Test-retest reliability and validity of distortion product oto-acoustic emissions and transient evoked oto-acoustic emissions in normal hearing adults

by

Carina Grové

Research study submitted in fulfilment of the requirements for the degree

M.Communication Pathology (Audiology)

in the

Department of Speech-Language Pathology and Audiology

Faculty of Humanities

UNIVERSITY OF PRETORIA

December 2016

ACKNOWLEDGEMENTS

The author would like to acknowledge the contribution of the following individuals, without whom this project would never have been completed:

- Dr. Maggi Soer for her unwavering support and loyalty throughout the whole project, and for her expert knowledge and guidance delivered not only on the field of audiology.
- My parents and sisters for their unconditional love, support and prayers.
- Konstant Petzer, for your sacrifices and belief in my abilities.
- My Saviour, Jesus Christ.

Phil 4:13 “I can do all things through Christ who strengthens me”

Isa 41:10 “So do not fear, for I am with you; do not be dismayed, for I am your God. I will strengthen you and help you; I will uphold you with my righteous right hand.”

TABLE OF CONTENTS

List of Tables .... **vii**

List of Figures .... **ix**

List of Abbreviations .... **x**

Abstract .... **xi**

1. Chapter One: Introduction and orientation **1**

1.1 Introduction 1

1.2 Background and development of OAEs 1

1.3 Problem statement and rationale 7

1.4 Research question 8

1.5 Outline of the thesis 8

  1.5.1 Chapter One: Introduction and orientation 8

  1.5.2 Chapter Two: Literature study 8

  1.5.3 Chapter Three: Research methodology 9

  1.5.4 Chapter Four: Presentation of results and interpretation of findings 9

  1.5.5 Chapter Five: Conclusion 10

  1.5.6 References 10

1.6 Terminology 10

1.7 Summary 15

2. Chapter Two: Oto-acoustic emissions – a literature review 16

2.1 Introduction 16

2.2 Assessment procedures 17

2.3 The value of OAEs 22

2.4 Sensitivity and specificity of OAEs 24

2.5 Test-retest reliability of OAEs 26

2.6 Current limitations 31

2.7 Summary 32

3. Chapter Three: Research methodology 35

3.1 Introduction 35

3.2 Research aims 36

3.2.1 Main aim 36

3.2.2 Sub aims 36

3.3 Research design 37

3.4 Participants 39

3.4.1 Criteria for participant selection 39

3.4.2 Material and apparatus for selection of participants 43

3.4.3 Sample size and selection procedure 44

3.4.4 Procedure for participant selection 45

3.5 Ethical considerations 51

3.6 Pilot study 52

3.6.1 Changes to participant selection procedures 52

3.6.2 Changes to data collection procedures 53

3.6.3 General changes 53

3.7 Data collection 54

3.7.1 Material and apparatus for data collection 54

3.7.2 Data collection and data recording procedures 54

3.7.3 Sequencing of data collection procedures 56

3.7.4 Procedure for data collection: OAEs 56

3.8 Data analysis 62

3.8.1 Statistical analysis using ANOVA 63

3.8.2 Statistical analysis using ICC 66

3.9 Summary 69

4. Chapter Four: Results and discussion 71

4.1 Introduction 71

4.2 Presentation and discussion of results according to sub-aims 71

4.2.1 Sub-aim 1: Test-retest reliability of DPOAEs and TEOAEs with short term re-assessment within five minutes of initial OAE tests (M1 – M2). 75

4.2.2 Sub-aim 2: Test-retest reliability of DPOAEs and TEOAES with short term re-assessment one hour after initial OAE test (M1 – M3)....92

4.2.3 Sub-aim 3: Test-retest reliability of DPOAEs and TEOAES with long term re-assessment one week after initial OAE test (M1 – M4) 110

4.2.4 Sub-aim 4: Test-retest reliability of DPOAEs and TEOAES with long term re-assessment one month after initial OAE test (M1 – M5)....127

4.2.5 Summary of results of sub-aims 143

4.3 Presentation and discussion of results according to the main aim 143

4.3.1 DPOAE Descriptive Statistics 144

4.3.2 TEOAE Descriptive Statistics....158

4.3.3 OAE-gram Descriptive Statistics....172

4.3.4 Reliability of results....188

4.4 Summary 189

5. Chapter Five: Conclusion and recommendations 191

5.1 Summary of research findings....191

5.2 Implications of study....193

5.3 Analysis of strengths of study....194

5.4 Analysis of weaknesses of study 195

5.5 Recommendations for future research....195

5.6 Final conclusion 198

6. References 199

7. Appendices 207

7.1 Appendix A: Questionnaire 208

7.2 Appendix B: Letter of informed consent 215

7.3 Appendix C: Letter confirming ethical clearance obtained for study 218

LIST OF TABLES

Table 1: Instruments and apparatus used for participant selection 44

Table 2: Criteria for participant selection 45

Table 3: Description of participants 49

Table 4: Instruments used for data collection 54

Table 5: DPOAE ICC comparison of M1 – M2 77

Table 6: TEOAE ICC comparison of M1 – M2 81

Table 7: SD M1 and M2 TEOAE amplitude 84

Table 8: OAE-gram ICC comparison of M1 – M2 89

Table 9: DPOAE ICC comparison of M1 – M3 94

Table 10: TEOAE ICC comparison of M1 – M3 99

Table 11: SD TEOAE M1 and M3 amplitude 102

Table 12: OAE-gram ICC comparison of M1 - M3 107

Table 13: DPOAE ICC comparison of M1 - M4 111

Table 14: TEOAE ICC comparison of M1 – M4 116

Table 15: TEOAE SD M1 and M4 amplitude 120

Table 16: OAE-gram ICC comparison of M1 – M4 124

Table 17: DPOAE ICC comparison of M1 – M5 128

Table 18: TEAOE ICC comparison of M1 – M5 132

Table 19: SD TEOAE M1 and M5 amplitude 136

Table 20: OAE-gram ICC comparison of M1 – M5 140

Table 21: DPOAE Descriptive statistics 145

Table 22: DPOAE ANOVA......153

Table 23: TEOAE Descriptive statistics......159

Table 24: TEOAE ANOVA......170

Table 25: OAE-gram Descriptive statistics......172

Table 26: OAE-gram ANOVA statistical analysis......183

LIST OF FIGURES

Figure 1: Distribution of age of participants 50

Figure 2: Sequence of data collection 55

Figure 3: SD of amplitude for measurements of M1 and M2 according to frequency 86

Figure 4: SD of amplitude for measurements of M1 and M3 according to frequency. 104

Figure 5: SD of amplitude for measurements of M1 and M4 according to frequency 121

Figure 6: SD of amplitude for measurements of M1 and M5 according to frequency 137

Figure 7: Test-retest reliability of DPOAEs 151

Figure 8: ICC DPOAE single measures comparison 155

Figure 9: ICC DPOAE Average measures 156

Figure 10: ICC DPOAE Average measures comparison 156

Figure 11: Test-retest reliability of TEOAEs 163

Figure 12: TEOAE mean results at different time intervals 165

Figure 13: ICC TEOAE average measures comparison 167

Figure 14: ICC TEOAE single measures comparison 168

Figure 15: Test-retest reliability of OAE-gram 181

Figure 16: ICC OAE-gram single measures comparison 185

Figure 17: ICC OAE-gram average measures comparison 187

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
daPa	Decca Pascal
dB	Decibel
dB SPL	Decibel Sound Pressure Level
df	Degrees of freedom
DPOAE	Distortion product oto-acoustic emissions
EOAE	Evoked oto-acoustic emission
FSOAE	Frequency specific oto-acoustic emissions
Hz	Hertz
ICC	Intraclass-correlation coefficient
kHz	Kilo Hertz
LSD	Least significant difference
M1	Measurement 1 (Initial OAE measure)
M2	Measurement 2 (OAE measure within five minutes of M1)
M3	Measurement 3 (OAE measure within one hour of M1)
M4	Measurement 4 (OAE measure within one week of M1)
M5	Measurement 5 (OAE measure within one month of M1)
MDD	Minimum detectable difference
min	Minutes
ml	Millilitre
N	Number of participants (unless otherwise specified)
OAE	Oto-acoustic emissions
OAE-gram	Oto-acoustic emissions gram
SANS	South African National Standards
SD	Standard Deviation
SEM	Standard error of measurement
SNR	Signal-to-Noise Ratio
SOAE	Spontaneous oto-acoustic emission
SPL	Sound pressure level
TEOAE	Transient evoked oto-acoustic emissions

ABSTRACT

Title:	Test-retest reliability and validity of distortion product oto-acoustic emissions and transient evoked oto-acoustic emissions in normal hearing adults
Author:	Carina Grové
Supervisor:	Dr. Maggi Soer
Co-supervisor:	Prof. Bart Vinck
Department:	Speech-Language Pathology and Audiology, University of Pretoria
Degree:	M.Communication Pathology (Audiology)

The clinical value of oto-acoustic emissions (OAEs) for the identification of the integrity of outer hair cell function has been proven numerous times in research studies, and OAEs are commonly included in the audiological test battery used by audiologists. The technological advances in this field, however, require continuous research to determine the value of new equipment. The main objective of this study was to determine the test-retest reliability of distortion product oto-acoustic emissions (DPOAEs) and transient evoked oto-acoustic emissions (TEOAEs). The subjects were female adults between 18 and 25 years of age, and 30 ears were used for this study. All participants had normal middle-ear function, normal hearing sensitivity (pure tone thresholds) and present OAEs. DPOAEs and TEOAEs were performed at specific time intervals to determine the test-retest reliability, and an OAE-gram was generated by combining DPOAE and TEOAE test results. These measures were performed as an initial measure (M1), after five minutes (M2), after an hour (M3), after a week (M4) and after a month (M5). The test-retest reliability was calculated by analysing the statistics quantitatively. Results were analysed to determine the variance between each participant individually at different time intervals, and also to determine the variance of the group as a whole at different time intervals. Through this analysis high test-retest reliability between the different tests was proven. Inferential statistical Analysis of Variance (ANOVA) and Intraclass Correlation Coefficient (ICC) proved that the test-retest reliability of DPOAEs, TEOAEs and the OAE-gram (generated through a combination of the DPOAE and TEOAE results) was consistent throughout all periods of testing. The highest correlation existed between immediate re-measurements, while the largest difference was observed between M1 and M3. High test-retest reliability for all tests conducted at different time intervals from original measures were confirmed. Recommendations that arose from this study are further research in the normal hearing population for better generalization, research into specific variables between subjects, research with longer time between different tests, as well as the test-retest reliability in pathologic populations.

Keywords: Distortion product oto-acoustic emissions (DPOAEs); transient evoked oto-acoustic emissions (TEOAEs); test-retest reliability

1. CHAPTER ONE: INTRODUCTION AND ORIENTATION

This chapter provides the motivation and rationale for the study.

1.1 Introduction

The aim of this chapter is to describe the context in which the problem originated, which then serves as the rationale for conducting the research. The research question is formulated and explained within the field of audiology. All concepts, terms, and abbreviations relevant to the study are defined and explained.

1.2 Background and development of OAEs

The cochlear travelling wave was first described by Bekesy in 1940, and mention of oto-acoustic emissions (OAEs) emerged only later from these descriptions (Kemp, 1978). In 1978 Kemp postulated that a cochlear origin would be confirmed for the exciting phenomenon he observed during the scientific experiment which he was conducting on the human ear. His scientific experiment in 1978 led to the development of OAE measurements which are widely used today in audiological environments, and are the subject of this study.

OAEs are "echoes" generated in the inner ear, as a result of a small amount of energy loss during the conversion of sound energy in the inner ear (Robinette & Glattke, 2007). These resulting sounds move through the middle ear, and can be measured in the outer ear with a sensitive microphone. Early studies already suggested that OAEs are produced by the motile activity of the outer hair cells of the cochlea (Brownell, 1990). The confirmed existence of OAEs is concrete evidence that the cochlea participates in the processing of acoustic signals, and that the movement of the outer hair cells enhances the sensitivity of the cochlea to certain sounds (Lonsbury-Martin, McCoy, Whitehead, & Martin, 1993). As early as in 1960 Bekesy noted that, when damage to the outer hair cells occurs, the sensitivity and the sharp-tuning of the basilar membrane's vibration are reduced, and that this can be confirmed if absent or reduced amplitude OAEs are noted in ears with a confirmed sensory neural hearing loss (Bekesy, 1960).

OAEs can be classified according to two main types, namely spontaneous OAEs (SOAEs) and evoked OAEs. SOAEs are not considered to have significant clinical value when used in isolation. Since SOAEs can be recorded in two thirds of the normal hearing population, they cannot be regarded as providing conclusive evidence on cochlear functioning. The presence of SOAEs is regarded as a positive indicator of normal cochlear function, whereas the absence of SOAEs is not necessarily indicative of abnormal cochlear function (Hall, 2000). Another reason for the low clinical value accorded to SOAEs would be the elaborate and specifically designed instrumentation that is required to separate the continuously present SOAEs from normal internal noise levels in a test subject, as there is no stimulus presented during recording (Berger, Royster, Royster, Driscoll, & Layne, 2003). Thus, much of the recent research in the field has been aimed at investigating the applications of evoked OAEs.

Evoked OAEs (EOAEs) can further be divided into Transient Evoked OAEs (TEOAEs), Distortion Product OAEs (DPOAEs) and Frequency Specific OAEs (FSOAEs) (Hall & Mueller, 1997). TEOAEs were first mentioned in the literature, and received the most attention as potential clinical instrument initially, because early studies confirmed that TEOAEs were measurable in almost all ears with normal hearing sensitivity and normal middle ear functioning (Kemp, 1978). TEOAEs are measured after presentation of a click stimulus to the ear (Hall, 2000; Prieve & Fitzgerald, 2002). This response measures the outer hair cell sensitivity of the low frequencies up to 5000 Hz (Hall, 2000) in terms of amplitude, percentage of wave reproducibility, and signal to noise ratio (Prieve & Fitzgerald, 2002; Backus, 2007). DPOAEs are measured during the presentation of two pure tone stimuli simultaneously. When the input of these two waves are close to one another in frequency, interaction takes place that results in traveling waves in the cochlea at discrete points. These discrete points (the distortion products) are mathematically related to the frequencies of the two primary waves. Therefore, DPOAEs can be measured by using narrow band filtration techniques that focus on the specific predictable frequency result of the two initial pure tones (Prieve & Fitzgerald, 2002). This result is then captured on a DPOAE-gram.

The functioning of the outer hair cells is affected by almost all pathologies of the cochlea (Hall, 2000). OAEs measure the integrity and functioning of the outer hair cells of the cochlea (Prieve & Fitzgerald, 2002). Therefore, almost all pathologies of the cochlea can be identified by OAE measures. OAEs are only sensitive to cochlear pathologies involving the functioning of outer hair cells and require normal middle ear status for reliable results. OAEs are therefore, not defined as a hearing test, but a supplementary test to complement the results of the conventional test battery with added information on the functioning and integrity of the inner ear (Hall &

Mueller, 1997; Hall, 2000). The sensitivity of TEOAEs to distinguish between normal hearing and hearing loss has been reported to be up to 90 per cent at 2000 Hz and 4000 Hz. DPOAEs have proven similiar sensitivity at 4000 Hz (Lonsbury-Martin, McCoy, Whitehead, & Martin, 1993; Hall, 2000).

DPOAEs have several clinical advantages . The configuration of the audiogram can be estimated with some difficulty by using DPOAE results, since DPOAEs can be absent in the frequency area where the hearing loss occurs, and present in adjacent frequencies (Dannhauer, 1997). It is stable over time and can be used to detect minimal changes in cochlear status (Probst, Harris, & Hauser, 1993). DPOAEs can also be specifically use for the early detection of cochlear damage, as with ototoxicity and noise induced cochlear damage (Hall, 2000)

Previous research has explored the different clinical applications of OAEs. The following applications have been confirmed (Hall & Mueller, 1997; Hall, 2000):

- Objective evaluation as part of the audiological test battery where functional hearing loss is predicted;
- Differentiation between cochlear and retro-cochlear pathologies (because OAEs are lesion specific);
- Monitoring of ototoxicity (can be used as objective measures to determine whether dosage of ototoxic medication affects the cochlea);
- OAEs supply frequency-specific information related to the frequency area associated with tinnitus;

- Noise exposure: (OAEs can supply early warning signs even before threshold shifts are detected on the audiogram) (Berger, Royster, Royster, Driscoll, & Layne, 2003);
- DPOAEs are especially valuable in newborn hearing screening and the identification of cochlear pathology very early in life (Robinette & Glattke, 2007);

With these proven clinical values in mind, it is important to consider the reliability of these measures which are trusted by audiologists. In order for any audiological test procedure to have clinical and widespread accepted value, its results have to be reliable. One of the most important measures of a test procedure or clinical equipment's value is the reliability and repeatability of the test results. In order to assess the repeatability of results, test-retest reliability must be assessed. Test retest reliability can be defined as a measure of consistency for tests and other instruments (www.adler.edu, 2009), or the extent to which two administrations of the same test to the same group of subjects yield consistent results (www.ec.wmich.edu/glossary, 2009). The concept of OAEs in general has shown much development from the time of its first discovery in 1978, and although some tests have been performed to assess the test-retest reliability of specific measures in specific groups (Sockalingham, Kei, & Ho, 2007; Beattie, Kenworthy, & Luna, 2003), recent studies have not dealt with this type of assessment of equipment and test procedures. This has resulted in a gap in research applicablity, as it is questionable how much value there is in a specific test or equipment if its results can not be repeated over time. It is also questionable to apply these measures to a specific population if baseline results have not been established on a normal hearing population.

Even though OAEs have been in use for almost 30 years, there are still areas within the field of OAEs that lack adequate detail, in particular with regard to the description and practical usage of these measures. However, one of the biggest problems in the application of OAEs has always been the lack of concrete evidence to quantify these measures and draw some correlation between OAEs (thus cochlear functioning) and the conventional audiological test battery (hearing levels) (Wagner & Plinkert, 1999; Heitmann, Waldmann, & Plinkert, 1996).

Equipment for evoking OAEs was used for measuring cochlear functioning in the past, but limitations were found where monitoring over time, exact measures, and time-effective results were concerned. New equipment can be utilized in this regard, if its reliability can be proven. An example is the Hearing Conservation Program using ILO V6 which has not been used test-retest reliability studies, but is available for the use of audiologists. This equipment measures DPOAEs and TEOAEs separately and also combines the results in an OAE-gram (Vinck, Van Cauwenberge, Leroy, & Corthals, 1999).

In order for OAEs to be more widely accepted in clinical practices as assessment tool, results need to be proven to be reliable, and have to be presented in easily understandable format that is quantifiable and directly related to a client’s cochlear functioning. Some studies have been conducted on the test-retest reliability of DPOAEs and TEOAEs, but rarely have they been combined into one study to yield reliable results that can be quantified (Sockalingham, Kei, & Ho, 2007).

1.3 Problem statement and rationale

It is well documented that OAE testing is a widely used clinical measure in most audiology practices in addition to the traditional audiological test battery (Gelfand, 2009). This is mainly due to the fact that OAE measurements are very site-specific to cochlear damage in the outer hair cells, and assist in differential diagnoses.

In order for any additional measure (be it in the field of audiology or any other clinical field) to add value to existing measures it is essential that its results be transparent, understandable to the clinician administering the tests, reliable, and valid (Orlikoff, Schiavetti, & Metz, 2014). Though there have been many advances in the field of audiology and specifically regarding OAEs, these results have not always been accompanied by the proof of reliable and valid test results.

From the existing literature it is clear that the lack of data available on the reliability, and specifically the test-retest reliability, of OAEs is one of the main problems identified (Chan & McPherson, 2000). This problem may have far-reaching implications as audiologists' trust in objective OAE measures should increase as technology advances and the measures become more advanced. However, without supporting data to prove the test-retest reliability of these measures the value of adding OAEs to a conventional test battery may be doubted.

An investigation into the test-retest reliability of a combination of TEOAEs and DPOAEs in a normal hearing population will provide a better understanding of these measures and the quality of the results gained by audiologists using these measures. These results need to be obtained in a normal hearing population before they can be applied to a population with specific auditory pathology affecting cochlear function.

1.4 Research question

In order to fill the perceived gaps in recent research, this study will aim to address some of the most important and pressing questions in OAE research. (Chan & McPherson, 2000).

Thus, the following research question arises: What is the test-retest reliability of a combination of DPOAEs and TEOAEs over the short term and long term?

1.5 Outline of the thesis

1.5.1 Chapter One: Introduction and orientation

The aim of this chapter is to describe the context wherein the problem originated, and this then serves as the rationale for conducting the research. The research question is formulated and explained within the field of audiology. All concepts, terms, and abbreviations relevant to the study are defined and explained.

1.5.2 Chapter Two: Literature study

In this chapter the researcher expands on the theoretical underpinnings of the concepts found in the related literature so that observations and conclusions related to this specific study can be made. Abstract concepts found in the literature are described in concrete terms and the clinical value related to this study explained, so that the prior knowledge and theory required to conduct and understand this research, and to draw conclusions from this study's results, are summarised and clarified.

1.5.3 Chapter Three: Research methodology

The aim of the third chapter is to describe the complete methodology involved in this research project. The design of the research, the main aim and the sub aims are described and the methods to reach each of these aims are explained. The participants, selection procedures, data collection instruments and apparatus, and analysis are explained in such a way that any future researcher can duplicate the study in all aspects.

1.5.4 Chapter Four: Presentation of results and interpretation of findings

In the fourth chapter all the collected and processed data are presented and interpreted. As this study is quantitative in nature, the data relating to each sub-aim is presented graphically, and the results and conclusions drawn from this representation discussed immediately afterwards in the same chapter.

1.5.5 Chapter Five: Conclusion

The fifth and final chapter of the dissertation presents final conclusions based on the results of the collected data. The clinical implications of the results of each sub aim are discussed. A critical evaluation of the study is also included, as well as recommendations for future research.

1.5.6 References

All the references mentioned in this thesis appear alphabetically in the reference list.

1.6 Terminology

A list of all the relevant terminology used in the study is included alphabetically with an explanation of these terms.

Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is used in the statistical analysis of the results obtained in this study. ANOVA is a quantitative statistical method used to analyse the variance between the means (this means ANOVA is a test to detect any overall statistical significant differences between related means) (Dallal, 2013). The related means in this study is the mean of the results obtained from each of the tests (DPOAE, TEOAE and OAE-gram) for each participant at different points in time. This particular test requires one independent variable and one dependent variable. In repeated measures ANOVA, the independent variable has categories called levels or related groups. Where measurements are repeated over time, such as when measuring changes in OAE results at different time intervals, the independent variable is time. Each level (or related group) is a specific time point. Hence, for this study, there would be five time points and each time point is a level of the independent variable. The independent variable is more commonly referred to as the within subjects or within groups factor as in this study. The independent variable is the change in mean results for each of the different subjects as well as for the subjects classified as a single group. The ANOVA therefore, investigates the changes in mean OAE results for each participant over five different points in time. (Dallal, 2013)

Distortion product OAEs

DPOAEs are measured during the presentation of two pure tone stimuli simultaneously. When the input of these two waves are close to one another in frequency, interaction takes place that creates resulting traveling waves in the cochlea at discrete points. These discrete points (the distortion products) are mathematically related to the frequencies of the two primary waves. Therefore, DPOAEs can be measured by using narrow band filtration techniques that focus on the specific predictable frequency result of the two initial pure tones (Prieve & Fitzgerald, 2002). This result is then captured on a DPOAE-gram.

Intraclass Correlation Coefficient (ICC)

The ICC is one of the main statistical methods used to analyse and interpret the data. The ICC is a statistical measure that assesses the reliability of ratings by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The ratings are quantitative. (Dallal, 2013). Since the measurements are on a continuous scale, reliability is evaluated by means of the ICC. The ICC is calculated using the DPOAE response, TEOAE response and OAE-gram of each measurement to determine the relative consistency which represents the consistency of the position on individual results relative to other results The ICC is then calculated for each frequency and as comparison the first OAE measurement (M1) is compared to each subsequent measurement made.

Normal hearing adults

The population used for this study was termed “Normal hearing adults”. For the duration of this study this group was defined as participants older than the age of 18 years with no permanent or temporary impairment in their hearing. This means that this group had to comply with the following criteria:

Normal outer ear upon otoscopic examination

Normal middle ear function at all stages where measurements were made

Normal pure tone thresholds (pure tone thresholds at or above 20 dB)

Present OAEs that were detectable with the Hearing Coach ILO V6 as DPOAEs and TEOAEs above the noise floor indicated at all times measurements were required.

OAE-gram

The OAE-gram is a recent development in the field of OAEs. It is a representation of a combination of the results of DPOAEs and TEOAEs. The algorhythm of the OAE-gram generated by the Hearing Coach ILO V6 is not available in the public domain, therefore, the OAE-gram and generation of it was interpreted in conjunction with the results of the DPOAE and TEOAE measures. The current study incorporated the OAE-gram results in the interpretation of the different measures conducted at different times. (Hearing Coach User Manual, 2007)

Test-retest reliability

Test retest reliability can be defined as a measure of consistency for tests and other instruments (www.adler.edu, 2009), or the extent to which two administrations of the same test to the same group of subjects yield consistent results (www.ec.wmich.edu/glossary, 2009). Test-retest reliability was measured in this study as the reliability of one tester with different participants at different time intervals of the different tests conducted. The variable of a single tester remained the same throughout the study, to independently assess the reliability of different results between the same participants at different times, as well as different participants at the same time, and different participants at different times. Test-retest reliability was calculated using a variety of statistical calculations.

Standard error of measurement (SEM)

Further analyses after ICC is applied by calculating SEM. The SEM analyses the reliability within repeated measures in a specific subject. The formula used for calculating SEM is SEM= s√(1-ICC). In the formula "s" is the SD of all measures. Furthermore, the SEM was used to calculate the minimum detectable difference (MDD) which could be considered above the measurement error as a definite change in a participant's results (MDD95% = 1.96√2SEM). The MDD95% is used to calculate the reliability of the different tests over the specified periods of time by ensuring that the mean levels of the results fall within a certain level, and that a definite (statistical significant) change in a participant's results can be defined.

