between manual and automated threshold audiometry results obtained in the meta-analysis.

Most of the studies obtained in the systematic literature review reported no statistically significant difference for AC between manual and automated threshold audiometry (Almqvist & Aursnes, 1978; Fautsi et al., 1990; Fromby et al., 1996; Harris, 1979; Ho et al., 2009; Margolis et al., 2007; Margolis et al., 2010; Sakade Hirai & Itami, 1978; Sparks, 1972; Swanepoel et al., 2010). Margolis and Moore (2011) indicated a statistically significant difference between AC thresholds for manual and automated threshold audiometry. The difference was partly attributed to the different transducers used (manual- TDH 50, automated- Sennheiser HDA 200) and the differential effect of low and high frequencies being tested.

No statistical significant difference was reported for the three studies that utilized children (Delany et al., 1966; Hartley & Siengenthalar, 1964; McPherson et al., 2010). However, Margolis et al. (2011) had to utilize a quality assessment method in order to obtain automated threshold results that were comparable to manual threshold audiometry. Margolis et al. (2007) developed a quality assessment method (QUALIND) based on a comparison of audiograms obtained utilizing automated (AMTAS) and manual testing. A predictive equation was derived from a multiple regression of a set of quantitative quality indicators on a measure of test accuracy, defined as the average absolute difference between automated and manually tested thresholds. Margolis et al. (2011) used QUALIND to identify and exclude ‘poor’ audiograms and measured the accuracy of automated testing by comparing it to the normal variability between thresholds recorded by two different audiologists testing the same population (Margolis et al. 2007). In this particular study a group of adults (n=30 ears) as well as a group of children were used (n=136 ears). The differences obtained between automated (AMTAS/KIDTAS) and manual threshold audiometry

© University of Pretoria

(4.1dB) produced differences with variability that were comparable to thresholds obtained using manual testing by two audiologists (3.9 dB), only if QUALIND identified and excludes 'poor' audiograms.

Seven of the thirteen reports included findings on BC audiometry (Ho et al., 2009; Margolis et al., 2007; 2010; Margolis & Moore, 2011; Picard et al., 1993; Sparks, 1972; Wood et al., 1973). However, as previously indicated, studies utilised varied forms of analyses in terms of agreement and as a result weighted averages for BC threshold audiometry could not be determined across studies. None the less, no statistically significant difference between manual and automated BC threshold audiometry was noted across these studies when reviewed independently.

Findings on the application of automated threshold audiometry using the method of limits are promising. However, limited data is available for paediatric populations and BC testing.

4.2. Clinical implications and recommendations

Clinical audiologists usually see a number of patients a day and are often expected to complete full audiological and hearing aid evaluation with subsequent counselling and aural rehabilitation. These are often time-consuming responsibilities for a clinician and any reliable method, which could aid the audiologist in performing these duties, may be of great benefit to increase the number of patients served and perhaps even the quality of these services. A significant part of an audiologist's time may be spent in obtaining AC and BC thresholds. If these pure tone thresholds could be obtained reliably through automatic processes it would obviously free the tester to see more patients and spend more time on other tasks (Margolis & Morgan, 2008). Automated audiometry may therefore prove to be cost effective and time efficient.

Pure tone audiometry measures are especially suited to automation as they are based on predetermined sequenced steps (Margolis & Morgan, 2008). These steps are based on reliable algorithms and can be transferred to a computer capable of reproducing them. Reports included in the review have demonstrated that automated audiometry can provide reliable and efficient diagnostic AC results in adult

© University of Pretoria

populations. Integrated with a telemedicine approach, automated threshold audiometry can ensure that diagnostic audiometric services are provided in areas where specialist personnel may be limited or unavailable (Swanepoel, Clark, Koekemoer, Hall, Krum & Ferrari, 2010; Swanepoel et al., 2010). Swanepoel et al. (2010) has reported that the combination of tele-audiology and automated audiometry can be a powerful way of providing time and resource efficient audiometric evaluations, especially in regions where audiological services are unavailable, which is true for most of Africa. Furthermore, results from automated threshold audiometry can be uploaded to be reviewed by specialists in off-site clinics. In some complex or difficult to test cases, automated threshold audiometry may include quality monitoring indicators where patient response time can be recorded for each response, and if inconsistent or long response times, compared with normative ranges, are noticed, it may indicate a malingerer or patient requiring reinstruction (Swanepoel & Biago, 2010). Furthermore, automated threshold audiometry can utilize features such as the QUALIND to be able to identify ‘poor’ or inaccurate audiograms (Margolis et al., 2010).

Although there is great benefit in using automated threshold audiometry, there are also limitations that require consideration. Audiologists play a decisive role in determining reliable thresholds in children and other difficult-to-test patients. Automated testing is not a replacement for the highly skilled audiologist trained to determine threshold for a variety of special needs and difficult to test patients. In fact an important limitation of automated threshold audiometry is the lack of research data on its validity in assessing children. According to a study conducted by Margolis et al. (2010) not all children who participated in the study were able to complete the task. It was indicated that modification of the instructions with a demonstration (either live or pre-recorded) could increase interest and de-sensitise the child to the procedure, to obtain better and more reliable results. Other limitations include a shortage of research evidence for the validity of automated bone conduction audiometry and comparative studies in patients with various types and degrees of hearing loss.

© University of Pretoria

4.3. Critical evaluation

A critical evaluation of the research project is crucial in order to interpret the findings of research within the framework of its strengths and limitations. These are highlighted below:

Strengths of study

Apart from a number of isolated independent studies conducted on automated threshold audiometry, an apparent lack of summative evidence supporting its utilization has existed. This research project attempted to determine the validity of automated threshold audiometry in terms of test-retest reliability and accuracy. Firstly, this study reviewed all available literature across the decades dating back to the 1950's. Secondly, it attempted to combine and quantify all available data on its validity in order for a general consensus to be made regarding the use of automated threshold audiometry in the clinical setting. Thirdly, limitations in the use of automated threshold audiometry were determined. These limitations include lack of research evidence regarding automated threshold audiometry in the paediatric population as well as lack of data regarding automated BC audiometry as well as comparative studies in patients with various types and degrees of hearing loss. Fourthly, by identifying the gaps in available research findings regarding automated threshold audiometry, recommendations for future research could be determined.

Limitations of study

The greatest limitation of this study was that the statistical representation of data across studies varied, thus, only a limited number of studies could be utilized in the meta-analysis. Secondly, the use of automated threshold audiometry across various types and degrees of hearing status could not be determined. Thirdly, the effect of the different types of transducers utilized for manual and automated threshold audiometry could not be obtained. Lastly, due to the cut off period for data collection, July 2012, articles published after this date could not be considered for inclusion in this study.

© University of Pretoria

4.4. Future research

This study provided important information on the validity of automated threshold audiometry and its use in the clinical setting. Reported results created a potential for future research regarding a number of aspects.

Firstly, more research needs to be conducted on the use of BC automated threshold audiometry. Secondly, the use of automated threshold audiometry with the paediatric population needs to be investigated (Margolis et al., 2010). Thirdly, the effect that different transducers have on automated and manual threshold audiometry needs to be investigated (Margolis et al., 2011). Fourthly, it needs to be determined whether the type of hearing status of a patient can influence the validity of automated threshold testing. Lastly, it should be noted that further research findings regarding automated threshold audiometry be reported in terms of average (real and absolute) differences and standard deviations as recommended by Bland and Altman (1986) in order for comparison to be made.

4.5. Conclusion

Automated threshold testing (method of limits and method of adjustment) has existed for many decades and has evolved over time from method of adjustments (Békésy methods) to method of limits. Current evidence demonstrates similar test-retest reliability for automated and manual threshold audiometry and automated thresholds being within typical test-retest and inter-tester variability of manual thresholds. However, despite the long history of automated audiometry validation is still limited in a number of areas. These include automated BC audiometry; automated audiometry in children and difficult-to-test populations and the effect of different types and degrees of hearing loss on automated threshold audiometry.

© University of Pretoria

5. REFERENCES

Almqvist, B., & Aursnes, J. (1978). Computerized pure tone audiometry. *Scandinavian Audiology, Supplement, 8*, 193-196.

Békésy, V. G. (1947). A new audiometer. *Acta Oto-Laryngologica, 35*, 411-422.

Bland. J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307-310.

Burns, W., & Hinchcliffe, R. (1957). Comparison of the audiometry threshold as measured by individual pure tone and Bekesy audiometry. *Journal of Acoustic Soc. Am., 29*, 1274-1277.

Bryman, B., & Bell, E. (2007). *Business research methods*. New York: Oxford University Press.

Carhart, R., & Jerger, J. F. (1959). Preferred method for determination of pure tone thresholds. *Journal of Speech Hearing Disorders, 24*, 330-345.

Cooper, H. (2010). *Research Synthesis and Meta-analysis: A Step by Step Approach* (4th edition). Sage Publications, Inc.

Corso, J. F. (1956). Effects of testing methods on hearing thresholds. *AMA Archs Otolar, 63,* 78-91.

Delany, M. E., Whittle, L. S., & Knox, E. C. (1966). A note on the use of self-recording audiometry with children. *Journal of Laryngology and Otology, 80*, 1135-1143.

Erlandsson, B., Hakanson, H., Ivarsson, A., & Nilsson, P. (1979a). Comparison of the hearing threshold measured by manual pure-tone and by self-recording (Békésy) audiometry. *Audiology, 18,* 414-429.

© University of Pretoria

Erlandsson, B., Hakanson, H., Ivarsson, A., & Nilsson, P. (1979b). Comparison of the hearing threshold measured by pure-tone audiometry and by Bekesy sweep audiometry. *Acta Oto-Laryngologica, 86,* 54-55.

Fagan, J. J., & Jacobs, M. (2009). Survey of ENT services in Africa: Need for a comprehensive intervention. Global Health Action, DOI: 103402/gha.v2i0.1932.

Fausti, S. A., Frey, R. H., Henry, J. A., Knutsen, J. L., & Olson, D. J. (1990). Reliability and validity of high-frequency (8-20 kHz) thresholds obtained on a computer-based audiometer as compared to a documented laboratory system. *Journal of the American Academy of Audiology, 1*, 162-170.

Formby, C., Sherlock, L. P., & Green, D. M. (1996). Evaluation of a maximum likelihood procedure for measuring pure-tone thresholds under computer control. *Journal Am Acad Audiol, 7,* 125-9.

Frampton, M. C., & Counter, R. T. (1989). A comparison of self-recording audiometry in naval establishments and clinical audiometry in a hospital setting. *Journal of the Royal Naval Medical Service, 75,* 99-104.

Frank, T., & Ragland, A. E. (1987) Repeatability of high-frequency bone conduction thresholds. *Ear Hear, 8*, 343-6.

Gosztonyi, R. E., Vassailo, L. A., & Sataloff, J. (1971). Audiometric Reliability in Industry. *Arch Enviromental Health, 22*, 113-118.

Goulios, H., & Patuzzi, R. B. (2008). Audiology education and practice from an international perspective. *International Journal of Audiology, 47*, 647–664.

Harris, D. A. (1979). Microprocessor, self-recording and manual audiometry. *Journal of Auditory Research, 19*, 159-166.

Hartly, V. V., & Siengenthalar, B. M. (1964). Relationships between Bekesy fixed frequency and convential audiometry with children. *Journal of Audit. Res, 4,* 15-22.

© University of Pretoria

Ho, A. T. P., Hildreth, A. J., & Lindsey, L. (2009). Computer-assisted audiometry versus manual audiometry. *Otology and Neurotology, 30*, 876-883.

Hood, J. D. (1960). The principles and practice of bone conduction audiometry. *Laryngoscope, 70,* 1211-1228.

Hughson, W., & Westlake, H. D. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Trans. American Academy Optomology Oto-laryngology, Suppl pp,* 3-15.

Ishak, W. S., Zhao, F., Stephens, D., Culling, J., Bai, Z., & Meyer-Bisch, C. (2011). Test-retest reliability and validity of Audioscan and Békésy compared with pure tone audiometry. *Audiological Medicine, 9*, 40-46.

Jerlvall, L., Dryselius, H., & Arlinger, S. (1993). Comparison of manual and computer-controlled audiometry using identical procedures. *Scand Audiology, 12,* 209-213.

Jokinen, K. (1969). Presbyacusis. I. Comparison of manual and automatic thresholds. *Acta Oto-Laryngologica, 68*, 327-335.

Katz, J. (2002)*. Handbook of Clinical Audiology (*5th edition). Baltimore: Williams and Wilkins Co.

Kitchenham, B. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering (*Version 2.3). Durham: University of Durham.

Knight, J. J. (1965). Normal hearing threshold determined by manual and self-recording techniques. *Journal of the Acoustical Society of America, 39*, 1184-1185.

Krumm, M., Ribera, J., & Klich, R. (2007). Providing basic hearing tests using remote computing technology. *Journal of Telemedicine and Telecare, 13*, 406-410.

© University of Pretoria

Leedy, P. D., & Ormrod, J. E. (2001). *Practical Research: Planning and design.* (7th Ed.). USA: Pearson.

Luakli, E., & Fjermedal, O. (1990). Reproducibility of hearing threshold measurements. Supplementary data on bone-conduction and speech audiometry. *Scand Audiology,19,* 187-190.

Lutman, M.E., Cane, M. A., & Smith, P. A. (1989). Comparison of manual and computer-controlled self-recorded audiometric methods for serial monitoring of hearing. *British Journal of Audiology, 23*, 305-315.

Maiya, P. S., & Kacker, S. K. (1973). Comparison of threshold between pure tone and Bekesy audiometry. *Silent Wld, 8,* 16-20.

Margolis, R. H., Frisina, R., & Walton, J. P. (2011). AMTAS: Automated method for testing auditory sensitivity: II. Air conduction audiograms in children and adults. *International Journal of Audiology, 50*, 434-439.

Margolis, R. H., Glasberg, B. R., Creeke, S., & Moore, B.C. (2010). AMTAS: Automated method for testing auditory sensitivity: Validation studies. *International Journal of Audiology, 49*, 185-194.

Margolis, R. H., & Moore, B. C. J. (2011). AMTAS: Automated method for testing auditory sensitivity: III. Sensorineural hearing loss and air-bone gaps. *International Journal of Audiology, 50*, 440-447.

Margolis, R. H., & Morgan, D. E. (2008). Automated Pure-tone Audiometry: An analysis of capacity, need and benefit. *American Journal of Audiology, 17*, 109-113.

© University of Pretoria

Margolis, R. H., Saly, G. L., Le, C., & Laurence, J. (2007). Qualind: A method for assessing the accuracy of automated tests. *Journal of the American Academy of Audiology, 18*, 78-89.

McPherson, B., Law, M. M. S., & Wong, M. S. M. (2010). Hearing screening for school children: Comparison of low-cost, computer-based and conventional audiometry. *Child: Care, Health and Development, 36*, 323-331.

Meyer-Bisch, C. (1996). Audioscan: a high definition Audiometry Technique based on constant-level Frequency sweeps- A new method with hearing indicators. *Audiology, 35,* 63-72.

Nachtegaal, J., Smit J. H., Smits C., Bezemer P. D., & Van Beek J. H. M. (2009a). The association between hearing status and psychosocial health before the age of 70 years: Results from an internet-based national survey on hearing. *Ear Hear*, 30, 302 – 312.

Picard, M., Ilecki, H. J., & Baxter, J. D. (1993). Clinical use of BOBCAT: Testing reliability and validity of computerized pure-tone audiometry with noise exposed workers, children and the aged*. Audiology, 32*, 55-67.

Raza, S. N. (2008). Computer audiogram. *Journal of the College of Physicians and Surgeons Pakistan, 18*, 463-464.

Rintelmann, W. F. (1973). Manual and automatic audiometry - a comparison. *Nat safety news , 108,* 95-102.