Transient evoked OAEs

TEOAEs were first recorded in the literature in 1978, and early studies confirmed that TEOAEs were measurable in almost all ears with normal hearing sensitivity and normal middle ear functioning (Kemp, 1978). TEOAEs are measured after presentation of a click stimulus to the ear. (Hall, 2000; Prieve & Fitzgerald, 2002). This response measures the outer hair cell sensitivity of the low frequencies up to 5000 Hz (Hall, 2000) in terms of amplitude, percentage of wave reproducibility and signal to noise ratio (Prieve & Fitzgerald, 2002; Backus, 2007).

1.7 Summary

DPOAEs and TEOAEs are widely used measures in clinical audiology today. These measures have been developed and refined extensively since the discovery of their existence in 1978. OAEs are used to measure the integrity of the outer hair cells of the cochlea, and even though it is not a test of hearing, it provides the audiologist with valuable information on site-specific regions of hair cell damage. These measures are used in various contexts, but the importance of the test-retest reliability of OAE tests in the field of audiology still has to be confirmed.

This chapter provides a brief outline of the origin and progress of OAE measures over the years. It explains the relevance of test-retest reliability in the audiological environment and why a study of test-retest reliability will provide insight into the value OAEs add, or do not add, to the audiology practice.

2. CHAPTER TWO: OTO-ACOUSTIC EMISSIONS – A LITERATURE REVIEW

2.1 Introduction

Only through thorough research of the fundamental principles of OAEs can conclusions be drawn on any recent developments in this specific field of audiology. Therefore, the role and clinical relevance of OAEs in audiology are explained and the most recent developments of OAEs in practical audiology are also discussed. The test-retest reliability of OAEs is of the utmost importance for reliable clinical usage and results of previous studies are discussed accordingly (Heitmann, Waldmann, & Plinkert, 1996; Wagner, Heppelmann, Vonthein, & Zenner, 2008).

The aim of this chapter is to specify and explain the concepts found in the related literature so that observations and conclusions related to this specific study on OAEs can be made. Abstract concepts found in the literature will be described in concrete terms and related to this study so that the prior knowledge and theory required to conduct this research, and to draw conclusions from this study's results, can be summarised and clarified (Mouton, 2001). The current literature provides limited evidence on the test-retest reliability and of OAEs, but the available literature will be utilized to gain a better understanding of OAEs and why the test-retest reliability is important. (Keppler, et al., 2010; Wagner, Heppelmann, Vonthein, & Zenner, 2008).

Chapter 2 is organized in such a way that the theoretical basis of OAEs and how they relate to the ear and specifically the cochlea is explained before the elaboration on the development of knowledge concerning OAEs. Test-retest reliability studies conducted over the years since the initial discovery of OAEs are analysed to provide a better understanding of why test-retest reliability is still important today.

2.2 Assessment procedures

The cochlear travelling wave was first described by Bekesy in 1940, but OAEs were only described in 1978 (Kemp, 1978). In 1978 Kemp hoped that a cochlear origin would be confirmed for the exciting which he observed during the scientific experiment he was conducting on the human ear. This scientific experiment led to the discovery of OAEs which are widely used today in audiological environments.

OAEs are "echoes" that are generated in the inner ear, as a result of a small amount of energy loss during the conversion of sound energy in the inner ear (Robinette & Glattke, 2007). These resulting sounds move through the middle ear, and can be measured in the outer ear with a sensitive microphone. Studies have suggested that OAEs are produced by the motile activity of the outer hair cells of the cochlea (Hall, 2015; Brownell, 1990). The confirmed existence of OAEs is concrete evidence that the cochlea participates in the processing of acoustic signals, and that the movement of the outer hair cells enhances the sensitivity of the cochlea to certain sounds (Lonsbury-Martin, McCoy, Whitehead and Martin, 1993 (Johnson & Seaton, 2012). Bekesy (1960) noted when damage to the outer hair cells occur, the sensitivity and the sharp-tuning of the basilar membrane’s vibration are reduced, which can, in modern times, be confirmed with absent or unreliable OAEs in ears with a confirmed sensory-neural hearing loss.

OAEs can be classified according to two main types, namely spontaneous OAEs (SOAEs) and evoked OAEs. SOAEs are not considered to have significant clinical value when used in isolation. SOAEs can be recorded in two thirds of the normal hearing population, and can therefore, not give conclusive evidence on cochlear functioning. The presence of SOAEs is regarded as a positive indicator of normal cochlear function, whereas the absence of SOAEs is not necessarily indicative of abnormal cochlear function (Hall, 2000). Another reason for the low clinical value attached to SOAEs would be the elaborate and specifically designed instrumentation that is required to separate the continuously present SOAEs from normal internal noise levels in a test subject, as there is no stimulus presented during recording (Lonsbury-Martin, McCoy, Whitehead, & Martin, 1993). Thus, much of the recent research has been aimed at investigating the applications of evoked OAEs.

Evoked OAEs can further be divided into Transient Evoked OAEs (TEOAEs), Distortion Product OAEs (DPOAEs) and Frequency Specific OAEs (FSOAEs) (Hall & Mueller, 1997). The different types of OAEs are discussed and reviewed.

TEOAEs were first mentioned in the literature, and received the most attention as potential clinical instrument initially, because early studies confirmed that TEOAEs were measurable in almost all ears with normal sensitivity and middle ear functioning (Kemp, 1978). TEOAEs are measured after a short interval after presentation of a short click stimulus (Prieve & Fitzgerald, 2002). This response measures the outer hair cell sensitivity of the low frequencies up to 5000 Hz in terms of amplitude, percentage of wave replicability, and signals to noise ratio (Hall, 2015). DPOAEs are measured during the simultaneous presentation of two pure tone stimuli simultaneously. When the input of these two waves is close to one another in frequency, interaction takes place resulting in waves in the cochlea at discrete points. These discrete points (the distortion products) are mathematically related to the frequencies of the two primary waves. Therefore, DPOAEs can be measured by using narrowband filtration techniques that focus on the specific predictable frequency result of the two initial pure tones (Prieve & Fitzgerald, 2002). This result is then captured on a DPOAE-gram.

Pure-tone audiometry measures hearing through the outer ear, middle ear, cochlea, cranial nerve VIII, and central auditory system. However, OAE measures reflect only the peripheral auditory system, which includes the outer ear, middle ear, and cochlea, but none of the central systems involved with hearing. The OAE response emanates only from the cochlea, but the outer and middle ear must be able to transmit the OAE to the recording microphone placed in the ear canal. OAE testing can be used as a screening tool to confirm the presence or absence of cochlear function of the outer hair cells in the ear, and conclusive analysis can be made for individual frequencies in the cochlea. OAEs cannot be used to describe an individual's auditory thresholds, as there is no direct correlation between these emissions and pure tone thresholds, but they can help to validate pure tone threshold measures (e.g. with suspected malingering). OAEs can also provide information about the site of a possible hearing loss. Most recent research reports have described some correlation between frequency-specific analysis of TOAEs/DPOAEs and a cochlear hearing loss (Campbell, 2014).

The functioning of the outer hair cells is affected by almost all pathologies of the cochlea (Hall, 2000). OAEs measure the integrity and functioning of the outer hair cells of the cochlea (Prieve & Fitzgerald, 2002), therefore, almost all pathologies of the cochlea can be identified by OAE measures. OAEs are sensitive only towards cochlear pathologies that involve the functioning of outer hair cells, and require normal middle ear status for reliable results. OAEs are therefore, not defined as a hearing test, but a supplementary test to complement the results of the conventional test battery with added information on the functioning and integrity of the inner ear (Hall, 2000). The sensitivity of TEOAEs to distinguish between normal hearing and hearing loss has been reported to be up to 90 per cent at 2000 Hz and 4000 Hz. DPOAEs have similar sensitivity at 4000 Hz (Campbell, 2014).

DPOAEs demonstrate several clinical advantages. The configuration of the audiogram can be predicted by using DPOAE results (Dannhauer, 1997), since DPOAEs can be absent in the frequency area where the hearing loss occurs, while being present in adjacent frequencies. The phenomenon is stable over time and can be used to detect minimal changes in cochlear status (Probst, Harris, & Hauser, 1993). DPOAEs can also be specifically used for the early detection of cochlear damage, as with ototoxicity and noise induced cochlear damage (Shupak, Tal, Sharoni, Oren, & Ravid, 2007)

Clicks are the stimuli most commonly used when recording TEOAEs, though tone-burst stimuli may also be used. Usually, 80 to 85 dB SPL stimuli are used clinically. The stimulation rate in most cases is less than 60 stimuli per second, and TEOAEs are generally recorded over approximately 20 milliseconds. When present, TEOAEs generally occur at frequencies of 500-4000 Hz. Data is then converted to the frequency results, usually in octave band analysis (Campbell, 2014; Mc Pherson, Li, Shi, Tang, & Wong, 2006).

Description of DPOAEs

DPOAEs result from the simultaneous presentation of two pure tones (designated f1 and f2) where f2 is greater than f1 (Kemp, 1978). The response of the outer hair cells of the cochlea generates a response on the equipment which is measured as a DPOAE quantifiable in terms of amplitude and frequency. Stimuli always consist of two pure tones at two frequencies (i.e., f1, f2 [f2>f1]) and 2 intensity levels (ie, L1, L2). The relationship between L1-L2 and f1-f2 dictates the frequency response. An f1/f2 ratio yields the greatest DPOAEs at 1.2 for low and high frequencies and at 1.3 for medium frequencies. Lowering the absolute intensity of the stimulus renders the DPOAEs more sensitive to abnormality. A setting of 65/55 dB SPL L1/L2 is frequently used. Responses are usually most robust at, and therefore, recorded at, the emitted frequency of 2 f1–f2; however, they are generally charted according to f2 because that region approximates the cochlear frequency region generating the response (Campbell, 2014).

Description of OAE-gram

The OAE-gram is a recent development in the field of OAEs. It is a representation of a combination of the results of DPOAEs and TEOAEs. It aims to give quantifiable results in layman's terms to OAE-results, and uses all the data captured during DPOAE and TEOAE measurements to generate this graph depicting percentage of hair cell damage. Unfortunately, the algorhythm of the OAE-gram generated by the Hearing Coach ILO V6 is not available in the public domain; therefore, the OAE-gram is interpreted in conjunction with the results of the DPOAE and TEOAE measures. The current study incorporated the OAE-gram results in the interpretation of the different measures conducted at different times. (Hearing Coach User Manual, 2007; Hearing Coach Manual to interpret Oto-Acoustic emissions, 2007).

2.3 The value of OAEs

Previous research has explored the different clinical applications of OAEs. The following applications have been confirmed (Hall, 2000):

- Objective evaluation as part of the audiological test battery where functional hearing loss is predicted,
- Differentiation between cochlear and retro-cochlear pathologies (because OAEs are lesion specific),
- Monitoring of ototoxicity (can be used as objective measures to determine whether dosage of ototoxic medication affects the cochlea),
- Tinnitus (supplies frequency-specific information related to the frequency area associated with the tinnitus),

- Noise exposure (OAEs can supply early warning signs even before threshold shifts are detected on the audiogram),
- DPOAEs are especially valuable in newborn hearing screening and the identification of cochlear pathology very early in life.

With these proven advantages in mind, it is important to consider the reliability of these measures in which audiologists place so much faith (Havelock, Kuwano, & Vorländer, 2008). In order for any audiological test procedure to have clinical and widespread accepted value, its results have to be reliable. One of the most important measures of a test procedure or clinical equipment's value is the reliability and repeatability of the test results. In order to assess the repeatability of results, test-retest reliability must be assessed. Test-retest reliability can be defined as a measure of consistency for tests and other instruments (www.adler.edu, 2009) or the extent to which two administrations of the same test to the same group of subjects yield consistent results. Concepts relating to OAEs in general have shown much development since the time of the first discovery of the phenomenon in 1978, and although some tests have been performed to assess the test-retest reliability of specific measures in specific groups (Sockalingham, Kei, Ho 2007; Franklin, McCoy, Martin, Lonsbury-Martin, 1992; Beattie, Kenworthy, Luna, 2003), not many recent studies have dealt with this type of assessment of equipment and test procedures.

Even though OAEs have been in use for almost thirty years, there are still areas within the field of oto-acoustic emissions without adequate detail in the description and practical usage of these measures. However, one of the greatest problems in the application of OAEs has always been the lack of concrete evidence to quantify these measures and draw some correlation between OAEs (thus cochlear functioning) and the conventional test battery (hearing levels) (Manley, Fay, & Popper, 2008).

2.4 Sensitivity and specificity of OAEs

In order for any measure to have practical application in the field of audiology, the sensitivity and specificity of the measure needs to be determined (Hall, 2015). The sensitivity and specificity of any test determines the clinical or research situations in which the test measures can be used, as well as the population involved in the clinical application of these test measures.

Sensitivity (which can also be called the true positive rate) measures the proportion of positives that are correctly identified as such (e.g., the percentage of a population with cochlear damage that can be identified as such by OAEs). Specificity (also called the true negative rate) measures the proportion of negatives that are correctly identified as such (e.g., the percentage of a population with normal cochlear functioning that is identified as belonging to this group via OAEs). Sensitivity therefore, quantifies the avoiding of false negatives, and specificity quantifies the avoiding of false positives. For any test, there is usually a trade-off between the measures – in order to increase sensitivity, some false negatives may be increased, or, in order to reduce false negatives, the specificity might be reduced and some true positives may be missed (Field, 2009). To establish a valid test with sensitivity and specificity within expected norms, several different studies need to be conducted in a specific field, in this case OAEs in audiology (Shupak, Tal, Sharoni, Oren, & Ravid, 2007).

Several different research studies have been conducted on the sensitivity and specificity of OAEs involving DPOAEs and TEOAEs (Lycke, Maes, & Michem, 2011; Roester, Valente, & Hosford-Dunn, 2000). OAEs as a measure have been proven to have both high sensitivity and specificity to cochlear pathology, and specifically to the functioning of the outer hair cells of the cochlea (Hall, 2000).

Many different variables have been included in studies conducted to investigate the sensitivity and specificity of OAEs in order to determine which variables affect the results obtained from OAEs. These variables included

Different testers (Ng & MacPherson, 2005),
Retesting measures (Beattie, Kenworthy, & Luna, 2003),
Different test environments (Dreisbach, Long, & Lees, 2006),
Noise levels (Backus, 2007),
Cochlear pathologies (Dreisbach, Long, & Lees, 2006)
Sample sizes (Bleech & Beattie, 2011)
Age of population (Hoth, Gudmunsdottir, & Plinkert, 2010)

Researchers endeavoured to determine the effect these variables might have on the ability of OAEs to correctly identify cochlear pathology with high accuracy. Different studies have shown that even though external variables influence OAE results, the ability of OAEs to identify cochlear pathology qualifies it as an important measure in the audiology test battery. (Backus, 2007).

2.5 Test-retest reliability of OAEs

OAEs can be used to study cochlear function in an objective and non-invasive manner. These features of emitted responses have stimulated a great deal of investigation into the utility of evoked emissions as clinical tests of hearing. One practical and essential aspect of any clinical measure is the consistency of its result upon repeated testing of the same individual (i.e., its test-retest reliability). Test-retest reliability is defined as the degree to which the results are consistent over time.

More than two decades ago Franklin, McCoy, Martin and Lonsbury-Martin measured the short and long-term reliability of DPOAEs and TEOAEs in 12 normally hearing adults which confirmed that the test-retest reliability was generally excellent (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992). However, a single study involving 12 normal hearing adults cannot serve as a basis on which to assess the applicability of a certain test measure to a diverse population. Therefore, several further research projects on this topic were initiated into the area of this topic over the ensuing years to determine the test-retest reliability in various different situations.

A normative study was attempted shortly after the test-retest studies were published (Vinck, De Vel, Xu, & Van Cauwenberge, 1996) in order to apply these results to establish normative data for normal hearing adults. These inquiries would form the basis of further studies including those involving subjects with abnormal cochlear results. This early normative study laid the foundation for further in-depth analysis of not only DPOAEs, but also TEOAEs and the development of equipment for measuring OAEs.

In 2005 very short-term (after 20 minutes) and short-term (average of 15 days later) reliability was assessed by DP-gram protocols and user-defined DP spectrum protocols in 35 normal hearing young adults (Ng & MacPherson, 2005). DPOAE amplitudes in the default frequency range (1 to 7 kHz) between test and short-term retest were correlated, and the various resolution and retest conditions showed no significant differences in reliability. It was concluded that DPOAE measurement can be considered reliable for the monitoring of cochlear function in cases like noise exposure (Ng & MacPherson, 2005).

OAEs are also widely applied in industries where noise-induced hearing loss (NIHL) is a common occurrence. The test-retest reliability is extremely important in this area to monitor cochlear functioning over a period of time, the more so because this is one of the areas where monetary value is connected to hearing (or the loss thereof) in the form of compensation. OAE methods for screening early hearing impairment and objective prediction of pure-tone thresholds in normal and hearing-impaired ears, monitoring noise susceptibility, and determining of disability in cases of NIHL, are of great importance in heavy machinery industry. Efficacy criteria and matching technologies for hearing conservations programmes were identified and consolidated in a suggested model framework for a new standard. The Hearing Coach was included in the research project to establish a baseline hearing conservation programme, which is an indication that the proven test-retest reliability was sufficient to be included in this programme (Steenkamp, 2008).

The ever-present need to quantify the test-retest reliability of OAEs resulted in larger participant groups and more robust research designs, where immediate re-measurements were compared to re-measurements of DPOAEs after five to ten days. In these studies test–retest repeatability was reduced with decreasing primary tone levels, even though repeatability values were still classified as mostly satisfactory with the lower primary tone levels. Furthermore, the SNR did not have an influence on repeatability (as long as SNR was within 6 to 35 dB).

These results were collected under circumstances resembling a clinical practice setting and yielded high repeatability. Results from a study with the widely used criteria of a minimum SNR of 6dB confirmed this as the normatively accepted SNR (Wagner, Heppelmann, Vonthein, & Zenner, 2008).

Further test-retest reliability with probe removal and including a second tester was subsequently documented in the field of audiology. The results indicated that the levels of DPOAEs L_2-level and F_2 frequency were statistically significant ($p < .0001$) predictors of a DPOAE response (i.e., the presence of a DPOAE response was more likely to be observed at higher L_1 , L_2 levels and lower F_2 frequencies regardless of test condition). Furthermore, DPOAE levels were significantly affected by L_1, L_2 level and f_2 frequency ($p < .0001$) but not by the test conditions. Intra- and inter-tester test–retest differences were not significantly different.

Although these results came from a small study involving 16 participants, it was concluded that the prevalence of missing responses coupled with large inter-subject variability and intra-subject test–retest variability are a detriment to the clinical utility of DPOAEs evoked with low level stimuli (Stuart, Passmore, Culbertson, & Jones, 2009). Thus, test-retest reliability could not always be proven through studies, even though significant contributions increased the likelihood of the applicability of these tests in different clinical environments and with greater accuracy as audiologists became more familiar with the testing procedures and the kind of results that could be expected.

Keppler further investigated the test-retest reliability of DPOAEs and TEOAEs in combination in a study involving a group of normal hearing adults. The results indicated high test-retest reliability over a shorter period of time of seven days. This significantly increased the overall reliability of DPOAEs and TEOAEs analysed together over a period of time (Keppler, et al., 2010).

In later studies the results of the pure tone audiometry showed a significant difference between right and left ears with higher threshold levels for the right ear, whereas the OAE results showed no significance. The results of the TEOAE and the OAE-gram showed gender differences where the male population exhibited less favourable results in the high frequency region. After high test-retest reliability was proven, it was concluded that OAEs can be used as an early detector of NIHL (Lycke, Maes, & Michem, 2011). High test-retest reliability was recorded amidst the results of variance between different participants. This study involved a smaller number of participants, and highlighted the difference between results for male and female participants.

Another area of interest in the test-retest reliability of DPOAEs is the amplitude of the DPOAEs and its relation to the noise floor, since DPOAEs are highly sensitive to noise and results can be influenced by the presence of external and internal noise. The effects of sample size on the test-retest reliability of the amplitude of DPOAEs and on the noise floor were investigated.

The results revealed that sample size, frequency, and intensity had little effect on the SEM of measurement. Therefore, the DPOAE data combined across all conditions resulted in the conclusion that the difference between two DPOAEs was significant if it exceeded approximately 6dB (Bleech & Beattie, 2011).This study further quantifies the test-retest reliability into dB terms, and enables audiologists to judge the reliability of measures more accurately.

Further studies examined the test-retest reliability of several different measures repeated on different days with multiple probe fits to determine the test-retest reliability. The results indicated that DPOAEs were highly stable and repeatable over multiple testing sessions, but that test-retest reliability coefficients of DPOAE inhibition magnitudes did not deliver the same test-retest reliability (Kumar, Methi, & Avinash, 2013).

DPOAEs and TEOAEs have become part of routine audiological diagnostic test batteries over the last three decades, and are being widely implemented in practices across the globe. In all audiological diagnostic tests, knowledge about the procedure's test–retest reliability is extremely important, to allow clear distinctions between true changes in monitoring over time vs. measurement or equipment deviations. The ever-enlarging area of clinical OAE applications, such as screening of hearing in infants, distinction between cochlear and retro-cochlear origins of hearing loss, objective estimation of cochlear and hearing status, the monitoring of hearing and/or hearing loss during the use of ototoxic medication, and the identification of psychogenic hearing loss, illustrates the significance of this audiological tool.

2.6 Current limitations

Even though various studies over the years in the ever-developing field of audiology have examined different areas of OAEs, there still appear to be several gaps in the existing literature. The different studies have identified and explained the theory underlying the science of OAEs and its physiological origins. The different OAE measures have been identified and studied separately and grouped together. The different applications of OAEs have been proven and have been applied in differing testing environments to determine the effects these have on the results. The effect of external factors (including noise) have been extensively documented and quantified into acceptable circumstances where OAEs can be measured. All of these results have contributed extensively to the advancement of OAEs within the field of audiology, as well as the application of OAEs in everyday testing environments as audiologists' training include more focus on OAEs and audiologists' confidence in these measures increase.

Despite all the research conducted over a wide span of time, there are some limitations in the literature that can still be identified, and addressed through this study:

- Though several studies have looked at the test-retest reliability of either DPOAEs or TEOAEs, a limited amount of research seems to have been conducted on the test-retest reliability of a combination of DPOAEs and TEOAEs,
- Test-retest reliability has been the subject of multiple studies, but the period of time of follow-up is limited to a maximum of 15 days at most, which does not include longer term re-assessments of OAEs.

These limitations that were identified lead to the formulation of the research question and goals that are put forward in the following section.

2.7 Summary

Previous studies in the field of OAEs have explored the test-retest reliability over limited periods of time, but not with a lengthened time of one month between results. The most important advantage of this study is the quantitative analyses of results after tests, which can be compared over time to notice specific trends in cochlear functioning, especially in industries. This monitoring over time is, however, dependent on the test-retest reliability of the tests. If results from one person without any known possible cochlear damage are reliable, and yield the same results from different tests, then the Hearing Conservation Program ILO V6, DPOAEs and TEOAEs, can be used in a variety of settings with confidence over a period of time.

If the results from this study show a high test-retest reliability it will indicate that OAEs and the algorhythm with the OAE-gram can currently be used for monitoring cochlear damage, but these measures should not be used in isolation and all test results should be confirmed with conventional audiometry. The algorhythm used to create the OAE-gram may also be researched and used on populations with known cochlear pathology, such as a population with noise-induced hearing loss. Furthermore, it would not be advised that audiologists rely solely on these OAE-tests for the identification of cochlear disorders, as results would not be conclusive if they cannot be repeated over time.

If the results from this study indicate high test-retest reliability, several possible advantages could arise. These would mostly involve wider application of the OAE-gram in a variety of settings:

- In industries to monitor cochlear damage in workers exposed to high noise levels (Beattie, Kenworthy, & Luna, 2003; Bockstael, et al., 2008)
- As part of the test battery to assess cochlear integrity together with hearing ability (Dannhauer, 1997)
- Monitoring of other cochlear pathologies for possible degeneration over time (Hall & Lutman, 1999)
- Monitoring of cochlear function in populations exposed to ototoxic medication and the effect of medication on cochlear functioning (Dhooge, et al., 2006).

Overall, the confidence placed in DPOAEs, TEOAEs, and an OAE-gram representing percentage of hair cell damage in the cochlea, will be improved if it can be proven that the test-retest reliability of these OAEs is consistent in a population with normal hearing and can be relied upon to determine cochlear integrity and normal cochlear functioning with certainty (Hoth, 2005).

3. CHAPTER THREE: RESEARCH METHODOLOGY

This chapter aims to give a detailed description of the research methodology followed in order for another researcher to duplicate this study.

3.1 Introduction

Leedy (Leedy & Ormrod, 2005) explains the process of research as a circle that begins with a problem and the circle is only completed when that problem is solved. According to literature (Leedy & Ormrod, 2005), research originates with a problem, but in addition to this problem clear goals and a specific plan for proceeding are needed to divide the goal (completion of the circle) into more manageable sub-problems.

The existing problems prompting this research project have been defined and explained in Chapters 1 and 2. In summary it can be noted that several research studies using OAEs and even specifically DPOAEs and TEOAEs have been conducted (Keppler, et al., 2010). The origin and development of OAEs have been documented extensively (Dannhauer, 1997). Some studies have even attempted to compare OAEs with pure-tone audiometry and audiological test results (Vinck, Van Cauwenberge, Corthals, & De Vel, 1998). Several studies have focused on the relevance of OAEs in clinical practices and audiologists' use of these measures (Lonsbury-Martin, McCoy, Whitehead, & Martin, 1993; Vinck, De Vel, Xu, & Van Cauwenberge, 1996; Beattie, Kenworthy, & Luna, 2003). However, even though these studies are widely documented in South Africa and internationally, limited data is available on the reliability of most of these studies as there is very little evidence in the literature on the test-retest value of OAEs (Keppler, et al., 2010).