Robinson, D. W., & Whittle, L. S. (1973). A comparison of self-recording and manual audiometry: Some systematic effects shown by unpractised subjects. *Journal of Sound and Vibration, 26*, 41-62.

Sakabe, N., Hirai, Y., & Itami, E. (1978). Modification and application of the computerized automatic audiometer. *Scandinavian Audiology, 7*, 105-109.

© University of Pretoria

Sala, O., & Babighian, G. (1973). Automatic versus standard audiometry. *Audiology, 12,* 21-27.

Schmuziger, N., Probst, R., & Smurzynski, J.(2004). Test-retest reliability of pure-tone thresholds from 0.5 to 16kHz using Sennheiser HDA 200 and Etymotic Research ER-2 earphones. *Ear and Hearing, 25*, 127-132.

Smith-Olinde, L., Nicholson, N., Chivers. C., Highley. P., & Williams, D.K. (2006). Test-retest Reliability of In Situ unaided Thresholds in Adults. *American Journal of Audiology, 15*, 75-80.

Sparks, D. W. (1972). The feasibility of computerized audiometry. *The Journal of Auditory Research, 12,* 62-66.

Stuart, A., Stenstrom, R., Tompkins, C., & Vandenhoff, S. (1991). Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults. *Audiology, 30*, 82-90.

Swanepoel, D., & Biago, L. (2011). Validity of Diagnostic Computer-Based Air and Forehead Bone Conduction Audiometry. *Journal of Occupational and Environmental Hygiene, 8*, 210-214

Swanepoel, D. W., Clark, J. L., Koekemoer, D., Hall, J. W. I., Krum, M.,& Ferrari, D. V. (2010). Telehealth in audiology: The need and potential to reach underserved communities. *International Journal of Audiology*, *49*, 195-202. doi: 10.3109/14992020903470783.

Swanepoel, D., & Hall, J.W. III. (2010). A systematic review of Telehealth applications in Audiology. *Telemedicine Journal and E-health, 16*, 181-200.

Swanepoel, D., Mngemane, S., Molemong, S., Mkwanaze, H., & Tutshini, S. (2010). Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. *Telemedicine journal and e-health: the official journal of the American Telemedicine Association, 16,* 557-563.

© University of Pretoria

Swanepoel, D., Olusanya, B. O., & Mars, M. (2010). Tele-audiology in sub-Saharan Africa. *Journal of Telemedicine and Telecare, 16*, 53-56.

White, A., & Schmidt, K. (2005). Systematic literature reviews. *Complementary therapies in Medicine, 13*, 54-60.

Wood, T. J., Wittich, W. W., & Mahaffey, R. B. (1973). Computerized pure tone audiometric procedures. *Journal of Speech and Hearing Research, 16*, 676-684.

World Health Organization. (2010). Deafness and hearing impairment. Geneva: World Health Organisation.

© University of Pretoria

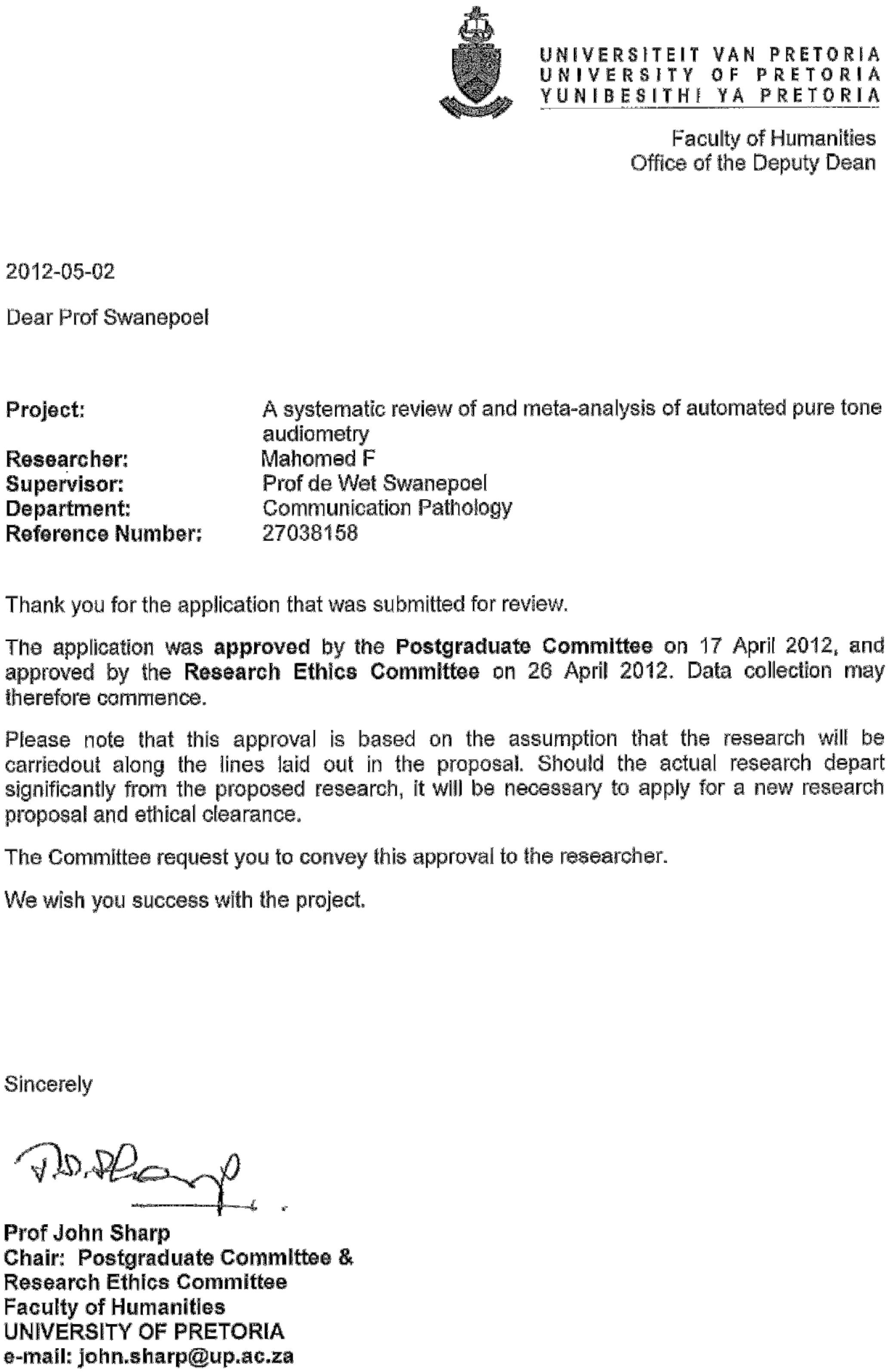
6. APPENDICES

© University of Pretoria

Appendix A.

Ethical clearance form

© University of Pretoria



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Office of the Deputy Dean

2012-05-02

Dear Prof Swanepoel

Project:	A systematic review of and meta-analysis of automated pure tone audiometry
Researcher:	Mahomed F
Supervisor:	Prof de Wet Swanepoel
Department:	Communication Pathology
Reference Number:	27038158

Thank you for the application that was submitted for review.

The application was **approved** by the **Postgraduate Committee** on 17 April 2012, and approved by the **Research Ethics Committee** on 26 April 2012. Data collection may therefore commence.

Please note that this approval is based on the assumption that the research will be carriedout along the lines laid out in the proposal. Should the actual research depart significantly from the proposed research, it will be necessary to apply for a new research proposal and ethical clearance.

The Committee request you to convey this approval to the researcher.

We wish you success with the project.

Sincerely

Prof John Sharp
Chair: Postgraduate Committee &
Research Ethics Committee
Faculty of Humanities
UNIVERSITY OF PRETORIA
e-mail: john.sharp@up.ac.za

© University of Pretoria

Appendix B.

Summary of reports included in review

(Arranged according to year of publication)

© University of Pretoria

Author	Year	Subject description	Test parameters	Automated audiometry threshold seeking method	Research findings		Conclusion
					Accuracy	Test-retest	
Corso	1956	105 subjects (210 ears), 17-25 years old. Normal hearing adults	Diagnostic AC audiometry. Frequencies: .25, .5, 1, 1.5, 2, 3, 4, & 8 kHz). Transducers: Auto- oscillator type 1011 manual- oscillator type 1304-A Audiometer: Manual- Bekesy type audiometer, Reager Model, Automated- ADC audiometer, Model 50-E2	Method of Adjustment- Békésy fixed frequency. Frequency range of 2- 8 kHz, starting at 40 dB. Testing time: 10min per ear was used with 0.5 dB rate per second. Thresholds obtained by the intersection of the midpoint curves and specific frequency lines.	- Average absolute thresholds and standard deviations -Test of significance (t-ration). -Difference in variability (F-ratio). -Pearson product-moment correlation coefficient.	-	Manual testing obtained thresholds that were lower than for automated testing (midpoint Békésy testing). Less variability in thresholds was noted between .25 and 2 kHz when manual testing was utilized. A low statistically significant positive correlation was noted at given frequencies between manual and automated audiometry.
Burns & Hichcliffe	1957	20 subjects (40 ears), 20-58 years of age. Hearing status not indicated	Diagnostic AC testing. Frequencies: .5, 1, 2, 3, 4, 6 kHz. Transducer: Standard Telephones Model 4026	Method of Adjustment - Békésy sweep frequency. Frequency range of .5-6 kHz was swept with a continuous tone, in 7 min 55 sec, paper speed of 1cm/min. Rate of change of intensity, increasing and decreasing, approximately 2 dB/sec. Thresholds obtained by the intersection of the midpoint curves and specific frequency lines.	- Average difference and standard deviation - t-Test values	- Average difference and standard deviations - Product moment correlation coefficients. -t-Test	Overall, manual and automated (Békésy) threshold audiometry gives essentially similar results. A significant difference was noted at 1000Hz, where Bekesy testing yielded a lower threshold of approximately 3 dB. Reliability was satisfactory at all frequencies utilizing both audiometric testing methods, besides at 500 Hz where the second automated test yielded a lowering of thresholds of 1-2 dB.
Hartely & Siengenthalar.	1964	30 subjects (60 ears) 13 children:	Diagnostic AC Testing. Frequencies: .25, 1,	Method of Adjustment- Békésy fixed frequency. 1 min fixed frequency	- Average thresholds	-	Better standard of acuity for manual compared to automated threshold audiometry were

© University of Pretoria

		4 - 5 years old; 17 children: 8-10 years old. Normal hearing children.	4 kHz. Audiometer: Manual- Audiovox Model 7-B, automated- Granson-Stadler Model E-800,	tracing (timed to begin after 3 reversals on the tracing) were obtained. Thresholds read using the mean mid-point between peaks and valleys.	- Average difference and Standard deviations -t-Test -Within subject variability – t-Test		obtained. The difference was greater for younger children than older children. Within subject variability for automated threshold testing was higher than manual testing. Significant difference of variability at .25 kHz for the older group and at 4 kHz for the younger group.
Delany *et al.*	1966	66 ears, 17-29 years old. Hearing status not indicated.	Diagnostic AC testing. Frequencies: .5, 1, 2, 3, 4, 6 kHz. Transducer: 4026A earphones Audiometer: Automated- mobile audiometric laboratory, manual- not indicated	Method of Adjustment - Békésy fixed frequency. Frequencies tested at kHz/sec. Tone burst presentation rate: 2 tones/sec.	- Average difference	-	Automated threshold audiometry gives results substantially in accord with manual audiometry. The differences over most frequencies are small, but automated threshold audiometry gives lower threshold levels.
Knight	1965	66 ears. Normal hearing subjects.	Diagnostic AC testing. Frequencies: .5, 1, 2, 3, 4, 6 kHz. Audiometer: Manual and automated- Grason-Stadler model E 800	Method of Adjustment. Attenuator speed: 5 dB/sec, tone pulsed 2/sec.	-Average difference and standard deviation	-	Manual and automated audiometry is equivalent, as they yield threshold levels on average that are within 1 dB.
Jokinen	1969	**4 groups*:*** 1) 19 subjects (30 ears), 19-24 years old, inexperienced, normal hearing subjects. 2)15 subjects	Diagnostic AC testing. Frequencies: .125 .25, .5, 1, 2, 3, 4, 6, 8 kHz. Audiometer: Manual- Madsen Model OB 60,	Method of Adjustment - Békésy fixed frequency. Tones presented for 30 sec at a frequency, first with 200 msec pulsed tones, secondly with a continuous tone. Tone pulse, rise and fall	-Average differences and standard deviations	-	Various differences were seen in the 4 groups. The normal hearing, inexperienced and experienced groups, obtained better results with automated testing (both continues and pulsed tones) than with manual testing.

© University of Pretoria

		(30 ears), 19-24 years old, experienced outpatients, normal hearing. 3) 9 subjects (17 ears), 52-73 years old, presbycusis with drop at 4000Hz indicating an acoustic trauma. 4) 22 patients (39 ears), 53-81 years old, subjects had presbycusis	Automated-Granson Stdler model E800	time of 25 msec, with on and off ratio of 1: 1. Intensity changes: 0.25dB steps, rate: 2.1 dB/sec.			The presbycusis group, with and without the acoustic trauma, indicated that manual and continues Békésy testing obtained the same results, however, pulsed Békésy testing obtained better thresholds than manual testing.
Gosztonyi et al.	1971	***Accuracy*** 19 subjects. ***Test-retest reliability*** 46 salaried employees and 25 hourly employees. All noise exposed adults.	Industrial screening AC testing. Frequencies: .5, 1, 2, 4, 8 kHz. Audiometer: Automated-self-recording audiometer, manual-standard clinical audiometer.	Method of Adjustment.	- Average thresholds	- Average difference	Manual testing produced better thresholds than automated testing , there was a difference of 10 dB between the two. Test- retest reliability for salaried employees indicated a difference no more than 10 dB. In this study it was investigated that the reason for the great difference between thresholds was as a result of subjects either being influenced to claim for HL or had compensation cases or had compensation legislations in progress.
Sparks	1972	15 subjects. Bi-modal population of mild or severe hearing loss	Diagnostic AC and BC testing, with masking. Frequencies: .25, .5, 1, 2, 4, 8 kHz. Transducers:	Method of limits. A computer program using Hughson-Westlake procedure for threshold seeking, masking programmed according to	-Average thresholds and standard deviations. -t-Test	-	It was apparent that if subjects were consistent in their response, automated testing could obtain thresholds similar to that of manual testing. The t-test: no significant

© University of Pretoria

		participants used.	AC- TDH-39 housed in a MX-41 AR cushion. BC- Radioear B-70A oscillator Audiometer: Manual and automated-Beltone 15-C	Hood (1960). Computer program provided instructions, which were followed by an assistant who was familiar with the use of Teletype system. If a response was elicited the assistant would type 1, no response the assistant would type 2. The computer would indicate next step.	conducted on mean values. -Product moment correlation coefficient.		difference between AC and BC values between two methods of testing. Correlation coefficients: high correlation between the two methods of testing.
Maiya, & Kacker.	1973	20 subjects, 15-30 years. Normal hearing subjects.	Diagnostic AC testing. Frequencies: .125, .25, .5, 1, 2, 4, 6, 8 kHz. Audiometer: Manual-Maico-MA-8, Automated-Grason-Stadler Company model E-800.	Method of Adjustment - Békésy sweep frequency. Rate: 1 octave/min, chart travel period of 6 2/3 min. Rate of change of intensity: 2.5dB/sec. Thresholds read using the mid-point mean value between ascending or descending tracing at the frequency level.	- Average thresholds	-	Automated and manual testing yielded similar thresholds, however automated testing seemed to be more sensitive than manual testing.
Robinson & Whittle	1973	***Accuracy:*** 64 subjects (128 ears), 26-73 years old. ***Test-retest reliability:*** 48 subjects (96 ears), 29-73 years old. Hearing status not indicated.	Diagnostic AC testing. Frequencies: .25, .5, 1, 2, 4, 6, 8 kHz. Transducers: TDH-39 earphones and MX-41-AR cushions. Audiometer: Manual and automated-Rudmose type ARJ-5	Method of Adjustment - Békésy fixed frequency. Pulsed tones with a repetition rate: 2 Hz, cycle consisting of a silent period of 185 ms and a tone pulse with 65 ms rise, fall times and a dwell of 185 ms at maximum amplitude, attenuator: 5dB/s. Thresholds read as the mid-point of the excursions, extraneous deviations being ignored.	- Average differences and standard deviations -Linear regression and correlation coefficients. - Estimation of asymptomatic data.	- Average differences and standard deviations of initial test - Average differences and standard deviations of second test	Automated threshold yield better results than manual testing, except at .25 kHz where no diff was noted. Test-retest reliability: manual and automated testing yield lower thresholds when tested for the second time.