The aim of the third chapter is to describe the complete methodology involved in this research project. The research design, the main aim, and the sub aims are described and the methods to achieve each of these aims explained. The participants, selection procedures, data collection instruments and apparatus, and recording, interpretation, and analysis of data are described in a replicable manner.

3.2 Research aims

These aims will be pursued to answer the specific research question.

3.2.1 Main aim

The main aim of this research project was to determine the test-retest reliability of DPOAEs and TEOAEs and a combined result of DPOAE and TEAOE results through a unique algorhythm within a group of normal hearing adults.

The following sub-aims were set up to investigate the different components of the main aim:

3.2.2 Sub aims

To determine the test-retest reliability of DPOAEs and TEOAEs with short term and long term assessments

- To determine the test-retest reliability of DPOAEs and TEOAES with short term re-assessment within five minutes of the initial OAE test (re-fit probe immediately);

- To determine the test-retest reliability of DPOAEs and TEOAES with short term re-assessment one hour after the initial OAE test;
- To determine the test-retest reliability of DPOAEs and TEOAES with long term re-assessment one week after the initial OAE test;
- To determine the test-retest reliability of DPOAEs and TEOAES with long term re-assessment one month after the initial OAE test.

3.3 Research design

According to Bless & Higson-Smith (2004) the research design is the set of procedures that guides the researcher during the research process to verify a particular hypothesis and exclude any other possible explanations not related to the research. Thus, the research design should be carefully selected to suit the research topic and question to be answered.

For the purpose of this study a quantitative research approach was employed, as the end-product data was presented in numerical format in order to determine the test-retest reliability of DPOAEs and TEOAEs in a normal hearing population (Keppler, et al., 2010). Detailed, accurate presentations of DPOAEs and TEOAEs, in combination with the OAE-gram, were obtained as test results. The raw data was analysed quantitatively as comparisons were made on the reliability and replicability of test results over short and longer periods of time. The separate DPOAE and TEOAE test results were compared over time, as well as the OAE-gram incorporating both these test results to calculate percentage of hair cell damage. The data was presented numerically in order to allow conclusions regarding the replicability of the different OAE tests as measured by the specific equipment.

This study used an explorative-descriptive design (Orlikoff, Schiavetti, & Metz, 2014). This type of research can be classified as explorative because, even though previous studies have investigated the test-retest reliability of OAEs, not many studies have dealt with DPOAEs and TEOAEs combined with an OAE-gram describing percentage of hair cell damage on the sample used in the current study. Furthermore, there are not many studies detailing the test-retest reliability of DPOAE and TEOAE tests over a specified time span in a homogenous population (Wagner, Heppelmann, Vonthein, & Zenner, 2008). A further description of the findings of this study, relating the test-retest reproducibility to the reliability of OAE equipment, is provided with reference to the time frame between different tests on the same subjects in this study.

According to (Orlikoff, Schiavetti, & Metz, 2014) this study can also be described as comparative because the test-retest results and protocol were compared with each other at different time intervals to determine the reliability of the equipment. Some qualitative aspects were included in the analysis of data after data collection. It was expected that not all test subjects would exhibit the exact same test-retest reliability over time, therefore, an in-depth analysis of the questionnaire, as well as audiological test results of each participant, were needed to explain discrepancies and to justify obtained results.

3.4 Participants

The selection criteria applied to participants are an integral part of the preparation for this study. The appropriate selection criteria for participants ensured that the research question could be answered without discrepancies. The selection criteria for participants were as follows:

3.4.1 Criteria for participant selection

Selection criteria were established to ensure that participants were appropriate for this study. Inclusion criteria are justified in 3.4.1, and detail on required test results for specific audiometric tests given in 3.4.4.

- Normal outer and middle ear function

  The sound stimulus used during OAE testing travels through the middle ear – normal middle ear functioning is a prerequisite for obtaining OAEs as the middle ear should conduct the stimulus to the inner ear for an emission to be elicited and for reliable OAE results to be obtained. Robinette and Glattke, (2007) mentioned that the middle ear pressure especially affects OAE measures and results, and this variable should therefore, be monitored during the study. Middle ear function was assessed during the selection of participants at initial evaluation, and also during the follow-up evaluations to ensure that middle ear pathology did not influence test results at any stage of data collection. Participants had to present with normal middle ear functioning confirmed by type A tympanograms at each testing with normal values for middle ear pressure (-50 daPa to 50 daPa) normal ear canal volume (0.8 – 1.2 ml) and normal compliance of the tympanic membrane (0.3 – 0.7 ml) (Katz, 2014). One acoustic reflex also had to be present at 1000Hz for participants to be selected for the study (Hall & Mueller, 1997; Katz, 2014). Normal outer ear function was confirmed by otoscopic examination and presence of light reflex on tympanic membrane and normal appearance of external meatus (Katz, 2014). Participants therefore, had to have normal outer and middle ear function.

- Hearing status

  As the goal of this study was to determine the test-retest reliability of OAEs in normal hearing adults, the hearing status of participants was an important factor in determining candidacy for participating in the research. Pure tone audiometry was conducted on all participants, and pure tone thresholds were required to fall within the normal range for adults (i.e. 0 – 20 dB for all frequencies across the frequency spectrum 500 Hz to 8000 Hz, including 3000 Hz and 6000 Hz) (Vinck, Van Cauwenberge, Corthals, & De Vel, 1998; Pickles, 2013; Gelfand, 2009)

- Gender

  According to Bless & Higson-Smith (2004) any sample used during research must have properties which make it representative of the population to which the study aims to generalize its results. In order for this study to have some generalization characteristics, it has to represent the general population. However, according to some studies, there are differences between OAEs in male and female populations (McFadden, Martin, Stagner, & Maloney, 2009; Keppler, et al., 2010). Therefore, only female participants were utilized during the study, although this was not indicated in the initial research proposal. This ensured that all data obtained during the study could be compared between subjects without taking gender differences into consideration to reach more reliable conclusions. As the aim of this study was to determine the test-retest reliability of DPOAEs and TEOAEs, it was imperative that the data collected for each participant could be compared between tests and with other participants to reach a conclusion about test-retest reliability. To ensure reliable results and easy comparison, this group needed to be as homogenous as possible. Only female participants were therefore, selected for this study.

- Language proficiency

A thorough and reliable case history was required of each participant in this study, as inaccurate information in the case history could lead to misleading results obtained in the test battery. It was therefore, imperative that there was a common language of communication established between the researcher and participants. To minimize the risk of a language barrier, all participants were therefore, required to be fluent in Afrikaans or English, as the researcher is proficient in these languages (Bless & Higson-Smith, 2004).

- Age

Participants had to be between 18 and 25 years of age to participate. This study required adults with normal hearing to participate, in order to collect reliable baseline data with as few variables as possible influencing results. In South Africa, the legal age at which a person is considered an adult is 18 years, therefore, an age of 18 years formed the bottom boundary for age of participation. Additionally, the OAE algorhythm utilized has not been developed for people under the age of 18 years. Presbycusis is the natural aging process of the inner ear, where the functioning of the hair cells of the cochlea deteriorates without other contributing factors. Although early deterioration is not detectable with conventional audiometry in hearing thresholds, OAEs are extremely sensitive to the functioning of the outer hair cells in the cochlea, and even small changes to these hair cells would have been detected by OAEs. To acquire reliable results as baseline data for the reproducibility in normal hearing adults, participants should have had as little damage to the cochlea as possible and therefore, the upper age limit for participants was set at 25 years of age (Hoth, Gudmunsdottir, & Plinkert, 2010).

- Noise exposure

  Participants should not have had excessive noise exposure in the past, nor any noise exposure in the 48 hours preceding each test. Noise exposure affects hearing and especially the functioning of the hair cells in the cochlea (Tlumak & Kileny, 2001; Vinck, Van Cauwenberge, Corthals, & De Vel, 1998; Vinck, Van Cauwenberge, Leroy, & Corthals, 1999). Continuous noise exposure over prolonged periods of time may cause permanent threshold shifts and participants with elevated thresholds were eliminated from this study by pure tone audiometry confirming normal thresholds. However, exposure to noise in excess of 85 dB for short periods of time may cause a small temporary threshold shift which might not be detected by pure tone audiometry, or heightened thresholds might still fall within the normal range for some participants even though a temporary emission shift may be present affecting the reliability of OAEs. It was therefore, necessary to ensure that all participants in this study had not been exposed to excessive noise in the 48 hours preceding both the initial and the follow-up evaluations. Noise exposure in the preceding 48 hours would still have had an effect on participants' hearing and especially on the functioning of inner hair cells, which had to be optimal at the time of testing to ensure reliable results (Vinck, Van Cauwenberge, Corthals, & De Vel, 1998). Participants therefore, had to have no history of noise exposure in the 48 hours preceding each test.

- Present and measurable OAEs

  All participants had to have present and measurable DPOAEs and TEOAEs in order to be included in this study, as the OAEs are the most important data collected during this study. Participants were selected based on all the previous criteria and OAEs were only recorded and measured if they were present during the initial visit.

3.4.2 Material and apparatus for selection of participants

The following material and apparatus were used to ensure that all participants selected for participation in the study fulfilled the requirements that were expected of the population studied during this research:

Table 1: Instruments and apparatus used for participant selection

Instruments and apparatus used:	**Justification:**
Questionnaire	Used to select participants and ensure that all pre-determined criteria are met (Appendix A).
Welch Allyn Otoscope	Used for otoscopic examination
GSI 33 immittance meter (Calibrated 2009)	Used for tympanometry including immittance measures and audiological reflexes
Milton fluid	Used to disinfect probes used during immittance testing and OAE measures
GSI 61 audiometer (Calibrated 2009)	Used in performing pure tone audiometry and speech audiometry.
Soundproof booth	Used for accurate and reliable audiometry results in a controlled environment.
Forms for recording of data	A standard audiogram form of the University of Pretoria was used to write up all results obtained.
HP Photosmart DS163	Used to print results obtained with the OAE tests.

3.4.3 Sample size and selection procedure

The sample size had to be sufficient to provide enough data in order for the research question to be answered (Maxwell & Satake, 2006). Ear-specific information of 30 participants was acquired, and described after thorough testing. The “rule of thumb” (Neuman, 1997) was used to determine the appropriate number of participants needed in this study. The method is based on previous experience with examples of sample sizes that yielded enough information to draw conclusions from statistical analysis. Previous studies (Erasmus & Grové, 2008) used a similar approach. A smaller sample size was used because the population associated with this study has homogenous characteristics, and therefore, conclusions could be drawn from a smaller group of participants. Several OAE measurements were made for each participant in this study spread over a pre-determined time lapse between each test.

3.4.4 Procedure for participant selection

The profile of the participants is presented in the following table. The information was obtained from the questionnaire and test battery:

Table 2: Criteria for participant selection

Criteria:	**Limits:**
Number of participants	30
Age:	**Participants had to be in the age group 18 to 25 years** Adult age group in which deterioration of cochlea is not yet present
Audiological status:	**Participants had to have:** **Normal outer ear,** Determined with Otoscopy: Visually normal outer ear and tympanic membrane with light reflex present. (Hall, 2000) **Normal middle ear function**, Determined with tympanometry and reflexes Type A tympanogram with values within normal range for Pressure -50 daPa to +50 daPa Compliance 0.3 ml to 0.7 ml Ear canal volume 0.8 ml to 1.2 ml Reflexes ipsi- and contralaterally present at 500 Hz, 1000 Hz and 2000 Hz, 70 dB to 90 dB above pure tone thresholds (Hall 2000). **Normal pure tone thresholds**, Determined with pure tone audiometry Thresholds 0 - 20 dB across the frequency spectrum from 125 Hz to 8000 Hz (Vinck, Van Cauwenberge, Corthals, & De Vel, 1998)
Language proficiency:	**Participants had to be Afrikaans or English speaking** As the researcher is proficient in these languages and clear instructions could be given at all times, to ensure a correct case history could be obtained from the participant without any language barriers that might influence understanding and answering some integral questions that may affect the outcome of the study.
Gender:	**Participants had to be female** This will ensure accurate representation and conclusions as population is more homogenous (McFadden et al, 2008).
Noise exposure:	**Participants may not be exposed to noise levels above 70 dB 48 hours preceding the testing** Exposure to noise may induce a temporary threshold shift which will be evident up to 48 hours after noise exposure, and will result in inaccurate results for a specific participant.
Ototoxic medication:	**No ototoxic medication exposure** Participants may not have been exposed to ototoxic medication, whether an influence is noted on the audiogram or not, as ototoxic medication affects the hair cells of the cochlea and results can not be deemed reliable.

3.4.4.1 Questionnaire

The questionnaire that each participant was required to complete contained basic identifying information about the participant, in order to distinguish participants' unique test results from each other and to determine if participants adhered to the selection criteria. Each participant was assigned a number between 1 and 40 to ensure the confidentiality of their information, and as only one ear of each participant was used for data collection a test ear was randomly assigned (either left or right).

The questionnaire further served as the first tool for identifying possible candidates to participate in the study. Essential information, such as age, language proficiency, and length of time with no noise exposure, had to be determined before testing commenced to select suitable participants for this study.

Their history of short and long term noise exposure might have affected OAE test results and was thus covered in the questionnaire. Participants were also expected not to have been exposed to excessive noise in the 48 hours preceding any appointment related to the study, as a temporary threshold shift due to noise exposure might still have been present (included as prerequisite selection criteria).

3.4.4.2 Otoscopic examination

A full otoscopic examination was performed on each participant to ensure normal outer and middle ear status. During the otoscopic examination, the status of the outer and middle ear was observed, and the following characteristics disqualified a participant from further participation

(Hall, 2000; Yost, 2006): Otitis externa, excessive cerumen in ear canal, collapsed external auditory canal, foreign objects in ear canal, perforation of tympanic membrane, tympanosclerosis. In other words, all physical observable landmarks of the outer ear and tympanic membrane should be representative of a healthy, normal external ear. A clear light reflex and pearly white tympanic membrane were also a prerequisite for all participants (Katz, 2014).

3.4.4.3 Immittance measures

Immittance measures performed included tympanometry and acoustical reflexes. Immittance measures were used to confirm normal middle ear function, and the following results were required for participants to continue with testing:

Type A tympanogram with values within normal range for:

Middle ear pressure	-50 daPa to +50 daPa (Katz, 2014)
Compliance	0.3 ml to 0.7 ml (Katz, 2014)
Ear canal volume	0.8 ml to 1.2 ml (Katz, 2014)

Reflexes ipsi- and contralaterally present at 500 Hz, 1000 Hz and 2000 Hz, 70 dB to 90 dB above pure tone thresholds (Yost, 2006; Katz, 2014; Hall, 2000).

Immittance measures were completed during the initial testing of each participant, but also repeated before every re-measurement to ensure that middle ear variability did not influence OAE results.

3.4.4.4 Pure tone audiometry

Pure tone audiometry was performed to determine participants' hearing sensitivity. Testing took place at the University of Pretoria in an audiometric booth approved by the SANS for pure tone audiometry (South African National Standard [SANS] 10154-1, 2004; South African National Standard [SANS] 10154-2, 2004; South African National Standards [SANS] 10182, 2006). Pure tone audiometry was carried out across the frequency spectrum from 125 Hz to 8000 Hz. All pure tone thresholds were expected to be between 0 and 20 dB, as even a minimal hearing loss may affect the presence of especially TEAOEs. (Vinck, Van Cauwenberge, Leroy, & Corthals, 1999; Gelfand, 2009).

The profile of the participants is presented in the following table. The information was obtained from the questionnaire and test battery. The most important identifying traits of the sample are highlighted through tables and figures in the following section.

Table 3: Description of participants

Participant	**Pure tone (in dB)**			**Pure tone average (dB)**	**Standard deviation (dB)**	**Ear selected**	**Age in years**
	500 Hz	**1000Hz**	**2000Hz**				
1	5	10	5	6.7	2.36	Right	20
2	5	10	5	6.7	2.36	Right	20
3	10	15	20	15	4.08	Left	21
4	10	0	5	5	4.08	Left	19
5	10	5	5	6.7	2.36	Right	18
6	10	5	0	5	4.08	Left	19
7	10	5	5	6.7	2.36	Right	19
8	10	10	10	10	0.00	Left	21
9	10	10	5	8.3	2.36	Left	21
10	0	0	0	0	0.00	Right	19
11	0	5	5	3.3	2.36	Right	18
12	10	5	15	10	4.08	Left	20
13	15	0	5	6.7	6.24	Right	21
14	10	15	0	8.3	6.24	Left	21
15	5	0	0	1.7	2.36	Right	20
16	5	5	0	3.3	2.36	Right	20
17	5	10	5	6.7	2.36	Left	19
18	5	10	10	8.3	2.36	Left	18
19	10	15	10	11.7	2.36	Left	20
26	5	10	15	10	4.08	Left	18
27	5	10	0	5	4.08	Right	18
28	10	10	5	8.3	2.36	Right	18
29	5	0	5	3.3	2.36	Right	18
30	10	10	10	10	0.00	Left	19
31	5	5	0	3.3	2.36	Left	18
32	5	5	0	3.3	2.36	Right	19
33	10	0	10	6.7	4.71	Right	18
34	20	15	15	18.3	2.36	Left	18
35	15	10	5	10	4.08	Left	18
36	10	5	15	10	4.08	Right	19
37	10	5	5	6.7	2.36	Left	19
Average	8.23	7.10	6.29	7.26	2.90		19.16
SD	4.12	4.71	5.38	3.73	1.47		1.08

In Table 3 the main features of the participants are summarised. In Table 3 it is indicated that all participants selected for this study met the selection criteria of normal hearing as measured with pure tone audiometry. Only thresholds for 500 Hz, 1000 Hz and 2000 Hz appear in Table 3 as these frequencies are used to calculate the pure tone average. Thresholds for all participants, at these frequencies as well as the other frequencies tested (250 Hz to 8000 Hz), were between 0 – 20 dB. In Table 3 an accurate description of the age of all the participants in this research study is also provided. It is apparent from Table 3 that all participants were of legal age (18 years) to participate in this study. All participants also adhered to the selection criteria whereby they had to be between the ages of 18 and 25 years to limit the percentage of hair cell damage already incurred due to the natural aging process. The average age and SD calculated also confirm that participants fall within the expected age range. A sequential number was assigned to each participant at the start of the study, therefore, participant numbers ranged from 1 to 37 as some participants were excluded due to selection criteria even though 31 ears were used.

The age of the participants of the study were considered as important selection criteria to ensure legal requirements were met and that hearing levels were not affected by presbycusis. The distribution of the age of the participants is explained in Figure 1.

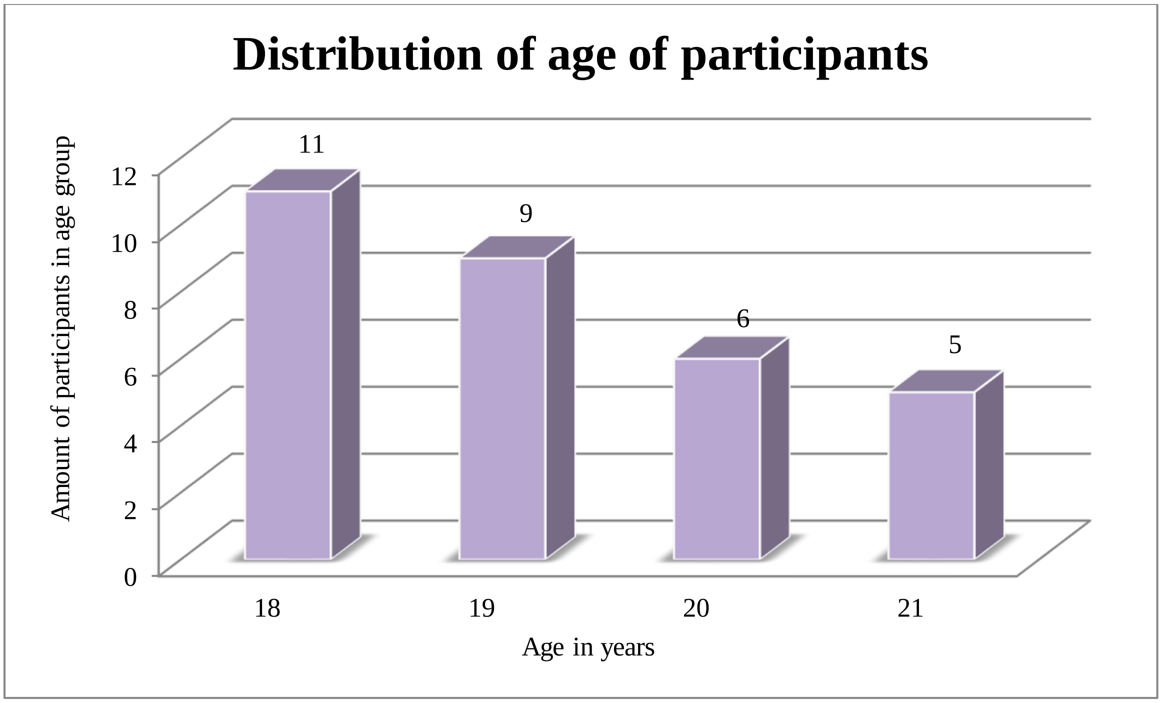


Figure 1: Distribution of age of participants

Figure 1 depicts the distribution of the age of participants. All participants were between 18 and 25 years of age with the majority of participants (n=11) aged 18 years at time of testing. This ensured that hair cell damage due to the aging process was minimized in the sample of participants used for this study.

3.5 Ethical considerations

According to Struwig and Stead, 2001, research ethics provide the researcher with guidelines so as to conduct research in a morally acceptable way. Research ethics can be divided into four categories (Leedy & Ormrod, 2005): Protection against harm, informed consent, confidentiality and anonymity, and honesty with colleagues.

In order to adhere to all requirements included in research ethics, ethical clearance was obtained from the Research Committee of the Department of Speech Language Pathology and Audiology, and the Research Proposal and Ethics Committee of the Faculty of Humanities at the University of Pretoria (Appendix C).

Informed consent was obtained from all participants in this study (all participants were of legal age – over 18 years). An informed consent letter (Appendix B) was provided to explain the procedures involved in the study in full detail to participants. Each participant received this letter of explanation, and then signed this letter as confirmation that all the aspects of the research were understood. The letter of consent clearly indicated that the participation in this study was entirely voluntary, and that any participant could voluntarily withdraw from the study at any time without any negative consequences. Specific further measures were taken by the researcher to ensure confidentiality, and personal records were organized according to a random number assigned to each of the participants. All personal information required was kept strictly confidential. Data obtained will be stored for 15 years as required by the Research Committee at the University of Pretoria.

3.6 Pilot study

A pilot study was completed prior to commencing data collection. For the pilot study 3 participants were selected and participant selection procedures as well as data collection procedures were performed. The pilot study highlighted some possible problems during participant selection and data collection procedures that resulted in some changes to the data collection sequence. The changes that were applied to the study included:

3.6.1 Changes to participant selection procedures

The following changes were applied to the questionnaire:

- A table detailing noise exposure history was added for more detailed information about participants' noise exposure as the original questions elicited mainly repetitive answers from participants and the table gave more in-depth knowledge about noise exposure.
- An additional section was added where participants could enter any general comments regarding their hearing as one participant had a particular concern unrelated to a specific question and asked where she could fill it in.

3.6.2 Changes to data collection procedures

The following changes were applied to data collection procedures:

- Participants were shown into a separate waiting room for the first follow-up (1 hour after initial measure), as a new participant could then undergo participant selection tests at the same time. Thus, the researcher could handle one follow-up test (M2) after each initial measure (M1) as participant selection tests were calculated to take 45 minutes to one hour.

3.6.3 General changes

- Participants were contacted (telephonically or via cellphone message) before each follow-up visit to remind them about appointments as one participant in the pilot study did not remember all her appointments. Participants were also given appointment cards with details of all their follow-up visits to encourage future attendance.

3.7 Data collection

A description of the instruments and apparatus used during the data collection for this study is included.