© University of Pretoria

Wood *et al.*	1973	20 subjects, 7-72 years old. Hearing status of subjects included: 1 normal hearing subject, 14 sensorineural, 4 conductive and 1 mixed hearing loss subject/s.	Diagnostic AC, BC testing with masking. Frequencies: .25, .5, 1, 2, 4, 8 kHz. Audiometer: Automated- Grason Stadler model 829E, manual- not indicated.	Method of limits. Functional generator controlled frequency of tonal signal. Rise and fall time: 30 sec, duration of the tone: 1500msec. ***Unmasked air and bone:*** Tones presented using an initial bracketing of 10 dB, then a bracketing of 5dB. ***Masking:*** AC Masking- 40dB gap between AC of test ear and BC of non-test ear. BC Masking- if AC of the test ear exceeded the midline BC by more than 10dB. Minimal effective masking (Martin 1976) was used / if patient did not respond to minimal masking than platue masking was administered.	- Average deviations	-	A high positive relationship between manual and automated testing for air and bone testing was noted. Automated testing reduces examiner bias and causes direct standardization of testing. Additionally, the use of computerized program will give the audiologist time for direct patient contact, counselling and aural rehabilitation.
Almqvist & Aursnen	1978	82 subjects (41 ears), 7-82 years. Hearing status not indicated.	Screening AC, Frequencies: .5, 1, 2, 3, 4, 6 Hz. Audiometer: Manual- not indicated, Automated- minicomputer, type PDP-8.	Method of limits. Computer program utilized principles based on manual audiometry.	-Standard deviation	-	Automated audiometry appeared to be a fast and a reliable method for screening audiometry. A total standard deviation of 4.8 dB was noted between manual and automated audiometry, standard deviation varied across frequencies and was the smallest in the speech frequencies.
Sakabe *et al.*	1978	***2 groups used:*** 1) 31 subjects (62 ears), 19- 22 years old. Normal hearing	Diagnostic AC testing. Frequencies: .125, .25, .5, 1, 2, 4, 6, 8 kHz.	Method of limits. Automatically interrupted tone, on-off time: 2sec, rise- fall time: 25ms. Tone presented at 30dB, if	- Error analysis	-	Automated audiometry has sufficient accuracy for practical use. Automated audiometry coincides with manual audiometry within 10

© University of Pretoria

		subjects. 2) 124 subjects (248 ears). Hearing status not indicated.		not heard, raised to 60dB, if heard lowered again to 30dB and increased by 5dB till heard again. The tone is lowered to 30dB again and raised in 5dB steps till a response is elicited. Once a response is obtained a comparison between the 2 'thresholds' are made. The smaller value is the threshold obtained at that frequency.			dB. Additionally it would take 5-15min to conduct.
Erlandsson *et al.*	1979	***Accuracy :*** 115 subjects (230 ears), 25 to 63 years. ***Test-retest reliability:*** 10 subjects (20 ears). All subjects were noise exposed shipyard workers.	Diagnostic AC. Frequencies: .25, .5, 1, 1.5, 2, 3, 4, 6, 8 kHz. Transducers: Manual- TDH-39M with MX-41/AR cushions. Automated- TDH-49P with MX-41/AR cushions. Audiometer: Manual- Madsen OB60, automated- Type Delmar 120.	Method of adjustment-Békésy sweep frequency. Attenuation rate: 2.5 dB/s, pulsed tone-presentation; sweep time from .25 -10 kHz was 400s.	-Regression equations and α and β coefficients. - Estimated standard deviations	-standard deviations	Automated audiometry yields a lower and more reliable hearing threshold than manual audiometry. Manual audiometry SD are about twice as much for automated testing. Test-retest reliability of automated audiometry indicated that the standard deviations between the 5 successive tests had their lowest values for 1 kHz, increasing slowly towards lower and higher frequencies.
Erlandsson *et al.*	1979	***Accuracy :*** 115 subjects (230 ears), 25 to 63 years. ***Test-retest reliability:*** 10 subjects (20 ears).	Diagnostic AC. Frequencies: .5, 1, 1.5, 2, 3, 4, 6, 8 kHz. Audiometer: Manual- Madsen OB60, automated- Type Delmar 120.	Method of adjustment-Békésy sweep frequency. Attenuation rate: 2.5 dB/s with a pulsed tone-presentation, sweep time from .25-1 kHz was 400s.	- Regression equation - Estimated standard deviations	- Average thresholds and standard deviations	Automated audiometry yields a lower and more reliable hearing threshold than manual audiometry. Test-retest reliability of automated audiometry indicated that the standard deviations between the 5 successive tests had their lowest values for 1 kHz,

© University of Pretoria

		All subjects were noise exposed shipyard workers.					increasing slowly towards lower and higher frequencies.
Harris	1979	12 subjects (24 ears), 20 - 26 years old. Hearing status not indicated.	Diagnostic AC. Frequencies: .5, 1, 2, 3, 4, 6, 8 kHz. Audiometer: Manual- Tracor Model RA-115, automated- Self-recording- Tracor Model ARJ-4C, Microprocessor- Tracor Moder RA-40 ** Two automated methods compared to manual testing.	Method of adjustment- Békésy fixed frequency. Tone pulse rate: 2.5pulses/sec was used; tones were presented for 30sec at each frequency. Attenuation rate of 5dB/sec in 0.25dB steps. Thresholds read as the mid-point of the excursions at each frequency. Method of limits. An 800msec tone presented at random intervals of 1,2, sec. The Hughston-westlake method was utilized by the computer program.	- Average threshold and standard deviation - Average differences	-	Automated audiometry, utilizing the method of limits, indicated results that agree more with manual than automated audiometry utilizing the method of adjustment. At all frequencies, automated audiometry utilizing the method of adjustment showed lower thresholds than the other 2 tests. Automated audiometry utilizing the method of limits showed higher thresholds for all frequencies except 4 KHz, over manual audiometry. The two automated audiometry tests differed significantly at the 0.01 level in all frequencies. Time differences between each test were less than a minute.
Frampton & Counter	1989	42 subjects (84ears). All subjects were noise exposed adults.	Diagnostic AC testing. Frequencies: .5, 1, 2, 3, 4, 6, 8 kHz. Audiometer: Manual- Grason Stadler GSI 10, automated- Grason Stadler 1703 B	Method of Adjustment - Békésy sweep frequency. 7 frequency sweep with a pulsed tone mode.	- Average differences	-	Automated audiometry produced lower thresholds than manual testing. Automated audiometry is reliable and sensitive in the 'real world' setting. It allows large numbers of audiograms to be collected quickly by medical assistants with no training.