3.7.1 Material and apparatus for data collection

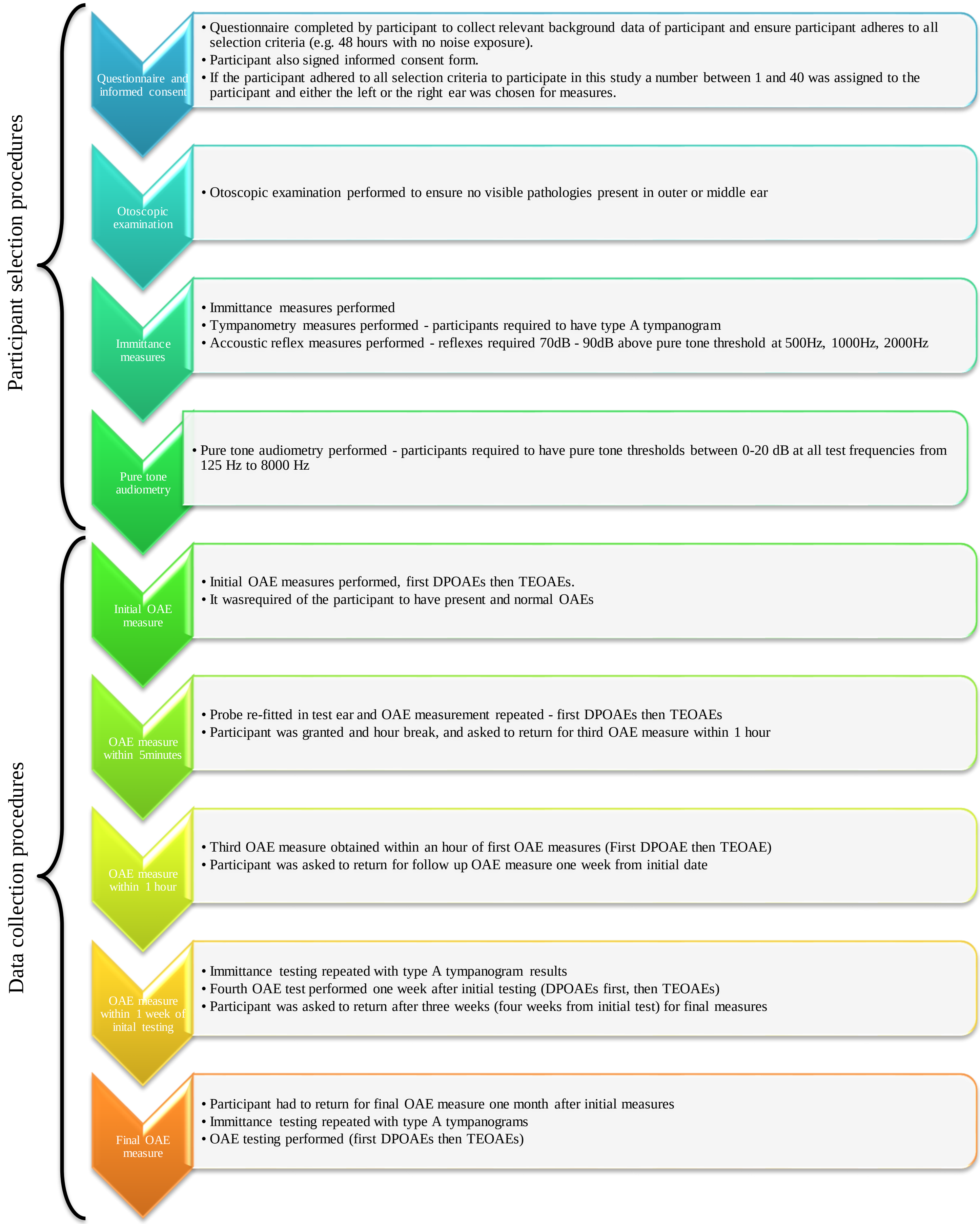
Table 4: Instruments used for data collection

Instruments and apparatus used	Justification
DP-echoport, ILO V6 software, Hearing Coach	Used to obtain DPOAEs and TEOAEs from selected participants, and to generate an OAE-gram determining percentage of hair cell damage
GSI 33 immittance meter	Used for tympanometry including immittance measures and audiological reflexes

In Table 4 the instruments used for data collection during this study are listed. These instruments were used for the data collection and data recording procedures and met all pre-requisites standards for reliable and repeatable data collection.

3.7.2 Data collection and data recording procedures

The sequence of events during data collection procedures is depicted in Figure 2. Sequencing of events during data collection and data recording played an important role during this study as different OAE measures had to be recorded at a pre-defined time lapse from initial measure for each participant in the study. The sequence in Figure 2 was followed for data collection for each participant after the participant had been selected through the criteria applied in participant selection.



Participant selection procedures

Questionnaire and informed consent
- Questionnaire completed by participant to collect relevant background data of participant and ensure participant adheres to all selection criteria (e.g. 48 hours with no noise exposure).
- Participant also signed informed consent form.
- If the participant adhered to all selection criteria to participate in this study a number between 1 and 40 was assigned to the participant and either the left or the right ear was chosen for measures.

Otoscopic examination
- Otoscopic examination performed to ensure no visible pathologies present in outer or middle ear

Immittance measures
- Immittance measures performed
- Tympanometry measures performed - participants required to have type A tympanogram
- Accoustic reflex measures performed - reflexes required 70dB - 90dB above pure tone threshold at 500Hz, 1000Hz, 2000Hz

Pure tone audiometry
- Pure tone audiometry performed - participants required to have pure tone thresholds between 0-20 dB at all test frequencies from 125 Hz to 8000 Hz

Data collection procedures

Initial OAE measure
- Initial OAE measures performed, first DPOAEs then TEOAEs.
- It wasrequired of the participant to have present and normal OAEs

OAE measure within 5minutes
- Probe re-fitted in test ear and OAE measurement repeated - first DPOAEs then TEOAEs
- Participant was granted and hour break, and asked to return for third OAE measure within 1 hour

OAE measure within 1 hour
- Third OAE measure obtained within an hour of first OAE measures (First DPOAE then TEOAE)
- Participant was asked to return for follow up OAE measure one week from initial date

OAE measure within 1 week of inital testing
- Immittance testing repeated with type A tympanogram results
- Fourth OAE test performed one week after initial testing (DPOAEs first, then TEOAEs)
- Participant was asked to return after three weeks (four weeks from initial test) for final measures

Final OAE measure
- Participant had to return for final OAE measure one month after initial measures
- Immittance testing repeated with type A tympanograms
- OAE testing performed (first DPOAEs then TEOAEs)

3.7.3 Sequencing of data collection procedures

A convenience sample of young volunteers in the appropriate age group 18 to 25 years was used for this study. A date and time that suited each participant were agreed upon. Each participant was required to meet the researcher on two additional dates for data collection, respectively one week and one month after the initial testing. Appointments were scheduled to be convenient for the participant yet still fall within the time frame required to collect the data.

On the pre-arranged date, the participant met the researcher at the Department of Speech-Language Pathology and Audiology. Upon her arrival, the letter of informed consent (Appendix B) was read and signed. After informed consent was obtained, the rest of the data collection procedures commenced.

The participant was required to complete the questionnaire (Appendix A) to ensure that the participant met all criteria to participate in the study. The researcher ensured that all the required criteria were met before any other tests commenced. Testing of the participant’s auditory skills then commenced (described below), and after normal results were obtained in all areas, OAEs were performed.

3.7.4 Procedure for data collection: OAEs

OAE measures were only performed if participants adhered to all set criteria during selection procedures for gathering of data. If all criteria for participant selection were met, OAEs were performed as the presence of OAEs was a prerequisite for continued participation (Hearing Coach Manual to interpret Oto-Acoustic emissions, 2007).

The OAE-measures were the most important measures for this study, but the results could not be used or interpreted without the additional information gained from the results of a full audiometric test battery. OAE-tests were conducted if all selection criteria were met. DPOAEs and TEOAEs were performed in the same order each time to ensure as little as possible variability between the different sessions.

DPOAEs were initially performed first in the selected ear, followed by TEOAEs. An OAE-gram was automatically generated by the software if DPOAE and TEOAE results were obtained. Thereafter the probe used for OAEs was removed, and subsequently re-fitted.
DPOAEs were again performed, followed by TEOAEs, and the generation of an OAE-gram which concluded the first short term re-measurement.

Thereafter participants were granted a one hour break. Participants were instructed to wait in the waiting area of the Department of Communication Pathology at the University of Pretoria until the researcher indicated that an hour had lapsed since initial measurements were recorded. When participants returned the second short term re-measurement was completed, again in the same order: DPOAEs, then TEOAEs, followed by OAE-gram. This concluded the short term re-measurement.

Participants were required to visit the Department of Speech-Language Pathology and Audiology one week after the initial measurements for the first of the long term re-measurements.

Following otoscopic examination and immittance measures, OAEs were again completed in the same order, DPOAEs then TEOAEs and an OAE-gram.

Participants were required to visit the Department of Speech-Language Pathology and Audiology for the last of the long term re-measurements one month after the initial measurements. On completion of otoscopic examination and immittance measures, OAEs were again completed in the same order, first DPOAEs, TEOAEs and then OAE-gram.

Although 40 participants initially agreed to participate in this study, not all participants attended all the assessments. The numbers assigned to active participants initially were, however, still utilized. Participants 20 to 25 and 38 to 40 completed the questionnaire and the selection test procedures, but failed to attend the further evaluations, thus the results of 30 participants could be analysed.

DPOAE and TEOAE measurements were obtained with the ILO 288 USB II (Otodynamics Ltd.) paired with the ILOv6 software on an accompanying laptop. The DPOAE probe used for data collection was calibrated using the 1 cc calibration cavity supplied with the ILO module before commencement of each session, and the check-fit procedure was completed for DPOAEs and TEOAEs to ensure correct probe fitting for each test (Hearing Coach User Manual, 2007).

TEOAE stimuli were rectangular pulses of 80 µs presented at a rate of 50 clicks per second. Click intensity of 85.9 ± 2 dB SPL were used (after adjusting the gain of the stimulus according to the detected peak stimulus by the microphone). The recording of TEOAEs was terminated after 260 accepted sweeps with a noise rejection setting of 49-50 dB SPL (4 mPa). Only TEOAE measures with stimulus stability of 90% and higher were used as valid measurements. The emission amplitudes and noise amplitudes were analysed in half-octave frequency bands at 1.0 kHz, 1.4 kHz, 2.0 kHz, 2.8 kHz, and 4.0 kHz, and the total emission amplitude across frequencies was also determined. General Diagnostic Mode was selected for TEOAE measures to obtain the selected results. For the purpose of this study, TEOAEs were only considered present if the emission amplitude was at least 6 dB greater than the noise amplitude in every half-octave frequency band. TEOAEs were measured using the non-linear differential method. (Hearing Coach User Manual, 2007)

The 2f1-f2 DPOAEs were measured after the simultaneous presentation of the two primary frequencies f1 and f2. For the purpose of this study f2/f1 equalled 1.22. A fixed primary tone level combination across frequencies was presented: L1/L2 was 75/70 dB SPL. Eight points per octave were used. f2 Ranged from 0.841 kHz to 8.0 kHz, and each consecutive centre frequency was calculated as $f_{I+1}= 2^{(1/8)}f_I$, where f_I represented the previous centre frequency. A noise artefact rejection level of 49.5 dB SPL was selected. Excessive noise at specific frequencies (noise was considered excessive if it exceeded -5 dB SPL) was reduced by temporal averaging. DPOAEs were considered present if the emission amplitude at all individual frequencies were at least 6 dB above than the correlating noise amplitude. If these criteria were not met, both the emission amplitudes and noise amplitudes were treated as absent. If DPOAEs were absent at all frequencies in half-octave frequency band, emission and noise amplitudes were considered as missing data in that half-octave frequency band. Only emissions and noise amplitudes considered as present were converted to pressure levels and averaged into half-octave frequency bands where f2 ranged from 0.841 kHz to 1.189 kHz (five frequencies), 1.297 kHz to 1.542 kHz (three frequencies), 1.682 kHz to 2.181 kHz (four frequencies), 2.378 kHz to 3.084 kHz (four frequencies), 3.364 kHz to 4.362 kHz (four frequencies), 4.757 kHz to 6.727 kHz (five frequencies), and 7.336 kHz to 8.0 kHz (two frequencies) respectively for half-octave frequency bands with centre frequencies 1 kHz, 1.4 kHz, 2 kHz, 2.8 kHz, 4 kHz, 6 kHz, and 8 kHz. (Hearing Coach User Manual, 2007; Roester, Valente, & Hosford-Dunn, 2000; Keppler, et al., 2010).

The following steps were followed to ensure reliable DPOAEs and TEOAEs (and the resulting OAE-grams were collected):

General procedures for OAE measures were performed in the following order:

- The ILO V6 desktop icon was selected and the program opened.
- Any noise sources in the testing room was minimized (including noise sources from the participant as these could also influence results obtained)
- The following steps were followed in the program: File > Options > Setup Options > Stims was selected. At DPOAE measures “F1 level” was set to 75dB and “F2 level” to 70 dB. This was saved as the standard test default so that it could be accessed with all participants in “General Diagnostic” and did not need to be changed for each participant.

- Before testing commenced, calibration was executed for each session of measures: The calibration icon was selected and "Start Calibration Test" selected to start the calibration. After calibration, the testing panel was closed before testing commenced.
- Relevant identifying information for participants was entered into the program in the "This patient" panel. No personal identifying information was entered into the program, instead, only the participant's number and the test (M1, M2, M3, M4 or M5) was entered as the "Patient name" for easy identification.
- The correct "Mode" was selected by "File > Options > Start / Stop / Score" and changing the protocol to "General diagnostic".
- The "tests" option was selected and thereafter "start test default action", to select the applicable DPOAE option.
- The correct size probe for the participant's ear was selected and connected with ILO V6 equipment, and placed securely in the participant's ear without supporting the connecting cable.
- Probes were changed and disinfected between every time results were recorded.

3.7.4.1 DPOAE data collection:

The "Start Test"- option was selected to perform the DPOAE, and "Continue" was selected if a correct probe fit was obtained. The test was then started and performed at least twice over the frequency spectrum (using "Auto Mode"). If the noise floor at a specific frequency was too high, the arrows were utilized to repeat the test at the selected frequency until the noise floor fell within the expected range. For reliable results the test was only accepted if the noise floor was

≤ 0dB was. The test was ended if reliable results were obtained by selecting the “End test” option. Hereafter results were carefully interpreted to ensure reliability and correct measures. Data was saved by pressing the “Save” option and choosing the appropriate ear. (Hearing Coach User Manual, 2007).

3.7.4.2 TEOAE data collection

The stimulus type was changed initially by choosing “Tests > Stimulus” and then selecting “Nonlinear click” as stimulation. This stimulus method uses a click stimulus repeated 50 times per second, resulting in a 20 millisecond response window. The “Start Test > TEOAE” option was selected to start the test. The “Checkfit peak stimulus” had to fall within the range of 84 ±3dB, with the arrow in the green part of the scale. When the stimulus fell outside of this area, the “Auto-adjust” option was selected to adjust the stimulus automatically. The “Continue”-option was selected to continue with testing. This test was automatically ended if the prerequisite criteria were met (with regards to noise floor, number of sweeps and reproducibility). Data was saved by choosing the “Save” option and the appropriate ear.

3.8 Data analysis

The main data used in this study was obtained from the DPOAE tests, the TEOAE tests as well as the generated OAE-gram. The questionnaire and results of the audiological test battery were mainly used as a reference to explain unexpected findings in the OAEs that would not correlate with a normal audiological evaluation. Data from each individual participant collected at the different points in time, was then analysed for possible discrepancies, and correlations.

A complete set of statistical measures were compiled to explore the test-retest reliability of the amplitudes of DPOAE and TEOAE responses. This statistical analysis of the data included descriptive statistics such as means and SD. Inferential statistics included one-way repeated measures Analysis of Variance (ANOVA).

3.8.1 Statistical analysis using ANOVA

A test to determine equivalence of variances and means between two or more samples is termed ANOVA or one way analysis of variance. ANOVA is used to determine if observed difference in means can be attributed to natural variation in population (Dallal, 2013).

ANOVA was used to determine the changes in emission amplitudes between the measures M1 to M5. For DPOAE, TEOAE and OAE-gram the mean signals were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Since the measurements were on a continuous scale, reliability was later also evaluated by means of the intraclass correlation coefficient (ICC). The level of significance was $\alpha = 0.05$. (Weir, 2005; Dallal, 2013). Before using one-way repeated measures ANOVA as statistical measure it is important to ensure that ANOVA is an appropriate statistical measure, by testing the data on the following five assumptions (Laerd, 2015):

Assumption 1: The dependent variable should be measured at the interval or ratio level (i.e., they are continuous).

Assumption 2: The independent variable (also known as the within-participants factor) should consist of at least two categorical, "related groups". "Related groups" indicates that the same participants are present in all groups. The reason it is possible to have the same participants in each group in this study is because each participant was measured at two or more different time intervals on the same dependent variable (measures). These repeated measurements (i.e., related groups) are also referred to as levels of the within-subjects factor.

Assumption 3: There should be no significant outliers. All outliers were removed before statistical analysis commenced.

Assumption 4: The dependent variables (in this study the different measures) should be approximately normally distributed for each category of the independent variable. The repeated measures ANOVA only requires approximately normal data because it is quite robust to violations of normality, meaning that the assumption can be a little violated and still provide valid results

Assumption 5: Known as sphericity, the variances of the differences between all combinations of related groups must be equal. The one-way repeated measures ANOVA is particularly susceptible to violating the assumption of sphericity, which causes the test to become too liberal (that is, the likelihood of detecting a statistically significant result when there isn't one).

Since the data collected in this study met all five requirements or assumptions to perform the one-way repeated ANOVA test, it was selected as statistical measure to add value to the data analysis (Field, 2009).

In SPSS 20, the following steps were followed to calculate the ANOVA:

In SPSS Statistics, the groups for analysis were separated by creating a grouping variable called Measures (i.e., the independent variable), and named the different time intervals (levels) M1, M2, M3, M4 and M5.

- Click Analyse then General Linear Model then Repeated Measures
- In the Within-Subject Factor Name box, replace "factor1" with "Measures" because it represents the different times OAE measures were made
- Enter into the Number of Levels box the number of times the dependent variable has been measured. In this case, enter 5, representing the five measures made at five different time intervals. Click the Add button. Put an appropriate name for your dependent variable in the Measure Name box, in this study the different frequencies measured were added. Click the Add button for each frequency.
- Click the Define button. Transfer "M1" M2, M3 M4 and M5 into the Within Subjects Variables box by drag-and-dropping.
- Click the Plots button. Transfer the “Measures” factor from the Factors box into the Horizontal Axis box by drag-and-drop. Click the Add button.
- Click the Continue button. Click the Options button. You will be presented with the following Repeated Measures: Options screen.
- Transfer the factor "Measures" from the Factor(s) and Factor Interactions box to the Display Means for box.
- Tick the Compare main effects checkbox and select "Bonferroni" from the drop-down menu under Confidence interval adjustment

- Tick the Descriptive statistics and Estimates of effect size checkboxes in the Display area.
- Click Continue.
- Click OK.

This produced the desired output for ANOVA in SPSS (Field, 2009).

3.8.2 Statistical analysis using ICC

The ICC was calculated using the DPOAE response, TEOAE response and OAE-gram of each measurement to determine the relative consistency, which represents the consistency of the position on individual results relative to other results. In each case, the estimator is the same whether the interaction effect is present or not. The type A ICC was calculated using an absolute agreement definition and the average measures estimation was computed assuming the interaction effect was absent because it was not estimable otherwise (Weir, 2005; Johnson & Danhauer, 2002).

The ICC was then calculated for each frequency and as comparison the first OAE measurement (M1) was compared to each subsequent measurement made. M2 was within five minutes of the initial measure, M3 within an hour of the initial measure, M4 within a week of the initial measure and M5 within a month of the initial measure.

ICC is an inferential statistic that can be used when quantitative measurements are performed of units that are organized into groups. In this study the individual measurements of each participant are organized into groups according to the participant, and also according to the time at which each measure was taken. The ICC describes how strongly units in the same group resemble each other. While it is viewed as a type of correlation, unlike most other correlation measures it operates on data structured as groups, rather than data structured as paired observations. This gives an accurate indication of the reliability of data collected during this study. The ICC was used to quantify the degree to which each participant's individual DPOAE, TEOAE, and OAE-gram results resembled the results of the same participant measured at other points in time (Weir, 2005).

For the ICC no consensus can be found in the literature regarding a value for a reliable ICC due to the used version and variability of the data (Weir, 2005). Therefore, the validity of the ICC could be compromised by a homogeneous distribution if the between-subjects variability does not reach significance ($\alpha = 0.05$). Due to these possible compromised results the SEM was added to the statistical analysis. The SEM analyses the reliability within repeated measures in a specific subject. The formula used for calculating SEM is $SEM= s\sqrt{(1 -ICC)}$. In this formula "s" is the SD of all measures. Furthermore, the SEM was used to calculate the minimum detectable difference (MDD) which could be considered above the measurement error as a definite change in a participant's results ($MDD95\% = 1.96\sqrt{2}SEM$). All statistical analyses were performed in IBM SPSS Statistics 20.

Quantitative conclusions were drawn from the raw data obtained with the OAE-tests performed at different stages to answer the main aim. The sub-aims assisted in the analysis and interpretation of the data. The raw data was analysed statistically to determine whether conclusive results can be obtained from this study on the test-retest reliability over different time intervals.

The comparison of results from the same participant at different points in time is the most significant contributor towards the findings of this study. However, results from different participants were also compared and analysed to explain specific trends in the results. The results are presented in the form of graphs, tables, figures, and charts to enhance clarity and facilitate interpretation.

In SPSS 20, the following steps were followed to calculate the ICC:

- Click on Statistics, then Scale, then Reliability, click on the Statistics button, and check the checkbox for Intraclass correlation coefficient.
- The structure of the data is as N cases or rows (N in this study being the different participants' results) and k variables or columns (which denote the different measurements of the participants, M1 to M5) The participants are assumed to be a random sample from a larger population, and the ICC estimates are based on mean squares obtained by applying analysis of variance (ANOVA) models to these data. A two-way mixed effects model was chosen, as the rater factor is treated as a fixed factor (because only one rater or researcher was used).
- In the dialog boxes, when the ICC checkbox is checked, a dropdown list is enabled that allows you to specify the appropriate model. If nothing further is specified, the default is

the two way mixed effect model selected for this study. These steps produce the desired output for ICC. The result includes two different ICC estimates: One for the reliability of a single rating, and one for the reliability for the mean or sum of k ratings. Multiple ratings were combined to produce more reliable measurements. The estimates for the reliability of a single rating under the mixed model are the same regardless of whether interactions are assumed, because there is only one rater or researcher present, which is therefore, assumed not to have an effect on results (Field, 2009).

3.9 Summary

In this chapter a comprehensive discussion of the procedure for data collection and data recording was provided. The need for investigating the test-retest reliability of DPOAEs and TEOAEs has become evident through a critical analysis of relevant literature (and lack thereof).

All participants in this study were between the ages of 18 and 25 years. Each participant completed a questionnaire to confirm that their audiological and medical history did not place them at risk for audiological or hearing problems. Each participant also indicated on this questionnaire that they had not been exposed to loud noise in the 48 hours preceding testing. Each participant also agreed to participate on a voluntary basis and signed an informed consent form. When a participant was selected to participate in this study she was assigned a random number between 1 and 40 (although only 30 participants were used) to ensure confidentiality and either her right or left ear was selected as test ear.

Following the conventional audiological test battery, data collection of OAE-measures commenced. Initial OAE measures were performed. The probe was replaced and OAE measures performed again. The participant was then granted an hour break and after an hour OAE measures were repeated again. These measures were again repeated one week and one month after the initial testing took place. At the one week and one month testing, immittance measures were also repeated to ensure that there was no middle ear pathology that could affect the subsequent results.

4. CHAPTER FOUR: RESULTS AND DISCUSSION

In this chapter the results obtained from the data processing are presented and discussed. The presentation of results and discussion are combined into one chapter to enhance the clarity and interpretation of the results.

4.1 Introduction

The results of the research study are discussed according to the research aims that were described in Chapter 3. In order to achieve the main aim to this study (to determine the test-retest reliability of DPOAEs and TEOAEs), variables among participants had to be restricted to a minimum, and the selection criteria are discussed in depth in Chapter 3. Repeated OAE tests were used to investigate the test-retest reliability over time of these measures (D'haenans, et al., 2008; Keppler, et al., 2010).

4.2 Presentation and discussion of results according to sub-aims

In this section the results are presented and discussed to gain a better understanding of the analysed data and statistical measures applied to the data. Results are presented and discussed according to the sub-aims, and subsequently presented and discussed according to the main aim.

DPOAE results, TEOAE results and the OAE-gram were analysed separately in order to compare the test-retest reliability for each type of test at different intervals of measurement and at the different frequencies measured. DPOAE test results, TEAOE test results, and OAE-grams were interpreted and analysed statistically in order to arrive at meaningful conclusions. Several different statistical methods of analysis were used to determine the test-retest reliability at different times and for each of the different frequencies measured.

The ICC is one of the main statistical methods used to analyse and interpret the data. The ICC is a statistical measure that assesses the reliability of ratings by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The ratings are quantitative (Dallal, 2013). Since the measurements were conducted on a continuous scale, reliability was evaluated by means of the ICC. The ICC was calculated using the DPOAE response, TEOAE response, and OAE-gram of each measurement to determine the relative consistency which represents the consistency of individual results relative to other results. In each case, the estimator is the same whether the interaction effect is present or not. The Type A ICCs were calculated using an absolute agreement definition, and the average measures estimation was computed assuming the interaction effect was absent (because it was not estimable otherwise). The level of significance was $\alpha = 0.05$ (Laerd, 2015; Weir, 2005). The ICC was then calculated for each frequency and the first OAE measurement (M1) was compared to each subsequent measurement made. M2 is within five minutes of the initial measure, M3 within an hour of the initial measure, M4 within a week of the initial measure, and M5 within a month of the initial measure.

For the ICC no consensus could be found in the literature regarding a value for a reliable ICC due to the various procedures utilised and variability of the data (Weir, 2005). Therefore, the validity of the ICC could be compromised by a homogeneous distribution if the between-subjects variability does not reach significance ($\alpha = 0.05$). Each frequency of the different tests was analysed individually in order to calculate the reliability of the measurements at different points in time. Using the ICC and α value calculated accordingly, the significance of difference between different measures can be deduced. Significant differences between measures is defined as an α value smaller than 0.05 (Dallal, 2013). An ICC with an α value smaller than 0.05 thus indicates significant statistical differences between measures, whereas a value of α larger than 0.05 indicated no statistical significant differences between measures. (Weir, 2005)

Due to the possibility of compromised results the SEM was added to the statistical analysis. The SEM analyses the reliability within repeated measures in a specific subject. The formula used for calculating SEM is $SEM= s\sqrt(1-ICC)$. In this formula s is the SD of all measures. Furthermore, the SEM was used to calculate the minimum detectable difference (MDD) which allowed the researcher to determine at a level above the measurement error whether there was a definite change in a participant's results. In this case $MDD95\% = 1.96\sqrt{2}SEM$.

The SEM coefficient was then calculated for each frequency and the first OAE measurement (M1) was compared to each subsequent measurement made. M2 was obtained within five minutes of the initial measure, M3 within an hour of the initial measure, M4 within a week of the initial measure and M5 within a month of the initial measure. The results of SEM with 95% confidence interval for each frequency and for each measurement will be discussed for each of the specific sub-aims.