© University of Pretoria

Lutman *et al.*	1989	120 subjects (240 ears), 40 – 65 years old. Hearing status not indicated. Longitudinal study, subjects retest 2-3 years later.	Diagnostic AC thresholds. Frequencies: .5, 1, 2, 3, 4 kHz. Transducers: Manual- TDH-39P with MX 41/AR cushions Automatic- TDH-49P with MX -41/AR cushions	Method of adjustment- Békésy fixed frequency. Stimulus tone pulsed at a rate: 2.5pulses/sec, with duration of 200ms (3dB down points). The tracking procedure : 2dB step occurring every 2 pulses. Tracking at each frequency lasted 40sec, 50 levels were visited for each frequency.	- Average thresholds and standard deviations - Ranges of thresholds - Average difference	- Average differences and standard deviations - Standard of variance	Automated audiometry produced better results than manual audiometry. Overall automated audiometry was 4.4 dB better than manual audiometry; the difference was lower at .5 kHz and increased as the frequency increased. Test-retest reliability- manual audiometry indicated a worsening of hearing at .5,1, 2 kHz and an improvement at 4 kHz. Automated audiometry produced correlation coefficients which were statistically significant, however it suggests the shift is due to random measurement error rather than actual shifts in the threshold.
Fausti *et al.*	1990	20 subjects (40 ears), 18-25 years old. Normal hearing adults.	Diagnostic AC testing. Frequencies: .25, 0.5, 1, 2, 4, 8 kHz. Audiometer: Manual- GS1701, Automated-V320	Method of limits. V 320 Audiometer used, tones presented: 50% duty cycle, duration: 250 ms , rise-fall time: 25-50ms. Modified Hughson Westlake Ascending-descending audiometric test technique .	- Two-way analysis of variance with repeated measures on frequency and system s - Sheffé's to determine statistical significance.	- Average absolute differences	No significant difference was noted between automated and manual testing over all test frequencies. Test-retest reliability: indicated no significant difference between the two tests conducted.
Picard *et al.*	1993	***3 groups used:*** 1) 420 subjects (840 ears), 18-64 years old. Noise exposed workers. 2) 36 elderly	Diagnostic AC and BC testing with masking. Frequencies: AC- .5, 1, 2, 3, 4, 6 kHz. BC- .5 ,1, 2, 4 kHz.	Method of limits- BOBCAT. Tone duration of 700ms, 2s time interval. The computer program made use of the ascending-descending method (ISO 6189).	- Reliability coefficients using Hoyt's solution. - Average thresholds	-	Manual and automated procedures produce similar results, regardless of subject age, degree of hearing loss or nature of hearing loss. Mean thresholds across the populations comparable between automated

		subjects (72 ears), 65-80 years old. Hearing status not indicated. 3) 12 subjects (24 ears), 7.5-12 years old. Normal hearing children.	Audiometer: Automated- MADSEN, Model OB 822, manual not indicated.	***Masking:*** Hood technique of masking used. AC Masking- 40dB gap between AC of test ear and BC of non-test ear. BC Masking- AC of the test ear exceeded the midline BC by more than 10dB.	and standard deviation - Dispersion relationships		and manual testing. Automated testing with the child population did not reveal consistent results when compared to manual audiometry, especially at 2 and 6 kHz. Automated testing takes longer to determine thresholds than manual testing (automated- 42 sec, manual- 34 sec). It was noted as population changed to 'difficult to test' patients (children) manual testing started to take more time. It was also noted that examiner takes shortcuts to obtain results but automated testing maintains rigid adherence to full procedure.
Fromby *et al.*	1996	***Accuracy:*** 101 subjects (202 ears), mean age of 43 years. Noise exposed workers. ***Test-retest reliability:*** 20 subjects (39 ears), Mean age of 43 years. Noise exposed workers.	Diagnostic AC testing. Frequencies: .25. .5, 1, 2, 3, 4, 6, 8 kHz Transducer: Telephonics TDH-39. Audiometer: Manual- Madsen, model OB822, automated- digital-to-analog converter (DAC) (TDT, model Quikki QDA1).	Method of limits- Maximum likelihood method was used (ML). Threshold for each frequency was measured in 15-trial block to yield 60% correct detection. On a trial, a 200msec pure-tone signal presented in a visually cued 200msec observation interval. Signals: 10-msec rise-fall times as part of the nominal durations. Subjects had 1000 msec to make a "yes-only" response which attenuated the signal level. If the subject did not respond during the 1000-msec response period, the	- Average threshold - Standard error bars	- Average threshold - Standard error bars	Automated testing and manual testing yielded similar results. Threshold differences between the two methods were not statistically significant at any test frequency except .25 kHz, automated threshold was higher, but was within 3 dB of the threshold obtained manually. Test- retest reliability for automated testing: no significant test-retest differences at any test frequency. Additionally, manual testing took less time than automated testing (manual- 3 min 46 sec, auto-6 min 43 sec).

© University of Pretoria

				computer assumed a "no" response for the trial, and the signal level was increased according to the ML algorithm.			
Margoliset *al.*	2007	***3 groups:*** 1) 120 subjects, 16-93 years old. Hearing status varied. 2) 8 subjects, 64- 85 years old. Varying degrees of hearing loss. 3) 6 subjects, 13- 86 years old. Varying degrees of hearing loss.	Diagnostic AC, BC and masking. Frequencies: not indicated. Transducers varied for different groups tested. ***Group 1 and 2:*** Manual- TDH-50, automated- prototype, non-occluding circumaural earphones ***Group 3:*** Manual- TDH-50 (not test ear occluded during BC testing), automated- insert earphones ER3A (both ears occluded during BC testing)	Method of limits- AMTAS. Tonal stimuli presented in a temporal observation interval that is visually marked for the listener, following the observation interval, the listener responds YES or NO by touching ‘buttons’ on a touchscreen monitor. The signal level is changed in an adaptive fashion to find the threshold of audibility. A threshold is obtained using a bracketing procedure. Masking noise presented to the non-test ear at levels that are selected to maximize the likelihood that neither under-masking nor over-masking will occur.	-Average absolute differences (QAave) - Regression coefficients - QUALIND -Correlation coefficients	-	The aim of this study was to develop a quality assessment method (QUALIND) based on a comparison of audiograms obtained utilizing automated (AMTAS) and manual testing. A predictive equation was derived from a multiple regression of a set of quantitative quality indicators on a measure of test accuracy, defined as the average absolute difference between automated and manually tested thresholds. For a large subject sample (n=120), a strong relationship was found between predicted and measured accuracy. The predictive equation was cross validated against two independent data sets. The results suggest that the predictions retain their accuracy for independent data sets if similar subjects and methods are employed, and that new predictive equations may be required for significant variations in test methodology. The method may be useful for automated test procedures when skilled professionals are not available to provide quality assurance.

© University of Pretoria

Ho *et al.*	2009	***3 groups used:*** 1) 16 subjects (32 ears), 20- 80 years old. 2) 16 subjects (32 ears), 23-80 years old. 3)16 subjects (32 ears), 23- 81 years old. Hearing status of all 3 groups unknown.	Diagnostic AC and BC testing with masking. Frequencies: AC- .25, .5, 1, 2, 3, 4, 6, 8 kHz. BC- .5,1, 2, 4 kHz Transducer: EAR 5A. Audiometer: Manual- not indicated, Automated- Otogram.	Method of limits- Otogram. Assesses AC and BC thresholds, administers masking when appropriate. Uses touch-screen technology programmed according to the Hughson-Westlake algorithm.	- Average Differences and standard deviations. - Levels of agreement were analysed and expressed by weighted □ coefficients, using SPSS version 15 and StatXact version 8.0.	- Average Differences and standard deviations. - Levels of agreement were analysed and expressed by weighted □ coefficients, using SPSS version 15 and StatXact version 8.0.	AC and BC results when tested with automated and manual testing produced similar results. AC thresholds when tested using automated and manual testing indicated 94% of automated thresholds that fell within 10 dB of those obtained manually and indicated 10 paired thresholds that fell within 15 dB of manual testing. BC unmasked thresholds showed that 93% of automated thresholds fell within 10 dB of each other and 96% fell within 15 dB of each other. BC masked thresholds between the 2 tests showed a lower level of agreement but still a good level of agreement. Test-retest reliability indicated good intrarater agreement between the automated and manual testing conducted.
McPherson *et al.*	2010	80 subjects (160 ears), 7-8 years old.	Screening AC tested. Frequencies: .5, 1, 2, 3, 4 kHz. Transducers: Manual- Circumaural ME-70 enclosures over TDH-39 supra-aural earphones. Automated- Circumaural headphone Ovann OV880V. Audiometer:	Methods of adjustment. Békésy fixed frequency. Continues tones of 1 sec were presented in left ear at .5 kHz at 40 dB, and were raised or lowered in 3dB steps depending on response. Thereafter 1-4 kHz tested.	-X^2-test -Sensitivity or specificity analysis - Individual test results for each ear was compared using kappa values of agreement.	-	Automated screening procedure produced higher referral rate than manual screening (56% versus 13%). However, when .5 kHz was excluded from the data the referral rate between the two methods indicated no significant difference. The reason for .5 kHz producing errors could be as a result of ambient environmental noise and that automated audiometry started at .5 kHz and subjects were unfamiliar to test