Furthermore, to enhance the statistical results, the amplitudes for TEOAE measures were also analysed with the aid of descriptive statistics. The minimum, maximum, and average values, and the SD between the different amplitudes were calculated for the different measures and different frequencies measured in order to draw conclusions regarding the test-retest reliability between the different tests at all measured frequencies.

Inferential statistics included one-way repeated measures Analysis of Variance (ANOVA). ANOVA was used to determine the changes in emission amplitudes between the measures M1 to M5. For DPOAEs, TEOAEs, and OAE-grams the mean signals were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Further descriptive statistics were calculated for each of the tests to give a clear picture of test-retest reliability. For each frequency the different measurements at the time intervals specified were compared according to the mean, SD, SEM, and the 95% confidence interval for the mean and minimum and maximum values recorded.

All statistical analyses were performed using IBM SPSS Statistics 20 software. Quantitative conclusions were drawn from the raw data obtained with the OAE-tests performed at different stages in order to achieve the main aim. The sub-aims assisted in the analysis and interpretation of the data.

The comparison of results from each individual participant at different points in time for the same individual is the most significant contributor towards the findings of this study. However, results from different participants are also compared and analysed to explain specific trends in the results. The results are presented in the form of graphs, tables, figures, and charts to enhance the interpretation and clarity of the material. The data obtained is analysed statistically according to these sub aims and explained accordingly in order to determine the test-retest reliability in a statistically sound way.

4.2.1 Sub-aim 1: Test-retest reliability of DPOAEs and TEOAEs with short term re-assessment within five minutes of initial OAE tests (M1 – M2).

The test-retest reliability of each test was analysed separately in order to draw statistically valid conclusions about results. DPOAEs, TEOAEs, and the OAE-gram were analysed according to ICC, SD and ANOVA.

These results were calculated using test data from the initial test (M1) compared to immediate re-fitting of the probe (M2). Thirty tests were used for this analysis as 30 test subjects completed the M1 and M2 measures with reliable results not influenced by noise or any other external factors. These results are presented and discussed according to the different tests performed, namely DPOAEs, then TEOAEs and lastly the OAE-gram.

4.2.1.1 DPOAEs M1 – M2

Table 5 provides the ICC data comparing M1 and M2 for all the frequencies measured with the DPOAE tests performed. The ICCs were calculated using an absolute agreement definition. The ICC for both single and average measures for all frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz) was calculated in SPSS. Two-way mixed effects models where participant effects are random and measures effects are fixed were used for this calculation. The estimator for the ICC of a single measure was the same, whether the interaction effect was present or not, and for the average measures of the ICC the estimate was computed assuming the interaction effect was absent, because it is not estimable otherwise.

Table 5: DPOAE ICC comparison of M1 – M2

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.965[a]	0.925	0.983	53.894	27	27	0
	Average Measures	.982[c]	0.961	0.992	53.894	27	27	0
1.5 kHz	Single Measures	.891[a]	0.781	0.948	17.478	27	27	0
	Average Measures	.943[c]	0.877	0.973	17.478	27	27	0
2 kHz	Single Measures	.967[a]	0.931	0.985	58.594	27	27	0
	Average Measures	.983[c]	0.964	0.992	58.594	27	27	0
3 kHz	Single Measures	.963[a]	0.922	0.983	51.195	27	27	0
	Average Measures	.981[c]	0.959	0.991	51.195	27	27	0
4 kHz	Single Measures	.927[a]	0.848	0.966	27.956	27	27	0
	Average Measures	.962[c]	0.918	0.983	27.956	27	27	0
6 kHz	Single Measures	.956[a]	0.908	0.979	45.524	27	27	0
	Average Measures	.977[c]	0.952	0.99	45.524	27	27	0
8 kHz	Single Measures	.944[a]	0.884	0.974	34.066	27	27	0
	Average Measures	.971[c]	0.938	0.987	34.066	27	27	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

The values obtained when comparing the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz between M1 (initial measure) and M2 (measure within five minutes of initial measurement) are summarised in Table 5. *Single measures* results refer to the calculation of ICC between M1 and M2 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M2 respectively across all participants at the two specified points in time. This means that the reliability of the averages of the 27 different measures is estimated by these results.

ICC is an inferential statistic that can be used when quantitative measurements are performed of units that are organized into groups. In this study the individual measurements of each participant are organized into groups according to the participant, and also according to the time at which each measure was taken. The ICC describes how strongly units in the same group resemble each other. While it is viewed as a type of correlation, unlike most other correlation measures it operates on data structured as groups, rather than data structured as paired observations. This gives an accurate indication of the reliability of data collected during this study. The ICC was used to quantify the degree to which each participant's individual DPOAE, TEOAE, and OAE-gram results resembled the results of the same participant measured at other points in time (Weir, 2005).

In simple terms, ICC is an indication of how reliable (or random) the different results for each participant are at the different points in time. An ICC of 0.6 would indicate that 60% of the variability of the results can be explained by the construct, and 40% represented random variation (Weir, 2005).

The ICC for single measures at the different frequencies measured was 0.965 at 1 kHz, 0.891 at 1.5 kHz, 0.967 at 2 kHz, 0.963 at 3 kHz, 0.927 at 4 kHz, 0.956 at 6 kHz and 0.944 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 11% variability between the results that could be described by random variation. More than 89% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1.5 kHz (0.891) and the highest reliability was observed at 2 kHz (0.967).

The ICC for average measures at the different frequencies measured was 0.982 at 1 kHz, 0.943 at 1.5 kHz, 0.983 at 2 kHz, 0.981 at 3 kHz, 0.962 at 4 kHz, 0.977 at 6 kHz and 0.971 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 6% variability between the results that could be described by random variation. More than 94% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 1.5 kHz (0.943) and the highest reliability was observed at 2 kHz (0.983). The difference between the lowest reliability and the highest reliability calculated was not statistically significant (p=0.04), and therefore, this does not indicate a statistical difference in reliability between the different frequencies measured for M1 and M2.

Previous studies (Sockalingham, Kei, & Ho, 2007) have proven the test-retest reliability of DPOAEs with ICC ranging from 0.64 to 0.89, indicating the high test-retest reliability achieved with this study (ranging between 0.891 and 0.983) as positive indicator of statistical significance.

In another study (Wagner, Heppelmann, Vonthein, & Zenner, 2008) the test-retest reliability was confirmed with SEM and confidence interval (CI) at lower levels than indicated by these tests conducted, thus confirming the high test-retest reliability. The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability between measures M1 and M2 during this study, which confirms the test-retest reliability. Therefore, the ICC results obtained in this study indicates high reliability between the measures made five minutes apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz). The ICC ranged between 0.891 and 0.965 for single measures and from 0.943 to 0.983 for average measures. The lowest lower-bound 95% confidence interval was 0.848 at 4 kHz, and the highest upper-bound 95% confidence interval was 0.992 at 1 kHz and 2 kHz.

Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference (Kumar, Methi, & Avinash, 2013). The results obtained by this study indicate that the ICC results above 0.9 for all frequencies measured are higher than this previous study’s results, confirming the high test-retest reliability.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the various frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M2 for the DPOAE results. To further quantify the test-retest reliability of M1 and M2, the TEOAE results were statistically analysed.

4.2.1.2 TEOAEs M1 – M2

Table 6 provides the ICC data comparing M1 and M2 for all the frequencies measured with the TEOAE results.

Table 6: TEOAE ICC comparison of M1 – M2

ICC									
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0				
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.934[a]	0.861	0.969	31.306	27	27	0
	Average Measures	.966[c]	0.926	0.984	31.306	27	27	0
1.5 kHz	Single Measures	.991[a]	0.956	0.997	329.348	27	27	0
	Average Measures	.995[c]	0.978	0.998	329.348	27	27	0
2 kHz	Single Measures	.984[a]	0.948	0.994	162.261	27	27	0
	Average Measures	.992[c]	0.973	0.997	162.261	27	27	0
3 kHz	Single Measures	.989[a]	0.974	0.995	213.394	27	27	0
	Average Measures	.995[c]	0.987	0.998	213.394	27	27	0
4 kHz	Single Measures	.995[a]	0.989	0.998	371.614	27	27	0
	Average Measures	.997[c]	0.994	0.999	371.614	27	27	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 6 the ICC measurements at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz between M1 (initial measure) and M2 (measured within five minutes of initial measurement) are summarised. *Single measures* results refer to the calculation of ICC between M1 and M2 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M2 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 27 different measures is estimated by these results.

The ICC, ranging between .934 and .995 for single measures and .966 and .997 for average measures, indicates high reliability between the measures made five minutes apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz). The lowest lower-bound 95% confidence interval was 0.861 at 1 kHz, and the highest upper-bound 95% confidence interval was 0.999 at 4 kHz.

The ICC for single measures at the different frequencies measured was 0.934 at 1 kHz, 0.991 at 1.5 kHz, 0.984 at 2 kHz, 0.989 at 3 kHz and 0.995 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was less than 7% variability between the results that could be described by random variation. More than 93% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.934) and the highest reliability was observed at 4 kHz (0.995).

The ICC for average measures at the different frequencies measured was 0.966 at 1 kHz, 0.995 at 1.5 kHz, 0.992 at 2 kHz, 0.995 at 3 kHz and 0.997 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was very little variability between the results that could be described by random variation. More than 96% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.966) and the highest reliability was observed at 4 kHz (0.997).

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability at all measurements during this study, which confirms the TEOAE test-retest reliability. Large differences in SD in TEOAEs across different participants were also observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) as in this study, which did not influence overall test-retest reliability.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M2 for the TEOAE results.

Table 7 and Figure 3 further explain the SD with the measured TEOAE amplitudes between M1 and M2 at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz.

Table 7: SD M1 and M2 TEOAE amplitude

Descriptive statistics M1

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	30	-2.50	16.60	7.8433	5.37476
Signal 1.5 kHz	30	-4.20	17.50	10.1100	5.75951
Signal 2 kHz	30	-3.60	18.70	8.3600	5.23895
Signal 3 kHz	30	-3.40	18.30	6.1200	5.14844
Signal 4 kHz	30	-5.80	15.70	3.9633	5.98374
Valid N (listwise)	30				

Descriptive Statistics M2

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	29	-3.20	17.00	8.3138	5.27966
Signal 1.5 kHz	29	-4.80	18.60	10.7034	5.91375
Signal 2 kHz	29	-2.90	18.90	9.1103	5.05713
Signal 3 kHz	29	-1.60	18.50	6.5517	4.93483
Signal 4 kHz	29	-6.80	15.80	4.0172	5.86558
Valid N (listwise)	29				

Table 7 summarises the descriptive statistics used when the amplitude of TEOAEs measured at M1 and the amplitude of TEOAEs measured at M2 were compared to each other. The minimum, maximum, mean, and SD of the amplitude of the TEOAE measures are similar for M1 and M2, indicating high test-retest reliability. The minimum wave amplitude, the maximum wave amplitude, the mean and the SD of the wave amplitudes compared between M1 and M2 showed that the results remained consistent with no statistical significant variability between the different measures. The minimum wave amplitude is the smallest at 4 kHz and the maximum wave amplitude is the largest at 2 kHz in both M1 and M2 measures. The SD is between 4.93 and 5.98 for all frequencies at both M1 and M2. This confirms that the signal level did not differ more than 5.98 dB on average between the TEOAE results at the time intervals measured at M1 and M2 indicating high test-retest reliability and limited external factors influencing results as with previous studies examining these variables (Kumar, Methi, & Avinash, 2013). The descriptive statistics can further be analysed and explained by the SD of measurements as seen in Figure 3.

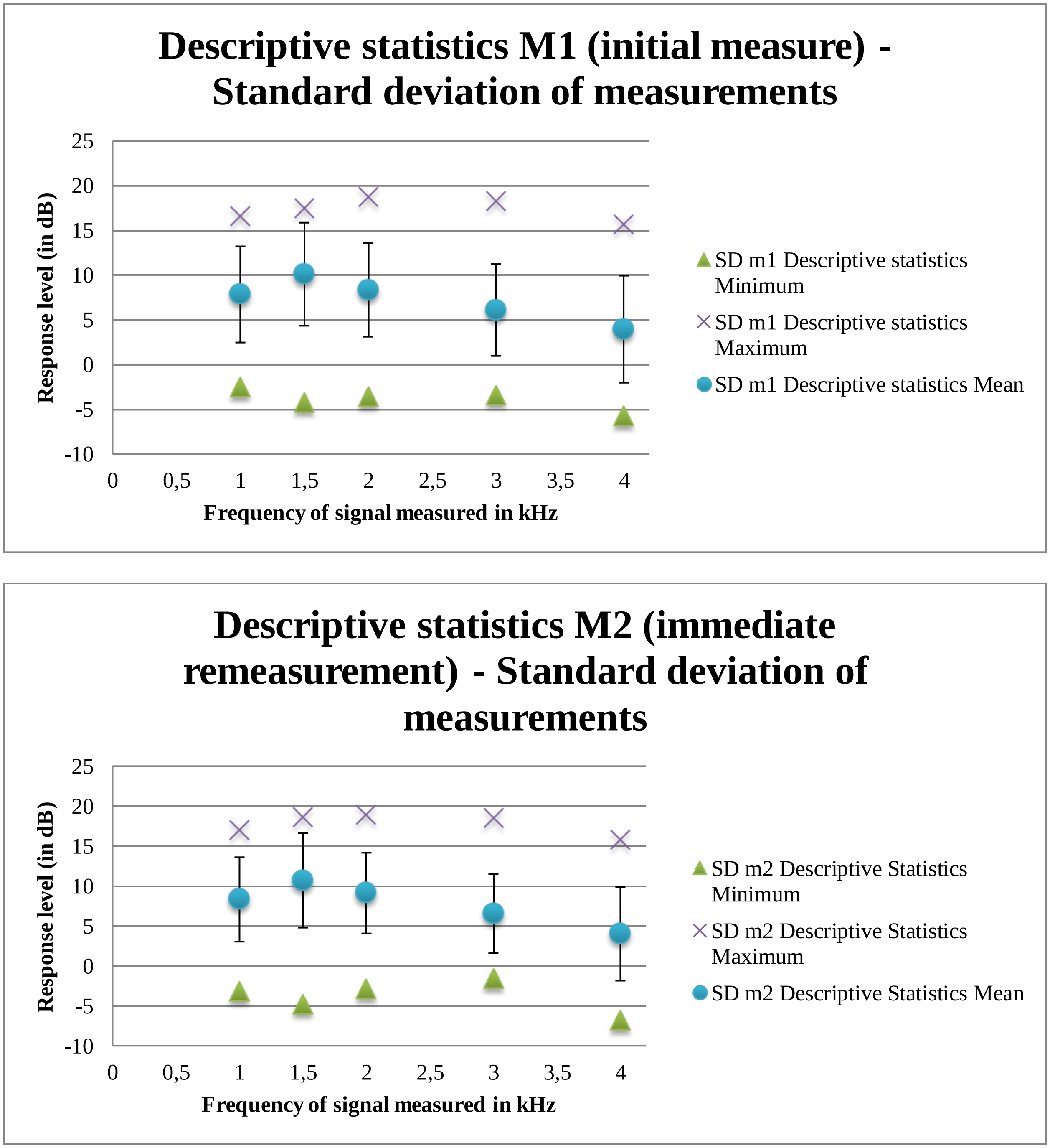


Figure 3: SD of amplitude for measurements of M1 and M2 according to frequency

As illustrated in Figure 3, the SD of amplitude between the different tests M1 and M2 showed very little variance indicating high test-retest reliability between these two tests. The mean response in dB is shown with the error bars indicating the SD observed. The minimum and maximum amplitude response for each frequency is also indicated in the figures.

This visual representation in Figure 3 supports the results summarised in Table 7. The comparison between the mean amplitudes of TEOAEs measured at M1 and M2 show consistency across all frequencies measured, as does the SD and the minimum and maximum amplitudes recorded.

Large differences in SD across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992). It was concluded that a SD of up to 13 dB in previous studies did not influence overall test-retest reliability, which was also the case in this study where differences of only up to 5.9 dB was observed as SD between M1 and M2 confirming test-retest reliability.

Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992; Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. This study confirmed that the SD between M1 and M2 differed only by up to 5.98 dB which was a closer resemblance than previous studies conducted. During this study between M1 and M2 the amplitudes were not reduced at the higher frequencies over longer periods of time as with previous studies indicating even higher test-retest reliability than previous studies conducted. Despite the statistical results obtained with previous studies the test-retest reliability was confirmed, which proves that with more encouraging statistical results obtained in this study, TEOAE test-retest reliability is statistically confirmed between M1 and M2. To further quantify the test-retest reliability of M1 and M2, the OAE-gram results were statistically analysed.

4.2.1.3 OAE-GRAM M1 – M2

Table 8 provides the ICC data comparing M1 and M2 for all the frequencies measured with the OAE-gram calculated.

Table 8: OAE-gram ICC comparison of M1 – M2

ICC

		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.840[a]	0.684	0.922	11.219	27	27	0
	Average Measures	.913[c]	0.812	0.96	11.219	27	27	0
1.5 kHz	Single Measures	.959[a]	0.913	0.981	48.637	27	27	0
	Average Measures	.979[c]	0.954	0.99	48.637	27	27	0
2 kHz	Single Measures	.980[a]	0.954	0.991	107.523	27	27	0
	Average Measures	.990[c]	0.976	0.995	107.523	27	27	0
3 kHz	Single Measures	.972[a]	0.941	0.987	73.714	27	27	0
	Average Measures	.986[c]	0.969	0.993	73.714	27	27	0
4 kHz	Single Measures	.909[a]	0.813	0.957	21.925	27	27	0
	Average Measures	.952[c]	0.897	0.978	21.925	27	27	0
6 kHz	Single Measures	.880[a]	0.76	0.943	15.908	27	27	0
	Average Measures	.936[c]	0.864	0.97	15.908	27	27	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 8 the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, and 6 kHz between M1 (initial measure) and M2 (measure within five minutes of initial measurement) are summarised of the combined DPOAE and TEOAE results according to the algorhythm (OAE-gram). *Single measures* results refer to the calculation of ICC between M1 and M2 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M2 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 27 different measures is estimated by these results.

The ICC which was used to quantify the degree to which each participant's individual DPOAE, TEOAE and OAE-gram results resemble the results of the same participant measured at other points in time (Weir, 2005), ranged from .840 and .980 for single measures and from .913 and .990 for average measures, indicating high reliability between the measures made five minutes apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz and 6 kHz). The lowest lower-bound 95% confidence interval was 0.684 at 1 kHz, and the highest upper-bound 95% confidence interval was 0.993 at 3 kHz.

The ICC for single measures at the different frequencies measured was 0.840 at 1 kHz, 0.959 at 1.5 kHz, 0.980 at 2 kHz, 0.972 at 3 kHz, 0.909 at 4 kHz, 0.880 at 6 kHz. All the ICC results calculated at all the different frequencies for the OAE-gram indicated that there was less than 16% variability between the results that could be described by random variation. More than 84% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.840) and the highest reliability was observed at 2 kHz (0.980).

The ICC for average measures at the different frequencies measured was 0.913 at 1 kHz, 0.979 at 1.5 kHz, 0.990 at 2 kHz, 0.986 at 3 kHz, 0.952 at 4 kHz, 0.936 at 6 kHz. All the ICC results calculated at all the different frequencies for the OAE-gram indicated that there was less than 10% variability between the results that could be described by random variation. More than 90% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 1 kHz (0.913) and the highest reliability was observed at 2 kHz (0.990).

The high ICC for the various frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M2 for the OAE-gram results.

No previous studies were found utilising the OAE-gram and algorhythm as comparison of test-retest reliability. It was included in this study with results resembling the test-retest reliability of the DPOAE and TEOAE statistical significance. It is therefore, concluded that the high ICC applied to the OAE-gram and the similarity of the results to the DPOAE and TEOAE ICC proves the test-retest reliability between M1 and M2 for the OAE-gram.

Test-retest reliability was proven between M1 (initial measure) and M2 (measure within 5 minutes) using various statistical methods and according to several previous studies. The ICC results obtained in this study when comparing M1 to M2 closely resembled results obtained in previous studies to confirm test-retest reliability of DPOAEs and TEOAEs. The same statistical principles were applied to the OAE-gram generated to confirm test-retest reliability. The test-retest reliability of DPOAEs, TEOAEs and the OAE-gram is statistically confirmed between the measures made five minutes apart.

4.2.2 Sub-aim 2: Test-retest reliability of DPOAEs and TEOAES with short term re-assessment one hour after initial OAE test (M1 – M3)

The test-retest reliability of each test was analysed separately in order to draw statistical conclusions about results. DPOAEs, TEOAEs, and the OAE-gram were analysed according to ICC and SD as well as ANOVA.

These results were calculated using test data from the initial test (M1) compared to retest within one hour (M3). Twenty-nine tests were used for this analysis as 29 test subjects completed the M1 and M3 measures with reliable results not influenced by noise or any other external factors. These results are presented and discussed according to the different tests performed, namely DPOAEs, then TEOAEs and lastly the OAE-gram.

4.2.2.1 DPOAEs M1 – M3

Table 9 provides the ICC data comparing M1 and M3 for all the frequencies measured with the DPOAE tests performed. The ICCs were calculated using an absolute agreement definition. The ICC for both single and average measures for all frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz) was calculated in SPSS. Two-way mixed effects models where participant effects are random and measures effects are fixed were used for this calculation. The estimator for the ICC of a single measure was the same, whether the interaction effect was present or not, and for the average measures of the ICC the estimate was computed assuming the interaction effect was absent, because it is not estimable otherwise.

Table 9: DPOAE ICC comparison of M1 – M3

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.310[a]	-0.04	0.598	1.94	28	28	0.043
	Average Measures	.473[c]	-0.084	0.749	1.94	28	28	0.043
1.5 kHz	Single Measures	.202[a]	-0.159	0.52	1.52	28	28	0.137
	Average Measures	.336[c]	-0.377	0.684	1.52	28	28	0.137
2 kHz	Single Measures	.286[a]	-0.071	0.582	1.823	28	28	0.059
	Average Measures	.445[c]	-0.154	0.736	1.823	28	28	0.059
3 kHz	Single Measures	.418[a]	0.071	0.676	2.435	28	28	0.011
	Average Measures	.590[c]	0.132	0.807	2.435	28	28	0.011
4 kHz	Single Measures	.297[a]	-0.065	0.593	1.853	28	28	0.054
	Average Measures	.458[c]	-0.139	0.744	1.853	28	28	0.054
6 kHz	Single Measures	.916[a]	0.83	0.96	22.184	28	28	0
	Average Measures	.956[c]	0.907	0.98	22.184	28	28	0
8 kHz	Single Measures	.916[a]	0.83	0.959	22.821	28	28	0
	Average Measures	.956[c]	0.907	0.979	22.821	28	28	0

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In

Table 9 the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz and 8 kHz between M1 (initial measure) and M3 (measure within one hour of initial measurement) are summarised. *Single measures* results refer to the calculation of ICC between M1 and M3 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M3 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 28 different measures is estimated by these results.

The ICC, which ranged between .202 and .916 for single measures and between .336 and .959 for average measures, indicates high reliability between the measures made one hour apart at some 6 kHz and 8 kHz. The lowest lower-bound 95% confidence interval was -0.377 at 1.5 kHz, and the highest upper-bound 95% confidence interval was 0.98 at 6 kHz.

The ICC for single measures at the different frequencies measured was 0.310 at 1 kHz, 0.202 at 1.5 kHz, 0.286 at 2 kHz, 0.418 at 3 kHz, 0.297 at 4 kHz, 0.916 at 6 kHz and 0.916 at 8 kHz. The ICC results calculated at 6 kHz and 8 kHz for the DPOAEs indicated that there was very little variability between the results that could be described by random variation. At the lower frequencies (1 kHz, 1,5 kHz, 2 kHz, 3 kHz and 4 kHz) lower reliability was indicated by the ICC. The lowest reliability was observed at 1.5 kHz (0.202) and the highest reliability was observed at 6 and 8 kHz (0.916). The lower reliability at lower frequencies could possibly be explained by higher ambient noise present at lower frequencies.

The ICC for average measures at the different frequencies measured was 0.473 at 1 kHz, 0.336 at 1.5 kHz, 0.445 at 2 kHz, 0.590 at 3 kHz, 0.458 at 4 kHz, 0.956 at 6 kHz and 0.956 at 8 kHz. The ICC results calculated at 6 kHz and 8 kHz for the DPOAEs indicated that there was very little variability between the results that could be described by random variation. At the lower frequencies (1 kHz, 1,5 kHz, 2 kHz, 3 kHz and 4 kHz) lower reliability was indicated by the ICC. The lowest reliability was observed at 1.5 kHz (0.336) and the highest reliability was observed at 6 and 8 kHz (0.956). The lower reliability at lower frequencies could possibly be explained by higher ambient noise present at lower frequencies (Keppler, et al., 2010).

Previous studies (Sockalingham, Kei, & Ho, 2007) have proven the test-retest reliability of DPOAEs with ICC ranging from 0.64 to 0.89, indicating the high test-retest reliability achieved with this study as positive indicator of statistical significance. In another study (Wagner, Heppelmann, Vonthein, & Zenner, 2008) the test-retest reliability was confirmed with SEM and confidence interval (CI) at lower levels than indicated by these tests conducted, thus confirming the high test-retest reliability. The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability between measures M1 and M3 during this study, which confirms the test-retest reliability.

Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference by recent studies (Kumar, Methi, & Avinash, 2013). The results obtained by this study indicate that the ICC results above 0.91 for 6 kHz to 8 kHz measured are higher than previous studies' results, confirming the high test-retest reliability.

The high ICC for all the high frequencies (6 kHz and 8 kHz) measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability for the 6 kHz and 8 kHz between M1 and M3 for the DPOAE results. To further quantify the test-retest reliability of M1 and M3, the TEOAE results were statistically analysed.

4.2.2.2 TEOAEs M1 – M3

Table 10 provides the ICC data comparing M1 and M3 for all the frequencies measured with the TEOAE results.