© University of Pretoria

			Manual- Madsen Micromate, automated- IBM ThinkPad laptop PC, model T22.				procedures.
Margolis *et al.*	2010	***Accuracy:*** 30 subjects (60 ears). Hearing status: 5 normal hearing subjects, 25 hearing loss subjects. ***Test-retest reliability:*** 18 subjects (36 ears). Hearing status: 3 normal hearing subjects, 15 sensorineural hearing loss subjects.	Diagnostic AC, BC and masking. Frequencies: AC- .25, .5, 1, 2, 3, 4, 6, 8 kHz. BC- .5, 1, 2, 4 kHz Transducer: AC- Sennheiser HDA200 BC manual- Radioear B71(mastoid placement) BC automated- B71 vibrator (forehead placement). Audiometer: Manual and automated- Madsen Conera.	Method of limits- AMTAS (see Margolis *et al*, 2007).	- Average differences -Average Absolute differences -Confidence intervals	-	The differences between automated and manual testing were compared to differences obtained when the same subjects are tested manually by two audiologists. AC thresholds obtained by manual and automated testing indicated similar differences that were obtained when the same patients were tested manually by two audiologists. BC thresholds obtained with automated testing were lower than thresholds obtained with manual testing. The difference could be due to the placement of the bone conductor.
Swanepoel *et al.*	2010	***2 groups used:*** 1) 30 subjects (60 ears), 18- 31 years old. Normal hearing adults. 2) 8 subjects (16 ears), average age of 55 years	Diagnostic AC and masking. Frequencies: .125, .25, .5, 1, 2, 4, 8 kHz. Audiometer: Manual and automated- KUDUwave 5000.	Method of limits. Modified Hughson-Westlake method. Software presented a tone for 1.25s, subjects had to respond within 1.5 s before the next tone was presented. Threshold was accepted if	- Absolute average differences and standard deviations - Two sided paired *t*-test - Pearson	- Absolute average differences and standard deviations - Two sided paired *t*-test	Thresholds determined by manual and automated testing were within 5 dB of each other, indicating no significant difference between the two test procedures, in both the hearing and hearing loss group. Test-retest reliability of automated testing indicated reliability equivalent to that of manual

© University of Pretoria

		old. Subjects had a sensorineural hearing loss ranging from mild to severe hearing loss.		there was a minimum of 3 responses. Software automatically determined if contralateral masking was necessary and applied when required in an adaptive manner.	correlation coefficients	- Pearson correlation coefficients	testing. Additionally, both manual and automated testing took more or less the same time to administer (manual- 7.2-7.7 min, automated- 7.2-7.4 min).
Ishak *et al.*	2011	***Accuracy:*** 13 subjects (13 ears), a8-60 years old. Normal hearing adults. ***Test-retest reliability:*** 21 subjects (21 ears), 18-60 years old. Normal hearing adults.	Diagnostic AC testing. Frequencies: .25, .5, .75, 1, 1.5, 2, 3, 4, 6, 8 kHz. Audiometer: Manual and automated- Essilor Audioscan system. ** Test-retest reliability was determined by testing subjects 4 times with each test producer.	Method of adjustment- Békésy sweep frequency and Audioscan. ***Békésy:*** Sweep rate: 15 s per octave, pulse rate: 2.5 pulses/s, attention rate: 2.5dB/s was used. Hearing thresholds determined by calculating averaged values of three consecutive audiometric data obtained around each octave or half-octave frequencies. These values were rounded to the nearest 5dB for the analysis. ***Audioscan:*** Sweep rate: 15sec/octave, tones swept 1- 8 kHz, back to 1 kHz and swept again from 1 kHz to .25 Hz. A straight line was produced when the subjects pressed the response button. The level was then increased by 5dB at frequencies to which the subjects did not respond.	- Repeated measures ANOVA - Contrasts analysis to compare mean thresholds.	- Thresholds from each test session were subtracted - Variance of hearing threshold (σ^2)	The results showed that the thresholds obtained with Békésy testing were significantly better than those obtained from the manual testing at most frequencies. Audioscan produces better thresholds than Békésy, showing no significant differences in hearing thresholds at frequencies from .5 kHz- 4 kHz. Hearing thresholds obtained from Audioscan were significantly poorer than manual testing at frequencies of .25, 6 and 8 kHz. This was probably due to the threshold seeking procedure, which does not allow the intensity level to go either higher or lower than the current screening intensity level. High test-retest reliability for manual and audioscan testing, however, Békésy testing indicated poor test-retest reliability.

© University of Pretoria

Margolis *et al.*	2011	***2 groups:*** 1) 68 subjects (136 ears), 4- 8 years old (1 group of 4-5 year olds and another group of 6-8 year olds). Normal hearing children. 2) 15 subjects , Adults. Hearing status: 11 normal hearing, 1 unilateral hearing loss, 3 mild-to-moderate bilateral hearing loss subjects.	Diagnostic AC testing. Frequencies: .5, 1, 2, 4, 8 kHz. Transducers: Automated- HDA 200 Manual- TDH-50. Audiometer: Manual and automated (children)- Benson CCA-100 Mini. Manua (adults)l- Grason Stadler, automated- Benson CCA. **Different transducers were only used in the adult population.	Method of limits- AMTAS was used for the adult group (see Margolis *et al*, 2007). KIDTAS was used for the child population. It differed from AMTAS, used a smiley and sad face and a visual reinforcement picture for a correct response. Additionally, QUALIND was used. QUALIND is a method for estimating accuracy by tracking variables that are known to predict agreement between automated and manual thresholds, and calculating the predicted average absolute difference with a formula derived from a regression analysis of the relationship between the quality indicators and the measured average absolute differences. The strength of the regression coefficient indicates the degree to which accuracy can be predicted by QUALIND.	- Average absolute average difference and standard deviation	-	The differences obtained between automated testing (AMTAS/KIDTAS) and manual testing produces thresholds with variability that is comparable to thresholds obtained using manual testing by two audiologists, only if QUALIND identifies and excludes 'poor' audiograms. No significant differences between manual and automated thresholds were noted when using different earphones in the adult subjects.
Margolis & Moore	2011	13 subjects (19 ears), 21- 65 years old. All subjects had a sensorineural	Diagnostic AC, BC and masking. Frequencies: .25, .5, 1, 2, 4, 8 kHz. Audiometer: Manual- Grason	Method of limits- AMTAS (see Margolis *et al*, 2007).	- Average thresholds -Average differences -Average	-	Automated testing produced thresholds similar to those obtained by manual testing results. Automated thresholds were higher than those obtained manual by 7 dB at .25, .5, 1, 2

© University of Pretoria

		hearing loss.	Stadler GSI 61, Automated- Madsen Aurical.		absolute differences -Analysis of variance (ANOVA)		kHz, with smaller differences at higher frequencies. According to Margolis et al (2010) results between manual and automated testing should be similar, thus it was concluded by this study that the difference noted between the two test results was due to the use of different earphones.

© University of Pretoria

Appendix C.