Table 10: TEOAE ICC comparison of M1 – M3

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.940[a]	0.877	0.971	31.402	28	28	0
	Average Measures	.969[c]	0.934	0.985	31.402	28	28	0
1.5 kHz	Single Measures	.974[a]	0.946	0.988	75.105	28	28	0
	Average Measures	.987[c]	0.972	0.994	75.105	28	28	0
2 kHz	Single Measures	.967[a]	0.931	0.984	57.623	28	28	0
	Average Measures	.983[c]	0.964	0.992	57.623	28	28	0
3 kHz	Single Measures	.983[a]	0.963	0.992	110.759	28	28	0
	Average Measures	.991[c]	0.981	0.996	110.759	28	28	0
4 kHz	Single Measures	.980[a]	0.956	0.991	112.973	28	28	0
	Average Measures	.990[c]	0.977	0.995	112.973	28	28	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 10 summarises the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz between M1 (initial measure) and M3 (measure within one hour of initial measurement). *Single measures* results refer to the calculation of ICC between M1 and M3 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M3 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 28 different measures is estimated by these results.

The ICC for single measures at the different frequencies measured was 0.940 at 1 kHz, 0.974 at 1.5 kHz, 0.967 at 2 kHz, 0.983 at 3 kHz and 0.980 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was less than 6% variability between the results that could be described by random variation. More than 94% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.940) and the highest reliability was observed at 4 kHz (0.980).

The ICC for average measures at the different frequencies measured was 0.969 at 1 kHz, 0.987 at 1.5 kHz, 0.983 at 2 kHz, 0.991 at 3 kHz and 0.990 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was very little variability between the results that could be described by random variation. More than 96% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.969) and the highest reliability was observed at 3 kHz (0.991).

The ICC, which gives an accurate indication of the reliability of data collected during this study, ranged between .940 and .980 for single measures and between .969 and .990 for average measures, indicating high reliability between the measures made five minutes apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz). The lowest lower-bound 95% confidence interval was 0.871 at 1 kHz, and the highest upper-bound 95% confidence interval was 0.981 at 4 kHz.

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability at all measurements during this study, which confirms the TEOAE test-retest reliability. Large differences in SD in TEOAEs across different participants were also observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) as in this study, which did not influence overall test-retest reliability between M1 and M3.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M3 for the TEOAE results.

Table 11

Table 11 and Figure 4 illustrate the SD with the measured TEOAE amplitude between M1 and M3

Table 11: SD TEOAE M1 and M3 amplitude

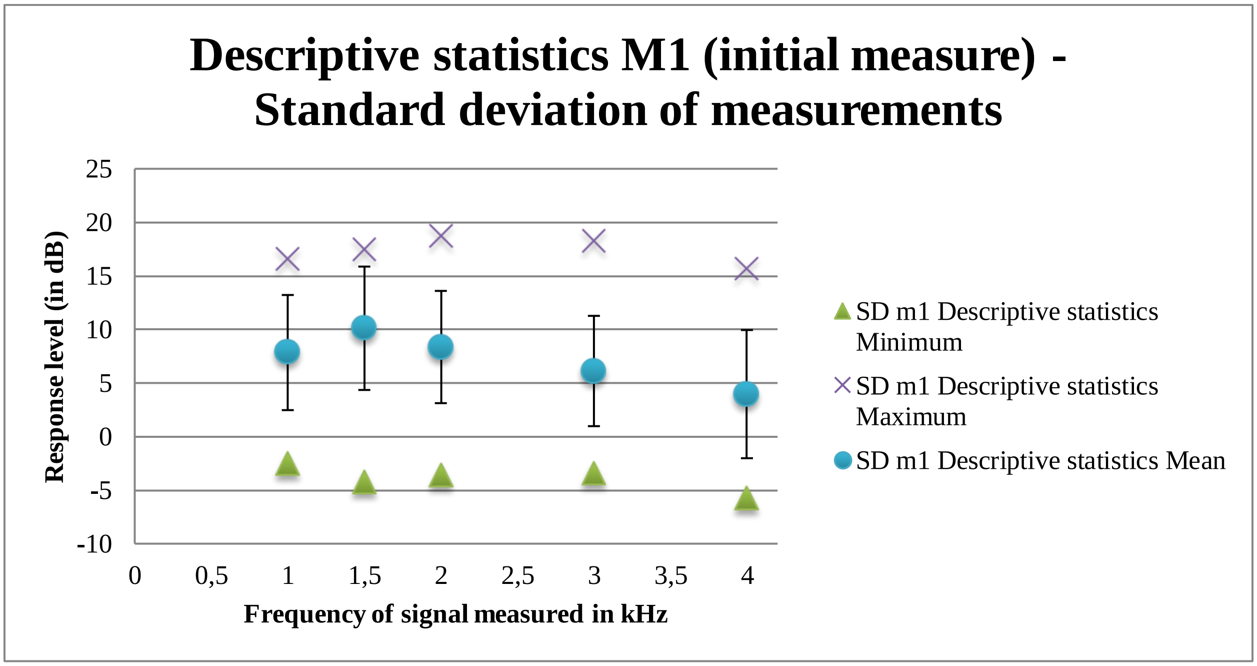
Descriptive statistics M1

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	30	-2.50	16.60	7.8433	5.37476
Signal 1.5 kHz	30	-4.20	17.50	10.1100	5.75951
Signal 2 kHz	30	-3.60	18.70	8.3600	5.23895
Signal 3 kHz	30	-3.40	18.30	6.1200	5.14844
Signal 4 kHz	30	-5.80	15.70	3.9633	5.98374
Valid N (listwise)	30				

Descriptive Statistics M3

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	29	-3.40	17.30	8.1966	5.35280
Signal 1.5 kHz	29	-4.80	20.40	10.3828	6.08488
Signal 2 kHz	29	-3.00	18.00	8.6000	5.19622
Signal 3 kHz	29	-2.10	18.00	6.0655	5.18316
Signal 4 kHz	29	-4.30	15.50	4.4276	5.92301
Valid N (listwise)	29				

Table 11 summarises the descriptive statistics used when the amplitude of TEOAEs measured at M1 is compared to the amplitude of TEOAEs measured at M3. The minimum, maximum, mean, and SD of the amplitude of the TEOAE measures are similar for M1 and M3, indicating high test-retest reliability. The minimum wave amplitude is the smallest at 4 kHz and the maximum wave amplitude is the largest at 2 kHz for M1, and the minimum wave amplitude is the smallest at 1.5 kHz and the maximum wave amplitude is the largest at 1.5 kHz for M3. The SD is between and 5.14 and 6.08 for all the different frequencies at both M1 and M3.



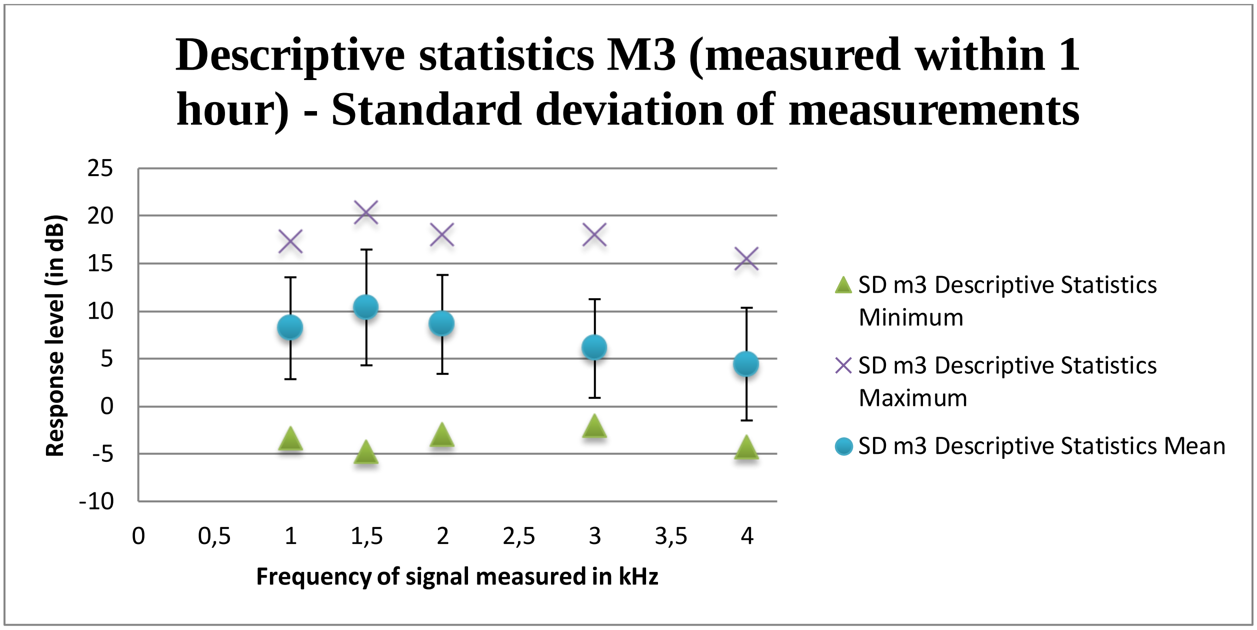


Figure 4: SD of amplitude for measurements of M1 and M3 according to frequency.

As illustrated in Figure 4, the SD of amplitude between the tests M1 and M3 showed very little variance indicating high test-retest reliability between these two tests. The mean response in dB is shown with the error bars indicating the SD observed. The minimum and maximum amplitude response for each frequency is also indicated in the figures.

This visual representation in Figure 4 supports the results summarised in

Table 11. The comparison between the mean amplitudes of TEOAEs measured at M1 and M3 show consistency across all frequencies measured, as does the SD and the minimum and maximum amplitudes recorded.

Large differences in SD across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992). It was concluded that the SD of up to 13dB in previous studies did not influence overall test-retest reliability, which was not the case in this study where differences of only up to 6.08 dB was observed as SD.

Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992; Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. This study confirmed that the SD between M1 and M3 differed only by up to 6.08 dB which was a closer resemblance than previous studies conducted. During this study between M1 and M3 the amplitudes were not reduced at the higher frequencies over longer periods of time as with previous studies indicating even higher test-retest reliability than previous studies conducted. Despite the statistical results obtained with previous studies the test-retest reliability was confirmed, which proves that with more encouraging statistical results obtained in this study, TEOAE test-retest reliability is statistically proven between M1 and M3. To further quantify the test-retest reliability of M1 and M3, the OAE-gram results were statistically analysed.

4.2.2.3 OAE-GRAM M1 – M3

Table 12 provides the ICC data comparing M1 and M3 for all the frequencies measured with the OAE-gram calculated.

Table 12: OAE-gram ICC comparison of M1 - M3

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.350[a]	-0.01	0.631	2.075	28	28	0.029
	Average Measures	.519[c]	-0.02	0.773	2.075	28	28	0.029
1.5 kHz	Single Measures	.402[a]	0.064	0.662	2.42	28	28	0.011
	Average Measures	.573[c]	0.12	0.796	2.42	28	28	0.011
2 kHz	Single Measures	.418[a]	0.074	0.675	2.452	28	28	0.01
	Average Measures	.590[c]	0.138	0.806	2.452	28	28	0.01
3 kHz	Single Measures	.524[a]	0.208	0.743	3.233	28	28	0.001
	Average Measures	.688[c]	0.344	0.852	3.233	28	28	0.001
4 kHz	Single Measures	.564	0.259	0.768	3.584	28	28	0.001
	Average Measures	.721	0.411	0.869	3.584	28	28	0.001
6 kHz	Single Measures	.874	0.749	0.938	14.483	28	28	0.001
	Average Measures	.932	0.857	0.968	14.483	28	28	0.001

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 12 the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, and 6 kHz between M1 (initial measure) and M3 (measure within one hour of initial measurement) are summarised of the combined DPOAE and TEOAE results according to the algorhythm. *Single measures* results refer to the calculation of ICC between M1 and M3 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average*

measures reflect the average measure calculated for M1 and M3 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 28 different measures is estimated by these results.

The ICC ranging between .350 and .874 for single measures and .519 and .932 for average measures indicates high reliability between some of the measures made one hour apart at some of the measured frequencies (3 kHz, 4 kHz, and 6 kHz). The lowest lower-bound 95% confidence interval was -0.02 at 1 kHz, and the highest upper-bound 95% confidence interval was 0.968 at 6 kHz.

The ICC for single measures at the different frequencies measured was 0.350 at 1 kHz, 0.402 at 1.5 kHz, 0.418 at 2 kHz, 0.524 at 3 kHz, 0.564 at 4 kHz, 0.874 at 6 kHz. The lowest reliability was observed at 1 kHz (0.350) and the highest reliability was observed at 6 kHz (0.874).

The ICC for average measures at the different frequencies measured was 0.519 at 1 kHz, 0.573 at 1.5 kHz, 0.590 at 2 kHz, 0.688 at 3 kHz, 0.721 at 4 kHz, 0.932 at 6 kHz. The ICC results calculated at 3 kHz, 4 kHz and 6 kHz for the OAE-gram indicated that there was less than 32% variability between the results that could be described by random variation. More than 68% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 1 kHz (0.519) and the highest reliability was observed at 6 kHz (0.932).

The high ICC for the higher frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at the higher frequencies between M1 and M3 for the OAE-gram results.

No previous studies were found utilising the OAE-gram and algorhythm as comparison of test-retest reliability. It was included in this study with results resembling the test-retest reliability of the DPOAE and TEOAE statistical significance. It is therefore, concluded that the high ICC applied to the OAE-gram and the similarity to the DPOAE and TEOAE ICC proves the test-retest reliability between M1 and M3 for the OAE-gram.

Test-retest reliability was proven between M1 (initial measure) and M3 (measure within one hour of initial measure) using various statistical methods and according to several previous studies. The ICC results obtained in this study when comparing M1 to M3 closely resembled results obtained in previous studies to confirm test-retest reliability of DPOAEs and TEOAEs. The lower frequencies of the DPOAE results did not correlate well with results at the higher frequencies, which could be explained by presence of lower frequency noise while performing measures. The same statistical principles were applied to the OAE-gram generated to confirm test-retest reliability. The test-retest reliability of DPOAEs, TEOAEs and the OAE-gram is statistically confirmed between the measures made one hour apart.

4.2.3 Sub-aim 3: Test-retest reliability of DPOAEs and TEOAES with long term re-assessment one week after initial OAE test (M1 – M4)

The test-retest reliability of each test was analysed separately in order to draw statistically valid conclusions about results. DPOAEs, TEOAEs and the OAE-gram were analysed according to ICC, SD, and ANOVA.

These results were calculated using test data from the initial test (M1) compared to retest within one week (M4). Twelve tests were used for this analysis as 12 test subjects completed the M1 and M4 measures with reliable results not influenced by noise or any other external factors. These results are presented and discussed according to the different tests performed, namely DPOAEs, then TEOAEs and lastly the OAE-gram.

4.2.3.1 DPOAEs M1 – M4

Table 13 provides a summary for all the frequencies measured with the DPOAE tests performed. The ICCs were calculated using an absolute agreement definition. The ICC for both single and average measures for all frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz) was calculated in SPSS. Two-way mixed effects models where participant effects are random and measures effects are fixed were used for this calculation. The estimator for the ICC of a single measure was the same, whether the interaction effect was present or not, and for the average measures of the ICC the estimate was computed assuming the interaction effect was absent, because it is not estimable otherwise. Table 13 provides the ICC data comparing M1 and M4 for all the frequencies measured with the DPOAE tests performed.

Table 13: DPOAE ICC comparison of M1 - M4

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.953[a]	0.813	0.987	53.533	11	11	0
	Average Measures	.976[c]	0.897	0.993	53.533	11	11	0
1.5 kHz	Single Measures	.989[a]	0.965	0.997	187.098	11	11	0
	Average Measures	.995[c]	0.982	0.998	187.098	11	11	0
2 kHz	Single Measures	.913[a]	0.732	0.974	20.811	11	11	0
	Average Measures	.955[c]	0.845	0.987	20.811	11	11	0
3 kHz	Single Measures	.920[a]	0.753	0.976	22.912	11	11	0
	Average Measures	.958[c]	0.859	0.988	22.912	11	11	0
4 kHz	Single Measures	.914[a]	0.741	0.974	22.447	11	11	0
	Average Measures	.955[c]	0.851	0.987	22.447	11	11	0
6 kHz	Single Measures	.974[a]	0.915	0.992	74.984	11	11	0
	Average Measures	.987[c]	0.955	0.996	74.984	11	11	0
8 kHz	Single Measures	.874[a]	0.617	0.962	13.743	11	11	0
	Average Measures	.933[c]	0.763	0.981	13.743	11	11	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 13 summarises the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz between M1 (initial measure) and M4 (measure within one week of initial measurement). *Single measures* results refer to the calculation of ICC between M1 and M4 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M4 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 11 different measures is estimated by these results.

The ICC for single measures at the different frequencies measured was 0.953 at 1 kHz, 0.989 at 1.5 kHz, 0.913 at 2 kHz, 0.920 at 3 kHz, 0.914 at 4 kHz, 0.974 at 6 kHz and 0.874 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 13% variability between the results that could be described by random variation. More than 87% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 8 kHz (0.874) and the highest reliability was observed at 1 kHz (0.989).

The ICC for average measures at the different frequencies measured was 0.976 at 1 kHz, 0.995 at 1.5 kHz, 0.955 at 2 kHz, 0.958 at 3 kHz, 0.955 at 4 kHz, 0.987 at 6 kHz and 0.933 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 7% variability between the results that could be described by random variation. More than 93% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 8 kHz (0.933) and the highest reliability was observed at 1.5 kHz (0.995).

The ICC, which ranged between .874 and .989 for single measures and between .933 and .995 for average measures, indicates high reliability between the measures made one week apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz). The lowest lower-bound 95% confidence interval was 0.617 at 1.5 kHz, and the highest upper-bound 95% confidence interval was 0.998 at 1.5 kHz.

Previous studies (Sockalingham, Kei, & Ho, 2007) have proven the test-retest reliability of DPOAEs with ICC ranging from 0.64 to 0.89, indicating the high test-retest reliability achieved with this study as positive indicator of statistical significance. In another study (Wagner, Heppelmann, Vonthein, & Zenner, 2008) the test-retest reliability was confirmed with SEM and confidence interval (CI) at lower levels than indicated by these tests conducted, thus confirming the high test-retest reliability. The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability between measures M1 and M4 during this study, which confirms the test-retest reliability.

Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference by recent studies (Kumar, Methi, & Avinash, 2013). The results obtained by this study indicate that the ICC results above 0.87 for all the different frequencies measured are higher than previous studies' results, confirming the high test-retest reliability.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the various frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M4 for the DPOAE results.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M4 for the DPOAE results. To further quantify the test-retest reliability of M1 and M4, the TEOAE results were statistically analysed.

4.2.3.2 TEOAEs M1 – M4

Table 14 provides the ICC data comparing M1 and M4 for all the frequencies measured with the TEOAE results.

Table 14: TEOAE ICC comparison of M1 – M4

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.992[a]	0.971	0.998	219.983	11	11	0
	Average Measures	.996[c]	0.986	0.999	219.983	11	11	0
1.5 kHz	Single Measures	.955[a]	0.855	0.987	41.202	11	11	0
	Average Measures	.977[c]	0.922	0.993	41.202	11	11	0
2 kHz	Single Measures	.943[a]	0.82	0.983	32.674	11	11	0
	Average Measures	.971[c]	0.901	0.991	32.674	11	11	0
3 kHz	Single Measures	.977[a]	0.923	0.993	80.179	11	11	0
	Average Measures	.989[c]	0.96	0.997	80.179	11	11	0
4 kHz	Single Measures	.992[a]	0.974	0.998	250.807	11	11	0
	Average Measures	.996[c]	0.987	0.999	250.807	11	11	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 14 summarises the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz between M1 (initial measure) and M4 (measure within one week of initial measurement). *Single measures* results refer to the calculation of ICC between M1 and M4 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M4 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 11 different measures is estimated by these results.

The ICC for single measures at the different frequencies measured was 0.992 at 1 kHz, 0.955 at 1.5 kHz, 0.943 at 2 kHz, 0.977 at 3 kHz and 0.992 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was less than 6% variability between the results that could be described by random variation. More than 94% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 2 kHz (0.943) and the highest reliability was observed at 4 kHz (0.992).

The ICC for average measures at the different frequencies measured was 0.996 at 1 kHz, 0.977 at 1.5 kHz, 0.971 at 2 kHz, 0.989 at 3 kHz and 0.996 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was less than 3% variability between the results that could be described by random variation. More than 97% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 2 kHz (0.971) and the highest reliability was observed at 1 kHz and 4 kHz (0.996).

The ICC ranging between .943 and .992 for single measures and between .971 and .996 for average measures indicate high reliability between the measures made five minutes apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz). The lowest lower-bound 95% confidence interval was 0.820 at 2 kHz, and the highest upper-bound 95% confidence interval was 0.999 at 1 and 4 kHz.

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability at all measurements during this study, which confirms the test-retest reliability. Large differences in SD in TEOAEs across different participants were also observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) as in this study, which did not influence overall test-retest reliability. This study (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and other studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. The results of this study confirmed that the amplitudes were not reduced at the higher frequencies over longer periods of time indicating even higher test-retest reliability than previous studies conducted. With these statistical results the test-retest reliability was confirmed with previous studies, which proves that with more encouraging statistical results obtained in this study, confidence in TEOAEs over time between M1 and M4 is warranted.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M4 for the TEOAE results.

Table 15: TEOAE SD M1 and M4 amplitude

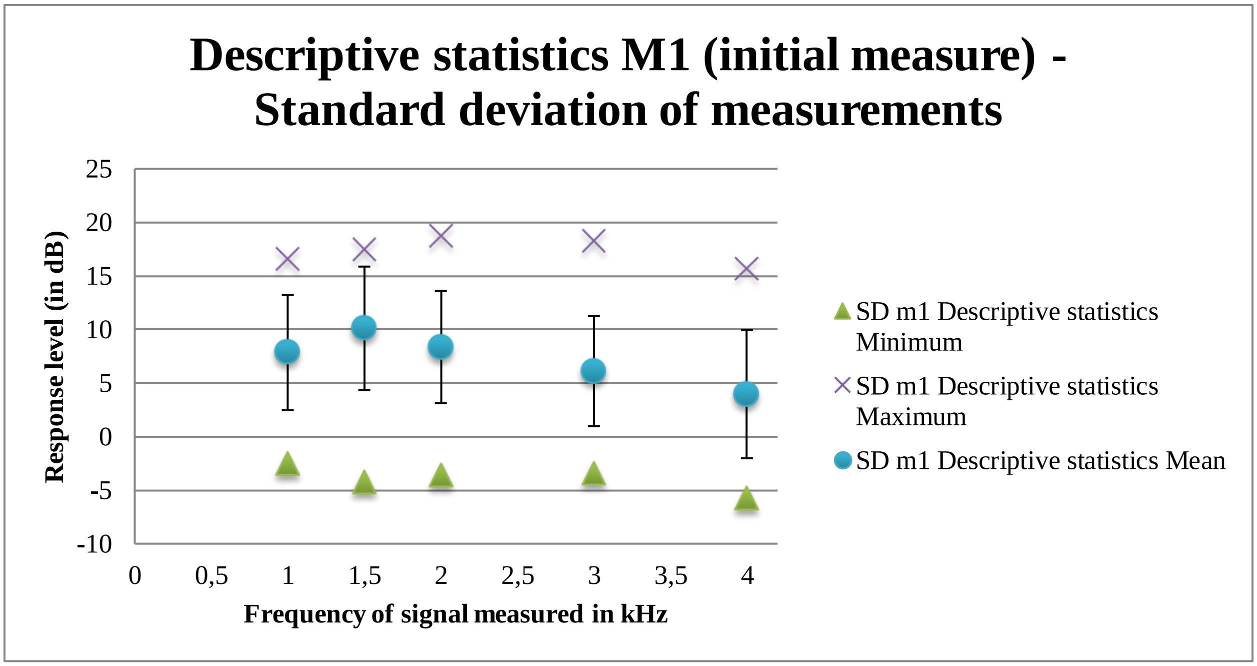
Descriptive statistics M1

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	30	-2.50	16.60	7.8433	5.37476
Signal 1.5 kHz	30	-4.20	17.50	10.1100	5.75951
Signal 2 kHz	30	-3.60	18.70	8.3600	5.23895
Signal 3 kHz	30	-3.40	18.30	6.1200	5.14844
Signal 4 kHz	30	-5.80	15.70	3.9633	5.98374
Valid N (listwise)	30				

Descriptive Statistics M4

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	12	-3.60	15.60	8.1583	5.81651
Signal 1.5 kHz	12	-5.40	16.00	8.3000	6.48915
Signal 2 kHz	12	2.50	18.10	9.2417	5.05056
Signal 3 kHz	12	-.20	17.30	7.4833	5.41023
Signal 4 kHz	12	-4.40	15.30	4.5833	5.54745
Valid N (listwise)	12				

Table 15 portrays the descriptive statistics used when the amplitude of TEOAEs measured at M1 is compared to the amplitude of TEOAEs measured at M4. The minimum, maximum, mean, and SD of the amplitude of the TEOAE measures for M1 and M4 are comparable, indicating high test-retest reliability. The minimum wave amplitude is the smallest at 4 kHz and the maximum wave amplitude is the largest at 2 kHz in M1, while the minimum wave amplitude is the smallest at 1.5 kHz and the maximum wave amplitude is the largest at 2 kHz for M4 measures. The SD is between and 5.05 and 6.48 for all the different frequencies between M1 and M4.