Summary of data included in meta-analysis (Test-retest reliability)

© University of Pretoria

	Author	Year	Number of ears	Statistical analysis	Frequencies (Hz)									
					125	250	500	1000	2000	3000	4000	6000	8000	All
MANUAL TESTING	AVEARAGE DIFFERENCES													
	Burns & Hichcliffe.	1957	40	Average differences	-	-	1.0	2.2	1.5	2.0	1.4	-1.7	-	1.0
				Standard deviation	-	-	4.9	4.2	4.7	4.7	4.7	7.6	-	5.1
	Lutman et al.	1989	460	Average differences	-	-	2.4	2.1	1.4	-	-0.5	-	-	1.3
				Standard deviation	-	-	6.9	4.8	5	-	7.1	-	-	6.1
	Ho et al.	2009	32	Average differences	-	-	-	-	-	-	-	-	-	1.8
				Standard deviation	-	-	-	-	-	-	-	-	-	6.6
	ABSOLUET AVERAGE DIFFERENCES													
	Fausti et al.	1990	20	Absolute Average difference	-	2.3	2	1.8	1.8	-	2.3	-	2.3	2.1
	Swanepoel et al.	2010	60	Absolute Average difference	4.8	3.8	3.3	3.7	3.0	-	3.0	-	3.3	3.6
				Standard deviations	5.0	3.7	3.7	3.4	3.6	-	3.5	-	4.3	3.9
AUTOMATED TESTING	AVERAGE DIFFERENCES													
	Burns & Hichcliffe.	1957	40	Average differences	-	-	1.0	2.0	1.0	2.1	1.2	1.7	-	1.5
				Standard deviation	-	-	6.4	5.2	3.8	6.2	6.4	10.4	-	6.4
	Lutman et al.	1989	460	Average differences	-	-	0.2	-1.3	-0.1	-	0.6	-	-	0.1
				Standard deviation	-	-	7.2	6.9	6.6	-	7.2	-	-	7.0
	Ho et al.	2009	32	Average differences	-	-	-	-	-	-	-	-	-	0.3
				Standard deviation	-	-	-	-	-	-	-	-	-	5.9
	ABSOLUTE AVERAGE DIFFERENCES													
	Fausti et al.	1990	20	Absolute Average difference	-	2.3	1.8	1.8	1.8	-	2.0	-	1.5	1.9
	Swanepoel et al.	2010	60	Absolute Average difference	4.9	3.8	3.2	2.8	2.8	-	2.4	-	2.2	3.2
				Standard deviations	4.8	3.5	3.6	3.2	4.1	-	3.0	-	3.2	3.8

© University of Pretoria

Appendix D

Summary of reports included in the Meta-analysis (Accuracy)

© University of Pretoria

	Author	Year	Number of ears	Statistical analysis	Frequencies									
					125	250	500	1000	2000	3000	4000	6000	8000	All
METHOD OF ADJUSTMENTS	AVERAGE DIFFERENCES													
	Burns & Hichcliffe	1957	40	Average differences	-	-	-1.1	3.2	1.3	1.6	1.2	-0.5	-	1.0
				Standard deviation	-	-	5.5	5.1	4.7	6.0	7.1	9.2	-	6.3
	Knight	1965	66	Average differences	-	-	-0.3	1.0	1.5	1.1	1.3	-0.1	-	0.8
				Standard deviation	-	-	4.2	4.9	4.9	4.9	3.8	5.3	-	4.7
	Delany et al.	1966	66	Average differences	-	-	1.2	-0.8	-0.9	-1.4	-1.5	-3.3	-	-1.1
	Jokinen	1969	30	Average differences	5.1	2.1	-0.6	-0.6	2.3	4.9	-2.5	-0.6	-2.7	0.8
				Standard deviation	7.0	6.0	5.0	5.6	7.0	4.4	5.4	6.6	6.4	5.9
			30	Average differences	-1.7	-3.1	-3.3	-2.8	-0.6	4.1	-4.4	-4.5	-5.5	-2.4
				Standard deviation	8.1	6.1	4.2	4.9	5.2	5.6	5.3	6.4	7.9	6.0
			17	Average differences	-8.4	-7.7	-5.4	-7.8	-4.4	2.6	-10	-5.6	-3.9	-5.6
				Standard deviation	6.3	6.4	6.4	3.8	6.2	7.4	5.6	6.7	6.4	6.1
			39	Average differences	-5.2	-7.0	-5.6	-6.4	-4.1	-1.0	-12.6	-8.1	-12.3	-6.9
				Standard deviation	9.9	9.5	6.7	7.0	7.6	7.4	7.0	9.0	11.5	8.4
			30	Average differences	4.3	0.3	-2.0	-0.4	1.2	5.6	-2.0	-0.7	-3.6	0.3
				Standard deviation	7.7	7.2	5.5	5.7	6.8	5.2	7.0	8.1	8.5	6.9
			30	Average differences	-4.0	-5.7	-4.1	-2.8	-1.5	3.2	-5.1	-6.1	-4.8	-3.4
				Standard deviation	8.9	6.1	4.0	5.2	5.1	6.8	5.9	7.4	8.2	6.4
			17	Average differences	-6.4	-6.0	-2.1	-3.0	0.9	8.5	-3.6	1.9	2.0	-0.9
				Standard deviation	6.7	6.7	8.4	5.0	7.3	6.8	4.8	7.2	6.1	6.6
			39	Average differences	-3.1	-4.7	-0.2	-1.7	-0.5	-4.2	-4.9	-2.8	-10.0	-3.6
				Standard deviation	11.4	10.0	7.9	7.7	6.6	8.1	8.2	10.1	14.6	9.4
	Robinson & Whittle	1973	128	Average differences	-	0.4	2.9	1.5	2.8	-	2.7	4.2	2.1	2.4
				Standard deviation	-	5.9	4.4	4.1	4.3	-	5.3	8.2	8.5	5.8
	Harris	1979	24	Average differences	-	-	-2.1	-4.0	-5.6	-4.0	-9.0	-1.0	-2.9	-4.1
	Lutman et al.	1989	240	Average differences	-	-	3.0	2.8	6.4	-	5.3	-	-	4.4
				Standard deviation	-	-	5.8	5.6	5.2	-	6.1	-	-	5.8

© University of Pretoria

METHOD OF LIMITS	**AVERAGE DIFFERENCES**													
	Harris	1979	24	Average differences	-	-	-3.5	-2.3	-1.3	-2.9	3.8	-4.4	-0.2	-1.5
				Standard deviation	-	-	6.4	5.2	3.8	6.2	6.4	10.4	-	6.4
	Ho et al.	2009	32	Average differences	-	-	-	-	-	-	-	-	-	0.76
				Standard deviation	-	-	-	-	-	-	-	-	-	5.7
	Margolis et al.	2010	60	Average differences	-	-0.4	0.4	1.5	1.4	-	0.1	-	-2.3	0.1
				Standard deviation	-	5.1	4.4	5.3	5.8	-	4.9	-	7.0	5.4
	ABSOLUTE AVERAGE DIFFERENCES													
	Sparks	1972	15	Absolute Average differences	-	-	-	-	-	-	-	-	-	4.5
	Swanepoel et al.	2010	60	Absolute Average differences	4.8	3.8	3.8	3.7	3.2	-	2.9	-	2.8	3.6
				Standard deviation	4.1	3.4	4.5	3.7	3.3	-	3.5	-	4.5	3.9
			60	Absolute Average differences	4.2	3.8	3.6	3.8	3.3	-	2.2	-	2.3	3.3
				Standard deviation	4.2	3.5	4.5	3.1	4.0	-	3.0	-	3.6	3.8
			16	Absolute Average differences	2.3	3.3	2.2	2.2	2.2	-	2.8	-	1.4	2.4
				Standard deviation	3.2	2.4	2.6	2.6	2.6	-	3.1	-	3.1	2.8
	Margolis et al.	2010	60	Absolute Average differences	-	3.2	3.0	3.3	4.0	-	3.7	-	4.5	3.6
				Standard deviation	-	4.0	3.2	4.4	4.4	-	3.2	-	5.8	4.2
	Margolis et al.	2011	15	Absolute Average differences	-	-	-	-	-	-	-	-	-	3.9
				Standard deviation	-	-	-	-	-	-	-	-	-	1.7

© University of Pretoria