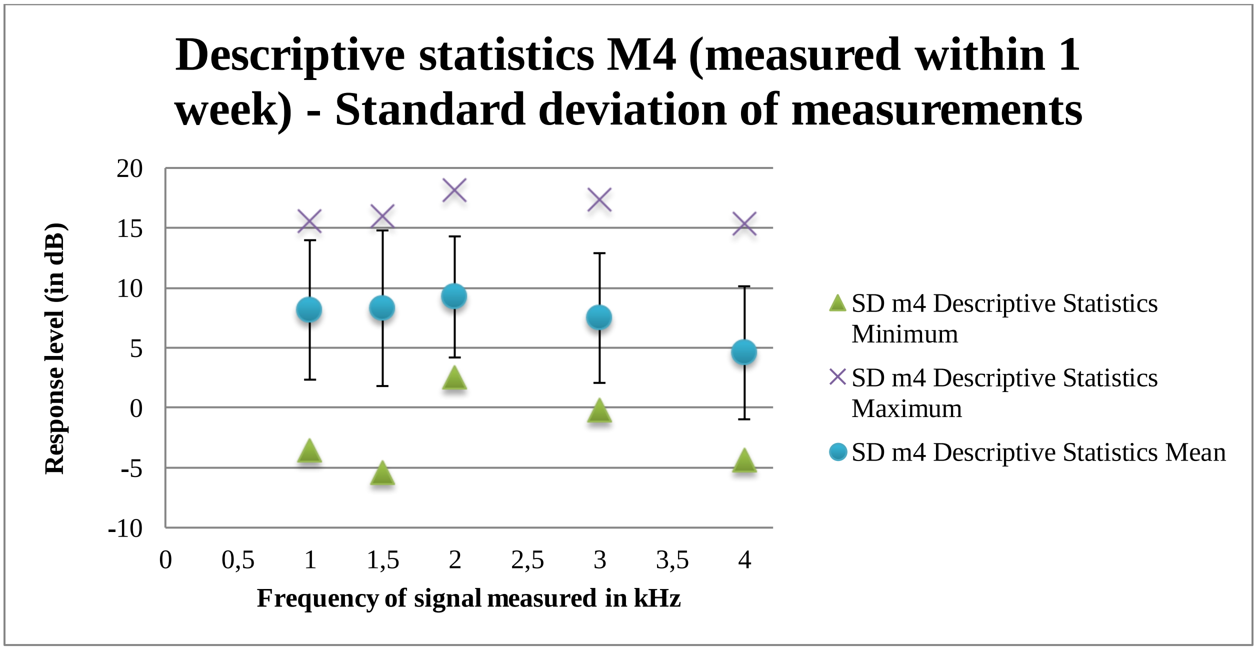


Figure 5: SD of amplitude for measurements of M1 and M4 according to frequency

As illustrated in Table 15 and Figure 5, the SD of amplitude between the tests M1 and M4 showed very little variance, indicating high test-retest reliability between these two tests. The mean response in dB is shown with the error bars indicating the SD observed. The minimum and maximum amplitude response for each frequency is also indicated in the figures.

This visual representation in Figure 5 supports the results summarised in Table 15. The comparison between the mean amplitudes of TEOAEs measured at M1 and M4 show consistency across all frequencies measured, as does the SD and the minimum and maximum amplitudes recorded.

Large differences in SD across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992). It was concluded that the SD of up to 13dB in previous studies did not influence overall test-retest reliability, which was not the case in this study where differences of only up to 5.9dB was observed as SD.

Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and more recent studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. This study confirmed that the SD between M1 and M2 differed only by up to 5.98dB which was a closer resemblance than previous studies conducted. During this study between M1 and M4 the amplitudes were not reduced at the higher frequencies over longer periods of time as with previous studies indicating even higher test-retest reliability than previous studies conducted. Despite the statistical results obtained with previous studies the test-retest reliability was confirmed, which proves that with more encouraging statistical results obtained in this study, TEOAE test-retest reliability is statistically proven between M1 and M4. To further quantify the test-retest reliability of M1 and M4, the OAE-gram results were statistically analysed.

4.2.3.3 OAE-GRAM M1 – M4

Table 16 provides the ICC data comparing M1 and M4 for all the frequencies measured with the OAE-gram calculated.

Table 16: OAE-gram ICC comparison of M1 – M4

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.943[a]	0.82	0.983	35.449	11	11	0
	Average Measures	.970[c]	0.901	0.991	35.449	11	11	0
1.5 kHz	Single Measures	.424[a]	-0.14	0.788	2.501	11	11	0.072
	Average Measures	.596[c]	-0.325	0.882	2.501	11	11	0.072
2 kHz	Single Measures	.410[a]	-0.22	0.789	2.291	11	11	0.092
	Average Measures	.581[c]	-0.564	0.882	2.291	11	11	0.092
3 kHz	Single Measures	.852[a]	0.573	0.955	11.988	11	11	0
	Average Measures	.920[c]	0.728	0.977	11.988	11	11	0
4 kHz	Single Measures	.963[a]	0.88	0.989	51.579	11	11	0
	Average Measures	.981[c]	0.936	0.994	51.579	11	11	0
6 kHz	Single Measures	.924[a]	0.76	0.978	23.649	11	11	0
	Average Measures	.961[c]	0.864	0.989	23.649	11	11	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 16 the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, and 6 kHz between M1 (initial measure) and M4 (measure within one week of initial measurement) are summarised of the combined DPOAE and TEOAE results according to the algorhythm. *Single measures* results refer to the calculation of ICC between M1 and M4 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average*

measures reflect the average measure calculated for M1 and M4 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 11 different measures is estimated by these results.

The ICC, which ranges between .410 and .963 for single measures and between .581 and .981 for average measures, indicates high reliability between the measures made one week apart at the measured frequencies 1 kHz, 3 kHz, 4 kHz, and 6 kHz, with slightly lower results at 1.5 kHz and 2 kHz. The lowest lower-bound 95% confidence interval was 0.564 at 2 kHz, and the highest upper-bound 95% confidence interval was 0.994 at 4 kHz.

The ICC for single measures at the different frequencies measured was 0.943 at 1 kHz, 0.424 at 1.5 kHz, 0.410 at 2 kHz, 0.852 at 3 kHz, 0.963 at 4 kHz, 0.924 at 6 kHz. All the ICC results calculated at 1 kHz, 3 kHz, 4 kHz, and 6 kHz for the OAE-gram indicated that there was less than 15% variability between the results that could be described by random variation. More than 85% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 2 kHz (0.410) and the highest reliability was observed at 4 kHz (0.963).

The ICC for average measures at the different frequencies measured was 0.970 at 1 kHz, 0.596 at 1.5 kHz, 0.581 at 2 kHz, 0.920 at 3 kHz, 0.981 at 4 kHz, 0.961 at 6 kHz. All the ICC results calculated at 1 kHz, 3 kHz, 4 kHz, and 6 kHz for the OAE-gram indicated that there was less than 8% variability between the results that could be described by random variation. More than

92% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 2 kHz (0.581) and the highest reliability was observed at 4 kHz (0.981).

The high ICC for 1 kHz, 3 kHz, 4 kHz, and 6 kHz remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all these frequencies between M1 and M4 for the OAE-gram results.

No previous studies were found utilising the OAE-gram and algorhythm as comparison of test-retest reliability. It was included in this study with results resembling the test-retest reliability of the DPOAE and TEOAE statistical significance. It is therefore, concluded that the high ICC applied to the OAE-gram and the similarity to the DPOAE and TEOAE ICC proves the test-retest reliability between M1 and M4 for the OAE-gram.

Test-retest reliability was proven between M1 (initial measure) and M4 (measure within one week of initial measure) using various statistical methods and according to several previous studies. The ICC results obtained in this study when comparing M1 to M4 closely resembled results obtained in previous studies to confirm test-retest reliability of DPOAEs and TEOAEs. The same statistical principles were applied to the OAE-gram generated to confirm test-retest reliability. The test-retest reliability of DPOAEs, TEOAEs and the OAE-gram is statistically confirmed between the measures made one week apart.

4.2.4 Sub-aim 4: Test-retest reliability of DPOAEs and TEOAES with long term re-assessment one month after initial OAE test (M1 – M5)

The test-retest reliability of each test was analysed separately in order to draw statistically acceptable conclusions about results. DPOAEs, TEOAEs, and the OAE-gram were analysed according to ICC, SD, and ANOVA.

These results were calculated using test data from the initial test (M1) compared to retest within one month (M5). Eighteen tests were used for this analysis as 18 test subjects completed the M1 and M5 measures with reliable results not influenced by noise or any other external factors. These results are presented and discussed according to the different tests performed, namely DPOAEs, then TEOAEs and lastly the OAE-gram.

4.2.4.1 DPOAEs M1 – M5

Table 17 provides the ICC data comparing M1 and M5 for all the frequencies measured with the DPOAE tests performed. The ICCs were calculated using an absolute agreement definition. The ICC for both single and average measures for all frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz) was calculated in SPSS. Two-way mixed effects models where participant effects are random and measures effects are fixed were used for this calculation. The estimator for the ICC of a single measure was the same, whether the interaction effect was present or not, and for the average measures of the ICC the estimate was computed assuming the interaction effect was absent, because it is not estimable otherwise.

Table 17: DPOAE ICC comparison of M1 – M5

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.897[a]	0.739	0.962	17.512	16	16	0
	Average Measures	.946[c]	0.85	0.98	17.512	16	16	0
1.5 kHz	Single Measures	.955[a]	0.88	0.983	46.847	16	16	0
	Average Measures	.977[c]	0.936	0.992	46.847	16	16	0
2 kHz	Single Measures	.968[a]	0.916	0.988	63.579	16	16	0
	Average Measures	.984[c]	0.956	0.994	63.579	16	16	0
3 kHz	Single Measures	.932[a]	0.823	0.975	26.964	16	16	0
	Average Measures	.965[c]	0.903	0.987	26.964	16	16	0
4 kHz	Single Measures	.896[a]	0.74	0.961	17.774	16	16	0
	Average Measures	.945[c]	0.851	0.98	17.774	16	16	0
6 kHz	Single Measures	.939[a]	0.839	0.977	29.932	16	16	0
	Average Measures	.968[c]	0.913	0.989	29.932	16	16	0
8 kHz	Single Measures	.906[a]	0.762	0.965	19.546	16	16	0
	Average Measures	.950[c]	0.865	0.982	19.546	16	16	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 17 summarises the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz between M1 (initial measure) and M5 (measure within one month of initial measurement). *Single measures* results refer to the calculation of ICC between M1 and M5 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M5 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 16 different measures is estimated by these results.

The ICC for single measures at the different frequencies measured was 0.897 at 1 kHz, 0.955 at 1.5 kHz, 0.968 at 2 kHz, 0.932 at 3 kHz, 0.896 at 4 kHz, 0.939 at 6 kHz and 0.906 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 11% variability between the results that could be described by random variation. More than 89% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 4 kHz (0.896) and the highest reliability was observed at 1.5 kHz (0.955).

The ICC for average measures at the different frequencies measured was 0.946 at 1 kHz, 0.977 at 1.5 kHz, 0.984 at 2 kHz, 0.965 at 3 kHz, 0.945 at 4 kHz, 0.968 at 6 kHz and 0.950 at 8 kHz. All the ICC results calculated at all the different frequencies for the DPOAEs indicated that there was less than 6% variability between the results that could be described by random variation. More than 94% of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 4 kHz (0.945) and the highest reliability was observed at 2 kHz (0.984).

The ICC ranging between .896 and .968 for single measures and .945 and .984 for average measures indicates high reliability between the measures made one month apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz and 8 kHz). The lowest lower-bound 95% confidence interval was 0.74 at 4 kHz, and the highest upper-bound 95% confidence interval was 0.994 at 2 kHz.

Previous studies (Sockalingham, Kei, & Ho, 2007) have proven the test-retest reliability of DPOAEs with ICC ranging from 0.64 to 0.89, indicating the high test-retest reliability achieved with this study as positive indicator of statistical significance. In other studies (Wagner, Heppelmann, Vonthein, & Zenner, 2008) the test-retest reliability was confirmed with SEM and confidence interval (CI) at lower levels than indicated by these tests conducted, thus confirming the high test-retest reliability. The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability between measures M1 and M5 during this study, which confirms the test-retest reliability.

Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference by recent studies (Kumar, Methi, & Avinash, 2013). The results obtained by this study indicate that the ICC results above 0.94 for all the different frequencies measured are higher than previous studies' results, confirming the high test-retest reliability.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M5 for the DPOAE results. To further quantify the test-retest reliability of M1 and M5, the TEOAE results were statistically analysed.

4.2.4.2 TEOAEs M1 – M5

Table 18 provides the ICC data comparing M1 and M5 for all the frequencies measured with the TEOAE results.

Table 18: TEAOE ICC comparison of M1 – M5

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.945[a]	0.813	0.981	45.938	16	16	0
	Average Measures	.972[c]	0.897	0.991	45.938	16	16	0
1.5 kHz	Single Measures	.949[a]	0.867	0.981	40.518	16	16	0
	Average Measures	.974[c]	0.929	0.991	40.518	16	16	0
2 kHz	Single Measures	.952[a]	0.875	0.982	39.216	16	16	0
	Average Measures	.976[c]	0.933	0.991	39.216	16	16	0
3 kHz	Single Measures	.956[a]	0.886	0.984	44.001	16	16	0
	Average Measures	.978[c]	0.94	0.992	44.001	16	16	0
4 kHz	Single Measures	.981[a]	0.95	0.993	111.298	16	16	0
	Average Measures	.991[c]	0.974	0.997	111.298	16	16	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 18 summarises the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz between M1 (initial measure) and M5 (measure within one month of initial measurement). *Single measures* results refer to the calculation of ICC between M1 and M5 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average measures* reflect the average measure calculated for M1 and M5 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 16 different measures is estimated by these results.

The ICC for single measures at the different frequencies measured was 0.945 at 1 kHz, 0.949 at 1.5 kHz, 0.952 at 2 kHz, 0.956 at 3 kHz and 0.981 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was less than 6% variability between the results that could be described by random variation. More than 94% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.945) and the highest reliability was observed at 4 kHz (0.981).

The ICC for average measures at the different frequencies measured was 0.972 at 1 kHz, 0.974 at 1.5 kHz, 0.976 at 2 kHz, 0.978 at 3 kHz and 0.991 at 4 kHz. All the ICC results calculated at all the different frequencies for the TEOAEs indicated that there was very little variability between the results that could be described by random variation. More than 97% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 1 kHz (0.972) and the highest reliability was observed at 4 kHz (0.991).

The ICC ranging between .945 and .981 for single measures and .972 and .991 for average measures indicates high reliability between the measures made one month apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz). The lowest lower-bound 95% confidence interval was 0.813 at 1 kHz, and the highest upper-bound 95% confidence interval was 0.997 at 4 kHz.

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also confirmed high test-retest reliability at all measurements during this study, which confirms the test-retest reliability. Large differences in SD in TEOAEs across different participants were also observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) as in this study, which did not influence overall test-retest reliability. This study (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and other studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. The results of this study confirmed that the amplitudes were not reduced at the higher frequencies over longer periods of time indicating even higher test-retest reliability than previous studies conducted. With these statistical results the test-retest reliability was confirmed with previous studies, which proves that with more encouraging statistical results obtained in this study, confidence in TEOAEs over time between M1 and M5 is warranted.

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M5 for the TEOAE results.

Table 19: SD TEOAE M1 and M5 amplitude

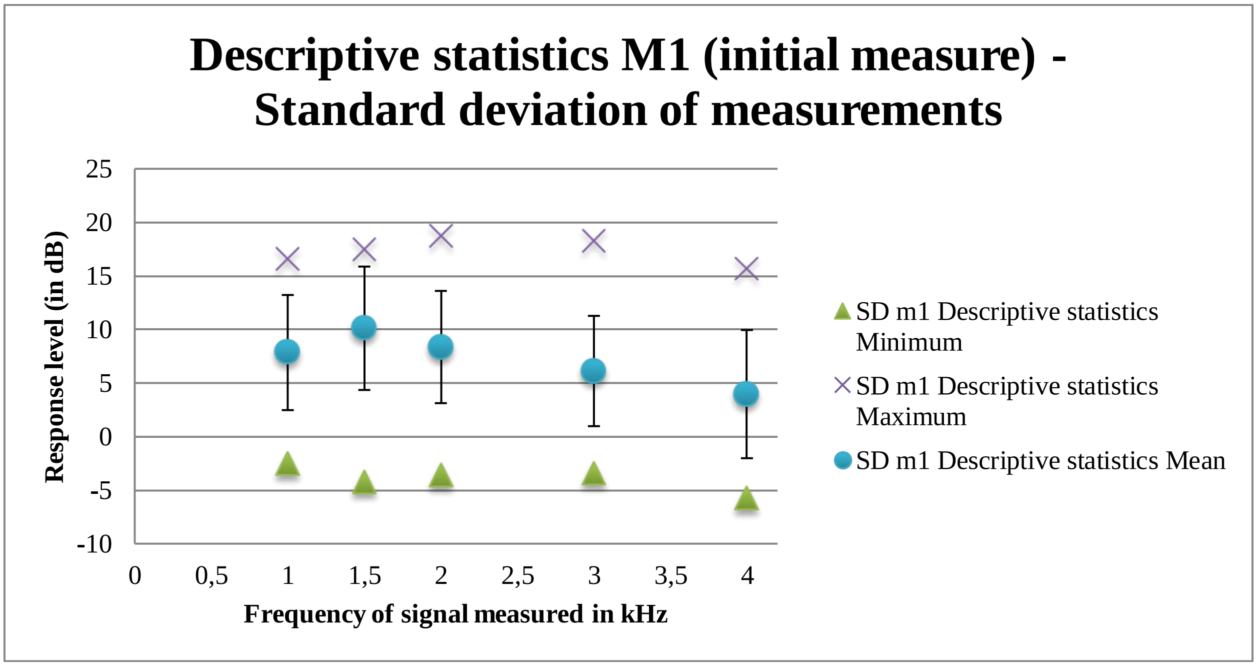
Descriptive statistics M1

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	30	-2.50	16.60	7.8433	5.37476
Signal 1.5 kHz	30	-4.20	17.50	10.1100	5.75951
Signal 2 kHz	30	-3.60	18.70	8.3600	5.23895
Signal 3 kHz	30	-3.40	18.30	6.1200	5.14844
Signal 4 kHz	30	-5.80	15.70	3.9633	5.98374
Valid N (listwise)	30				

Descriptive Statistics M5

	N	Minimum	Maximum	Mean	Std. Deviation
Signal 1 kHz	18	-3.80	17.30	7.4944	4.84592
Signal 1.5 kHz	18	-1.30	15.60	9.4444	4.87940
Signal 2 kHz	18	-1.40	17.60	8.3889	5.29816
Signal 3 kHz	18	-1.70	15.90	6.4278	5.04961
Signal 4 kHz	18	-4.50	14.50	3.8833	6.34714
Valid N (listwise)	18				

In Table 19 the descriptive statistics used when the amplitude of TEOAEs measured at M1 is compared to the amplitude of TEOAEs measured at M5. The minimum, maximum, mean, and SD of the amplitude of the TEOAE measures are similar for M1 and M5, indicating high test-retest reliability. The minimum wave amplitude is the smallest at 4 kHz and the maximum wave amplitude is the largest at 2 kHz at both M1 and M5. The SD is between and 4.84 and 6.34 for all the different frequencies between M1 and M5.



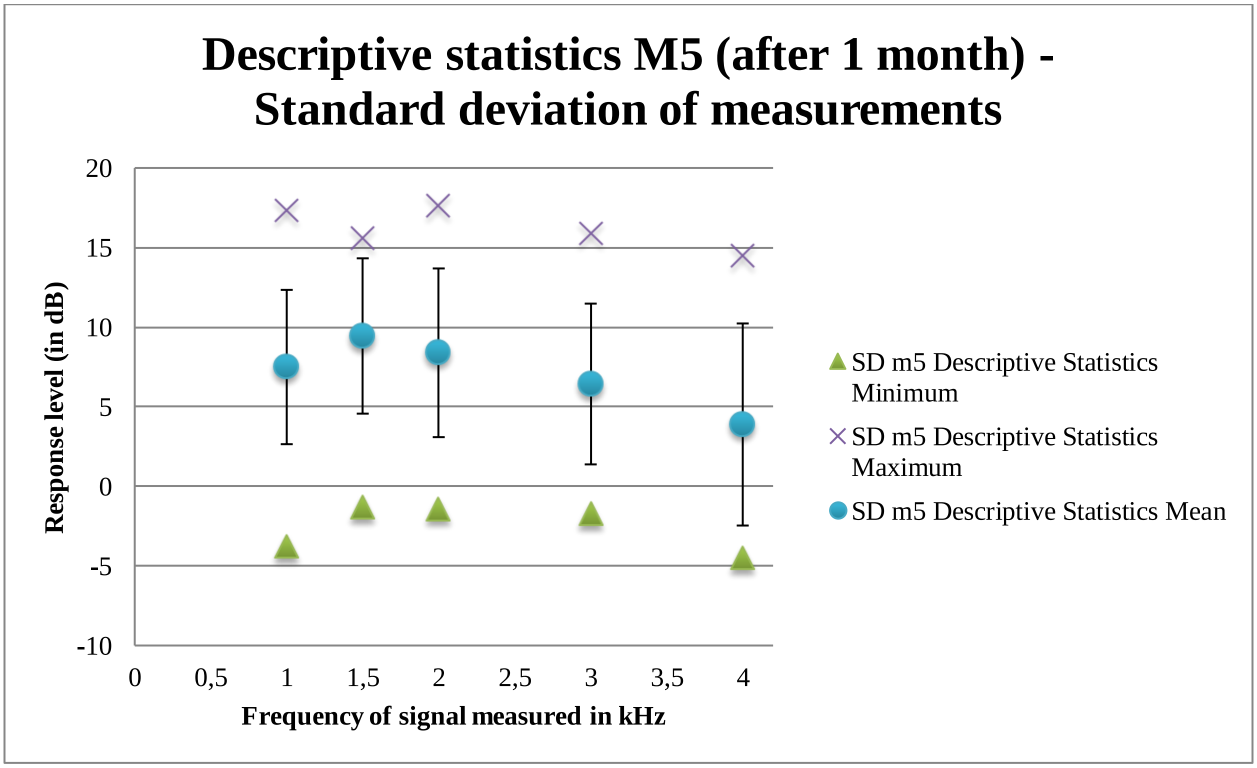


Figure 6: SD of amplitude for measurements of M1 and M5 according to frequency

As summarised in Table 16 and Figure 6, the SD of amplitude between the different tests M1 and M5 showed very little variance, indicating high test-retest reliability between these two tests. The mean response in dB is shown with the error bars indicating the SD observed. The minimum and maximum amplitude response for each frequency is also indicated in the figures.

This visual representation in Figure 6 supports the results summarised in Table 16. The comparison between the mean amplitudes of TEOAEs measured at M1 and M5 show consistency across all frequencies measured, as does the SD and the minimum and maximum amplitudes recorded.

Large differences in SD across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992). It was concluded that the SD of up to 13dB in previous studies did not influence overall test-retest reliability, which was not the case in this study where differences of only up to 5.9dB was observed as SD.

Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and more recent studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. This study confirmed that the SD between M1 and M2 differed only by up to 5.98dB which was a closer resemblance than previous studies conducted. During this study between M1 and M2 the amplitudes were not reduced at the higher frequencies over longer periods of time as with previous studies indicating even higher test-retest reliability than previous studies conducted. Despite the statistical results obtained with previous studies the test-retest reliability was confirmed, which proves that with more encouraging statistical results obtained in this study, TEOAE test-retest reliability is statistically proven between M1 and M5. To further quantify the test-retest reliability of M1 and M5, the OAE-gram results were statistically analysed.

4.2.4.3 OAE-GRAM M1 – M5

Table 20 provides the ICC data comparing M1 and M5 for all the frequencies measured with the OAE-gram calculated.

Table 20: OAE-gram ICC comparison of M1 – M5

ICC								
		Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	p
1 kHz	Single Measures	.809[a]	0.546	0.926	9.024	16	16	0
	Average Measures	.894[c]	0.707	0.962	9.024	16	16	0
1.5 kHz	Single Measures	.659[a]	0.269	0.862	4.665	16	16	0.002
	Average Measures	.794[c]	0.424	0.926	4.665	16	16	0.002
2 kHz	Single Measures	.632[a]	0.249	0.847	4.541	16	16	0.002
	Average Measures	.775[c]	0.399	0.917	4.541	16	16	0.002
3 kHz	Single Measures	.899[a]	0.746	0.962	18.125	16	16	0
	Average Measures	.947[c]	0.855	0.981	18.125	16	16	0
4 kHz	Single Measures	.902[a]	0.75	0.963	18.338	16	16	0
	Average Measures	.948[c]	0.857	0.981	18.338	16	16	0
6 kHz	Single Measures	.855[a]	0.645	0.945	12.234	16	16	0
	Average Measures	.922[c]	0.784	0.972	12.234	16	16	0

Two-way mixed effects model where participant effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A ICCs using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 20 the ICC at 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, and 6 kHz between M1 (initial measure) and M5 (measure within one month of initial measurement) are summarised of the combined DPOAE and TEOAE results according to the algorhythm. *Single measures* results refer to the calculation of ICC between M1 and M5 for each individual participant. This means that the reliability of measures per single participant is estimated by these results. *Average*

measures reflect the average measure calculated for M1 and M5 respectively across all different participants at the two specified points in time. This means that the reliability of the averages of the 16 different measures is estimated by these results.

An ICC of between .632 and .902 for single measures and between .775 and .948 for average measures indicates high reliability between the measures made one month apart at all measured frequencies (1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, and 6 kHz). The lowest lower-bound 95% confidence interval was 0.249 at 2 kHz, and the highest upper-bound 95% confidence interval was 0.981 at 4 kHz.

The ICC for single measures at the different frequencies measured was 0.809 at 1 kHz, 0.659 at 1.5 kHz, 0.630 at 2 kHz, 0.899 at 3 kHz, 0.902 at 4 kHz, 0.855 at 6 kHz. All the ICC results calculated at all the different frequencies for the OAE-gram indicated that there was less than 35% variability between the results that could be described by random variation. More than 65% of the variation in results could be explained by the construct. This indicated reliable results obtained for the single measures. The lowest reliability was observed at 2 kHz (0.630) and the highest reliability was observed at 4 kHz (0.902).

The ICC for average measures at the different frequencies measured was 0.894 at 1 kHz, 0.794 at 1.5 kHz, 0.775 at 2 kHz, 0.947 at 3 kHz, 0.948 at 4 kHz, 0.922 at 6 kHz. All the ICC results calculated at all the different frequencies for the OAE-gram indicated that there was less than 23% variability between the results that could be described by random variation. More than 77%

of the variation in results could be explained by the construct. This indicated reliable results obtained for the average measures. The lowest reliability was observed at 2 kHz (0.775) and the highest reliability was observed at 4 kHz (0.948).

The high ICC for all the different frequencies measured remained high for single measures as well as average measures across the frequencies, with the 95% confidence interval similarly high. This indicates high test-retest reliability at all frequencies between M1 and M5 for the OAE-gram results.

No previous studies were found utilising the OAE-gram and algorhythm as comparison of test-retest reliability. It was included in this study with results resembling the test-retest reliability of the DPOAE and TEOAE statistical significance. It is therefore, concluded that the high ICC applied to the OAE-gram and the similarity to the DPOAE and TEOAE ICC proves the test-retest reliability between M1 and M5 for the OAE-gram.

Test-retest reliability was proven between M1 (initial measure) and M5 (measure within one month of initial measure) using various statistical methods and according to several previous studies. The ICC results obtained in this study when comparing M1 to M5 closely resembled results obtained in previous studies to confirm test-retest reliability of DPOAEs and TEOAEs. The same statistical principles were applied to the OAE-gram generated to confirm test-retest reliability. The test-retest reliability of DPOAEs, TEOAEs and the OAE-gram is statistically confirmed between the measures made one month apart.

4.2.5 Summary of results of sub-aims

The results of the statistical analysis according to the sub aims show that the tests performed over different time periods had high test-retest reliability. The TEOAE test had higher test-retest reliability than the DPOAE test. M1 compared to M2 had the highest test-retest reliability, indicating that the longer time lapse between tests resulted in lower test-retest reliability. However, there was still not a statistical significant difference over time, indicating that the test-retest reliability was proven for all the time intervals measured. The lowest test-retest reliability was recorded between M1 and M3.

With regard to the different frequencies tested, the greatest test-retest reliability was statistically proven at 4 kHz with the lowest test-retest reliability at 1.5 kHz. Even at the lowest test-retest reliability at 1.5 kHz, however, no statistically significant difference was recorded, thus still indicating reliable test-retest results.

4.3 Presentation and discussion of results according to the main aim

The main aim of this research project was to quantitatively assess the test-retest reliability of DPOAEs and TEOAEs and the subsequently generated OAE-gram within a group of normal hearing adults. The results of the different measures were combined to reach conclusions pertaining to the main aim. Descriptive statistic tables and ANOVA tables were compiled along with illustrative graphs utilizing all data captured at different time intervals for each of the tests.

Inferential statistics included one-way repeated measures Analysis of Variance (ANOVA). ANOVA was used to determine the changes in emission amplitudes between the measures M1 to

M5. For DPOAE, TEOAE, and OAE-gram the mean values were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Further descriptive statistical calculations were made for each of the tests to obtain a clear picture of test-retest reliability. For each frequency, the different measurements at the time intervals specified were compared according to the mean, SD, SEM, and the 95% confidence interval for the mean and minimum and maximum values recorded. The DPOAE results are presented and discussed with all measures at different times taken into consideration. The results according to the main aim are presented and discussed in the order that the different measures were made, first DPOAE statistical results, then TEOAE statistical results and finally OAE-gram statistical results.

4.3.1 DPOAE Descriptive Statistics

The descriptive statistics for the DPOAE tests performed at all the different time intervals are presented and discussed in this section. The descriptive statistics provide a summary of the results for all participants with all the results recorded for each participant.

Table 21: DPOAE Descriptive statistics

DPOAE Descriptive statistics									
Frequency	**Measure**	**N**	**Mean**	**Std. Deviation**	**Std. Error**	**95% Confidence Interval for mean**		**Min**	**Max**
						Lower Bound	**Upper bound**		
1 kHz	M1	30	12.42	5.24	0.96	10.47	14.38	0.70	22.78
	M2	29	12.29	5.86	1.09	10.06	14.52	-1.60	22.03
	M3	29	9.76	12.16	2.26	5.14	14.39	-30.00	22.45
	M4	12	13.24	4.83	1.39	10.17	16.31	2.44	21.15
	M5	18	13.64	4.42	1.04	11.44	15.83	4.12	19.85
	Total	118	12.01	7.57	0.70	10.63	13.39	-30.00	22.78
1.5 kHz	M1	30	16.27	3.93	0.72	14.81	17.74	5.98	23.26
	M2	29	15.78	4.10	0.76	14.22	17.34	6.87	22.88
	M3	29	13.52	12.68	2.35	8.70	18.34	30.00	22.71
	M4	12	16.09	4.06	1.17	13.51	18.67	2.44	21.75
	M5	18	16.37	3.32	0.78	14.72	18.02	4.12	23.33
	Total	118	15.47	7.12	0.66	14.17	16.77	-30.00	23.33
2 kHz	M1	30	15.80	4.24	0.77	14.21	17.38	4.00	23.11
	M2	29	15.53	4.50	0.83	13.82	17.24	4.39	22.55
	M3	29	13.07	12.77	2.37	8.21	17.93	-30.00	22.52
	M4	12	15.75	4.63	1.34	12.81	18.65	7.89	22.18
	M5	18	16.21	3.46	0.82	14.49	17.93	9.86	22.97
	Total	118	15.12	7.31	0.67	13.79	16.45	-30.00	23.11
3 kHz	M1	30	14.47	4.66	0.85	12.73	16.22	3.50	21.54
	M2	29	14.73	4.26	0.79	13.11	16.35	4.35	21.83
	M3	29	13.03	9.54	1.77	9.40	16.66	-30.00	21.79
	M4	12	15.31	4.85	1.40	12.23	18.40	5.52	21.28
	M5	18	14.46	5.07	1.20	11.94	16.98	2.60	20.57
	Total	118	14.27	6.17	0.57	13.14	15.39	-30.00	21.83
4 kHz	M1	30	16.69	4.12	0.75	15.15	18.23	5.68	22.62
	M2	29	16.23	4.23	0.79	14.62	17.84	6.01	21.42
	M3	29	14.87	9.59	1.78	11.22	18.52	-30.00	23.40
	M4	12	17.37	4.73	1.36	14.37	20.38	9.05	23.04
	M5	18	16.44	4.37	1.03	14.27	18.62	7.30	21.39
	Total	118	16.16	6.00	0.55	15.06	17.25	-30.00	23.40
6 kHz	M1	30	15.32	5.24	0.96	13.36	17.27	2.09	25.85
	M2	29	14.46	5.70	1.06	12.29	16.63	2.64	25.24
	M3	29	15.50	4.96	0.92	13.61	17.38	1.58	22.67
	M4	12	15.18	6.39	1.84	11.12	19.24	0.29	22.09

	M5	18	14.49	6.64	1.56	11.19	17.79	-1.57	24.01
	Total	118	15.01	5.56	0.51	14.00	16.02	-1.57	25.85
8 kHz	M1	30	8.15	7.92	1.45	5.19	11.11	-8.82	24.07
	M2	29	7.19	8.14	1.51	4.10	10.29	-6.20	23.89
	M3	29	9.08	7.45	1.38	6.24	11.91	-9.52	19.89
	M4	12	7.87	7.02	2.03	3.41	12.33	-5.84	16.66
	M5	18	8.50	7.47	1.76	4.78	12.21	-5.77	18.90
	Total	118	8.17	7.61	0.70	6.78	9.55	-9.52	24.07

In Table 21 the DPOAE descriptive statistics are presented in tabular form. The mean, SD, SEM, 95% confidence interval for the mean with upper- and lower-bound values as well as minimum and maximum values are presented according to the different tests at all the different frequencies measured for DPOAEs. This summary of the descriptive statistics with all DPOAE tests included in this study (immediate re-testing, re-testing after one hour, re-testing after one week and re-testing after one month) proves the test-retest reliability between the different tests conducted at all the different frequencies. The 95% confidence interval for the mean is indicative of the test-retest reliability.

The statistical results obtained with this study indicated that the SD in dB calculated for M1 was 5.24 at 1 kHz, 3.93 at 1.5 kHz, 4.24 at 2 kHz, 4.66 at 3 kHz, 4.12 at 4 kHz, 5.24 at 6 kHz and 7.92 at 8 kHz. The SEM in dB calculated for M1 was 0.96 at 1 kHz, 0.72 at 1.5 kHz, 0.77 at 2 kHz, 0.85 at 3 kHz, 0.75 at 4 kHz, 0.96 at 6 kHz and 1.45 at 8 kHz.

The statistical results obtained with this study indicated that the SD in dB calculated for M2 was 5.86 at 1 kHz, 4.10 at 1.5 kHz, 4.50 at 2 kHz, 4.26 at 3 kHz, 4.23 at 4 kHz, 5.70 at 6 kHz and

8.14 at 8 kHz. The SEM in dB calculated for M2 was 1.09 at 1 kHz, 0.76 at 1.5 kHz, 0.83 at 2 kHz, 0.79 at 3 kHz, 0.79 at 4 kHz, 1.06 at 6 kHz and 1.51 at 8 kHz.

The statistical results obtained with this study indicated that the SD in dB calculated for M3 was 12.16 at 1 kHz, 12.68 at 1.5 kHz, 12.77 at 2 kHz, 9.54 at 3 kHz, 9.59 at 4 kHz, 4.96 at 6 kHz and 7.45 at 8 kHz. The SEM in dB calculated for M3 was 12.16 at 1 kHz, 12.68 at 1.5 kHz, 12.77 at 2 kHz, 9.54 at 3 kHz, 9.59 at 4 kHz, 4.96 at 6 kHz and 7.45 at 8 kHz.

The statistical results obtained with this study indicated that the SD in dB calculated for M4 was 4.83 at 1 kHz, 4.06 at 1.5 kHz, 4.634 at 2 kHz, 4.85 at 3 kHz, 4.73 at 4 kHz, 6.39 at 6 kHz and 7.02 at 8 kHz. The SEM in dB calculated for M4 was 4.83 at 1 kHz, 4.06 at 1.5 kHz, 4.63 at 2 kHz, 4.85 at 3 kHz, 4.73 at 4 kHz, 6.39 at 6 kHz and 7.02 at 8 kHz.

The statistical results obtained with this study indicated that the SD in dB calculated for M5 was 4.42 at 1 kHz, 3.32 at 1.5 kHz, 3.46 at 2 kHz, 5.07 at 3 kHz, 4.37 at 4 kHz, 6.64 at 6 kHz and 7.47 at 8 kHz. The SEM in dB calculated for M5 was 1.04 at 1 kHz, 0.78 at 1.5 kHz, 0.82 at 2 kHz, 1.20 at 3 kHz, 1.03 at 4 kHz, 1.56 at 6 kHz and 1.76 at 8 kHz.

As can be seen in Table 21 the variability at all frequencies and at all different time intervals remained in a narrow band between the lower-bound and upper-bound 95% confidence interval for the mean. This provides a further indication of the high test-retest reliability.

According to previous studies (Beattie, Kenworthy, & Luna, 2003) the SEM at 550 Hz (~4.6dB) were nearly twice as large as those found for 1000Hz, 2000Hz and 4000Hz (~2.5dB). The short-term test-retest data suggest that there is a 95% probability that an individual's true DPOAE will fall within 5 dB of the obtained distortion product at 1000-4000Hz and within 10dB at 550Hz. The SEM of the difference was calculated to assess whether two or more DPOAE measurements are significantly different. The data revealed that short-term differences (probe removed and subject retested on the same day or on different days) between two DPOAEs must exceed approximately 14 dB at 550 Hz and 7 dB at 1000-4000Hz to be statistically significant at the 0.05 level of confidence. According to these results, this study also confirms and indicates a 95% probability that an individual's true DPOAE will fall within 5dB of the distortion product at the frequencies measured between 1 kHz and 8 kHz.

Other previous studies have found greater variability at the higher frequencies (>8 kHz) for DPOAE level measurements (Dreisbach, Long, & Lees, 2006). The average DPOAE level differences between trials for the higher and lower frequencies for the four different stimulus level conditions was 5.15 (SD = 4.40 dB) and 2.80 (SD = 2.70 dB) dB, respectively. Individual subject analysis revealed that high frequency DPOAE levels varied no more than 10 dB for 87.5 and 83.1% of young adult subjects for the 70/75 and 60/50 dB SPL stimulus level conditions, respectively. These results were tested at the highest DPOAE results measured in this study with all (100%) of results obtained in this study exhibiting DPOAE levels that varied no more than 10dB with 70/75 dB SPL stimulus level conditions, confirming high test-retest reliability.

Previous studies analysing the DPOAE amplitudes in the default frequency range (1 to 6 kHz) between test and short-term retest were highly correlated (Ng & MacPherson, 2005), with average *r*=0.81. SEM was found to average 2.64 dB for DPOAE amplitude in the default frequency range. DPOAE amplitudes in the higher frequency range of 6.5 to 7 kHz were also significantly correlated between test and short-term retest; *r* ranged from 0.80 to 0.82 and SEM ranged from 2.59 to 3.04 dB. Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference (Kumar, Methi, & Avinash, 2013), and also found to have high test-retest reliability with immediate remeasurements. The various resolution and retest conditions showed no significant differences in reliability. DPOAE measurement in the 6.5 to 7 kHz range can be considered a reliable tool for monitoring cochlear function in cases such as exposure to ototoxic medication or noise. According to the results obtained with the study in 2005, this study conducted indicated significant correlation between short and longer term test-retest reliability (up to one month after initial measures) as SEM ranges confirmed less variance than any of the previous studies conducted (SEM ranged from 0.51 to 2.37).

The SD and SEM calculated with the DPOAE results closely resemble the results obtained with previous studies indicating that the results fall within the predictable range expected of DPOAE amplitudes with high test-retest reliability. In the previous section the test-retest reliability of DPOAEs were proven for each separate time frame when DPOAEs were conducted. The descriptive statistics confirm the test-retest reliability of the DPOAEs for all participants with all the different measures calculated together. Therefore, high test-retest reliability for DPOAEs can be statistically confirmed. This high test-retest reliability is further illustrated by Figure 7 where the different results for each participant are represented in the figure to illustrate the high correlation proven statistically in a visual manner.

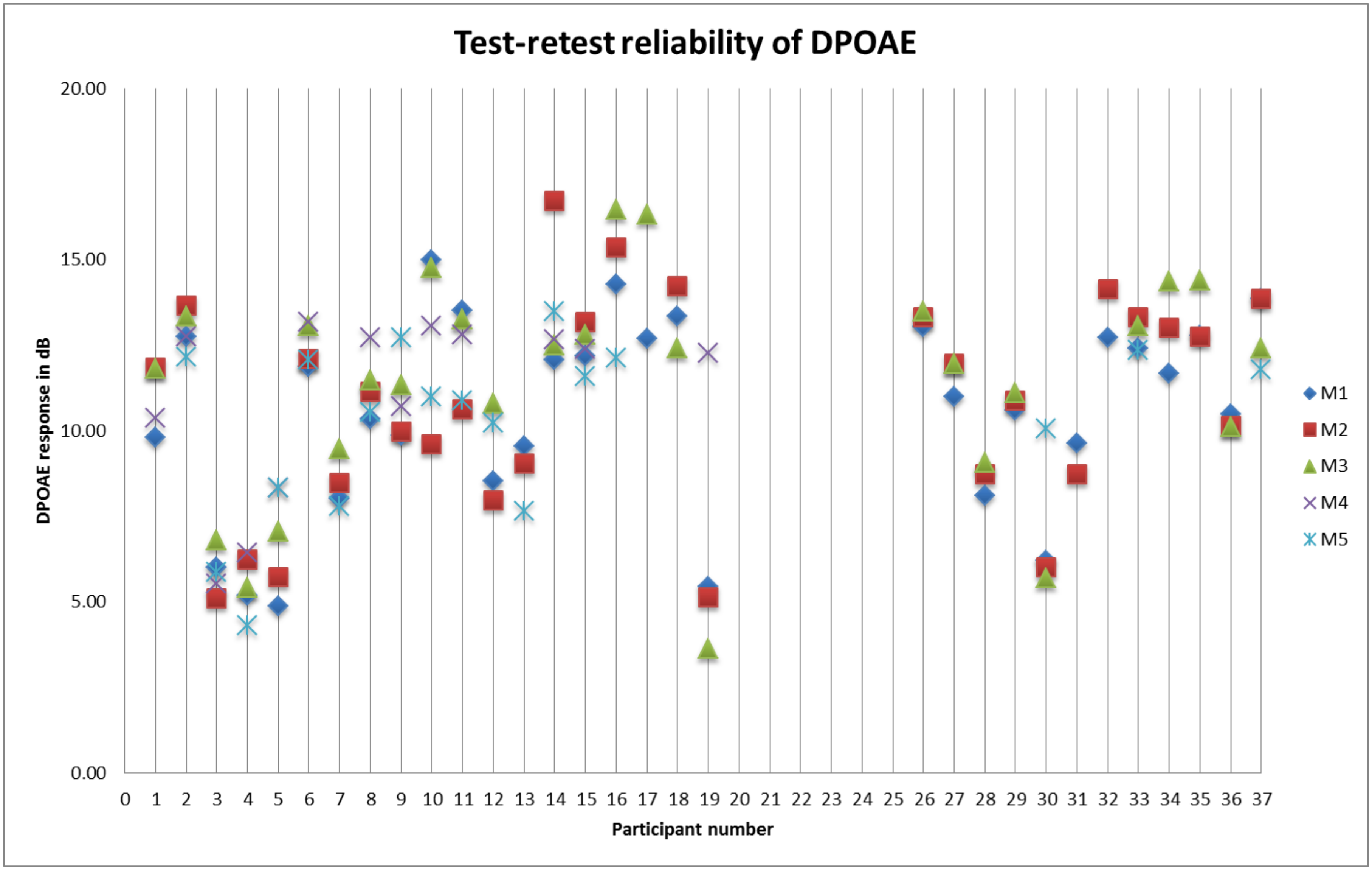

Test-retest reliability of DPOAE
DPOAE response in dB
20.00
15.00
10.00
5.00
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
Participant number
M1
M2
M3
M4
M5

Figure 7 illustrates the test-retest reliability of DPOAEs with the DPOAE amplitude for each test for all participants measured. The graph indicates the average DPOAE response in dB for each participant at all of the different time intervals measured. The deductions made from this graph confirm the limited variance in DPOAE response at the different time intervals for each participant. This highlights the high test-retest reliability of DPOAEs at the different time intervals as proven statistically in Table 21. A minimal change over time is observed between the different measures for each participant. Outliers illustrated by this graph are limited to three participants (number 14, 19 and 30). These participants showed significant variance in one of the tests measured (respectively M2, M4 and M5). A single measure that differed from other measures, even though variance observed was significant, did not statistically affect the overall reliability of the measures. Test-retest reliability was not only measured by variance between each participant, but also between the results classified as a single group per time interval and measured with ANOVA as can be seen by the p-value (Sig) also depicted in Table 22.

Table 22: DPOAE ANOVA

DPOAE ANOVA						
		Sum of Squares	df	Mean Square	F	p
1.0 kHz	Between Groups	219.251	4	54.813	.954	.436
	Within Groups	6489.809	113	57.432		
	Total	6709.060	117			
1.5 kHz	Between Groups	151.282	4	37.821	.739	.567
	Within Groups	5785.006	113	51.195		
	Total	5936.288	117			
2.0 kHz	Between Groups	166.447	4	41.612	.772	.546
	Within Groups	6093.839	113	53.928		
	Total	6260.286	117			
3.0 kHz	Between Groups	65.528	4	16.382	.422	.792
	Within Groups	4384.938	113	38.805		
	Total	4450.465	117			
4.0 kHz	Between Groups	76.074	4	19.019	.519	.722
	Within Groups	4140.863	113	36.645		
	Total	4216.938	117			
6.0 kHz	Between Groups	23.771	4	5.943	.187	.945
	Within Groups	3591.589	113	31.784		
	Total	3615.360	117			
8.0 kHz	Between Groups	54.554	4	13.638	.229	.921
	Within Groups	6716.765	113	59.440		
	Total	6771.318	117			

Inferential statistics for the DPOAE results included one-way repeated measures Analysis of Variance (ANOVA) as seen in Table 22. ANOVA was used to determine the changes in DPOAE results between the measures M1 to M5. For DPOAE results the mean signals were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Further descriptive statistical calculations were made for each of the tests to provide a clear picture of test-retest reliability. For each frequency the different measurements at the time intervals specified were compared between groups and within groups, according to the sum of squares, mean square, F value calculated, and statistical significance (Sig.). The null hypothesis states that the means do not differ across M1 to M5, while the alternative hypothesis states that there is indeed a significant difference between at least two of the time measurements. The p values (Sig) are 0.436 at 1 kHz, 0.567 at 1.5 kHz, 0.546 at 2 kHz, 0.792 at 3 kHz, 0.722 at at 4 kHz, 0.945 at 6 kHz and 0.921 at 8 kHz. Since the p-values (Sig) are greater than 0.05 at all the frequencies, it can be concluded that there are no statistically significant differences in the mean signals across M1 – M5.

One-way repeated measures ANOVA did not show a significant change in emissions between DPOAE measurements across frequencies (Field, 2009). With all DPOAE tests included in this study (immediate re-testing, re-testing after one hour, re-testing after one week, and re-testing after one month) there was no statistically significant difference in the time lapse to influence the test-retest reliability. The different tests over time showed no significant difference as the significance value (Sig) calculated was bigger than 0.05 in every case, indicating no statistical significance in the average results calculated for all tests (Laerd, 2015; Laerd, 2015; Dallal, 2013). This high value of Sig calculated between groups at all different frequencies tested confirms the high test-retest reliability of DPOAEs at different time intervals. Since the p-values have been proven to be greater than 0.05, at all the frequencies, it can be concluded that there are no significant differences in the mean results of DPOAE measures across M1 – M5 (Field, 2009).

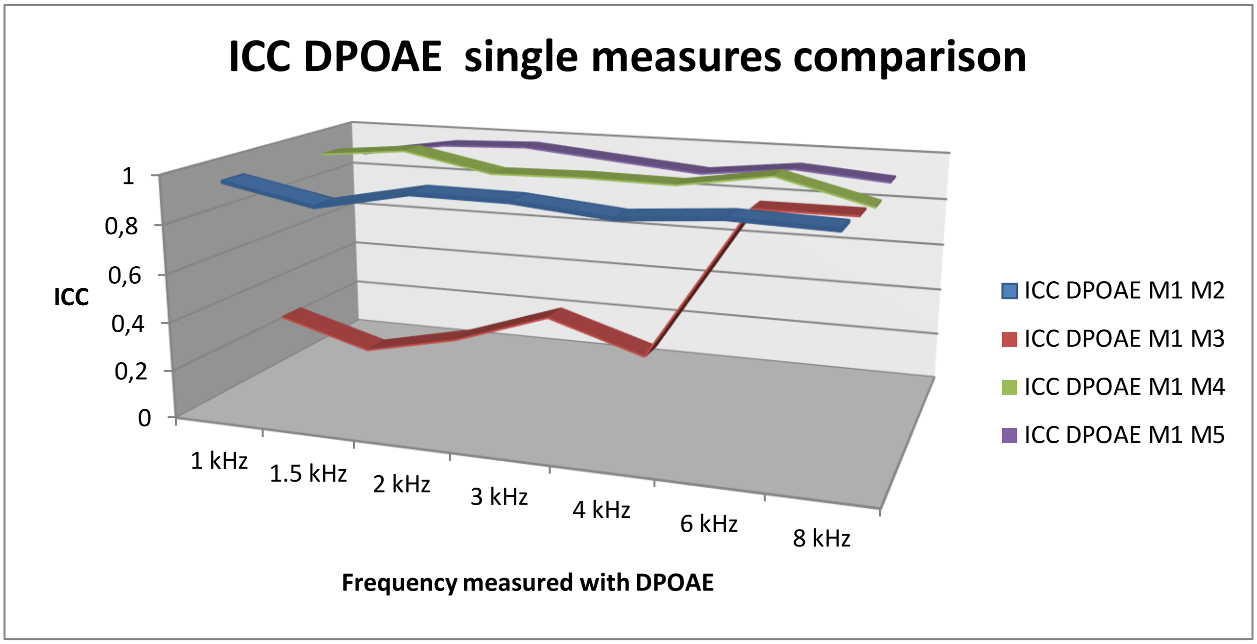


Figure 8: ICC DPOAE single measures comparison

In Figure 8 the ICC single measures for DPOAEs are compared over all the different time intervals. The different DPOAEs are all compared to the initial measured (M1) DPOAE to create an ICC. ICC single measures results refer to the calculation of ICC for each individual participant. This means that the test-retest reliability of measures per single participant is estimated by these results. It describes how strongly units in the same group resemble each other. This gives an accurate indication of the reliability of data collected during this study for each participant at different time intervals. The ICC is used to quantify the degree to which each participant's individual DPOAE results resemble the results of the same participant measured at other points in time.

An ICC of 1 indicates the highest possible correlation between different results, meaning that results are identical with no difference whatsoever. As can be seen in Figure 8, the ICC indicated a high reliability between tests M1 and M2, M1 and M4, and M1 and M5 at low and high frequencies measured. The ICC obtained for M1 compared to M3 did not indicate the same high reliability at lower frequencies as the other measures. However, ICC remained consistently high for high frequency measures of 6 kHz and above. This result could not be explained by a specific variable factor present at M3 (one hour after initial measure). Overall ICC still indicates high test-retest reliability for DPOAEs, but serves to emphasise that any single test should not be viewed in isolation and should be correlated, and results confirmed with other measures at different points in time. The high ICC at all other test points and specifically at the high frequencies confirms the high test-retest reliability.

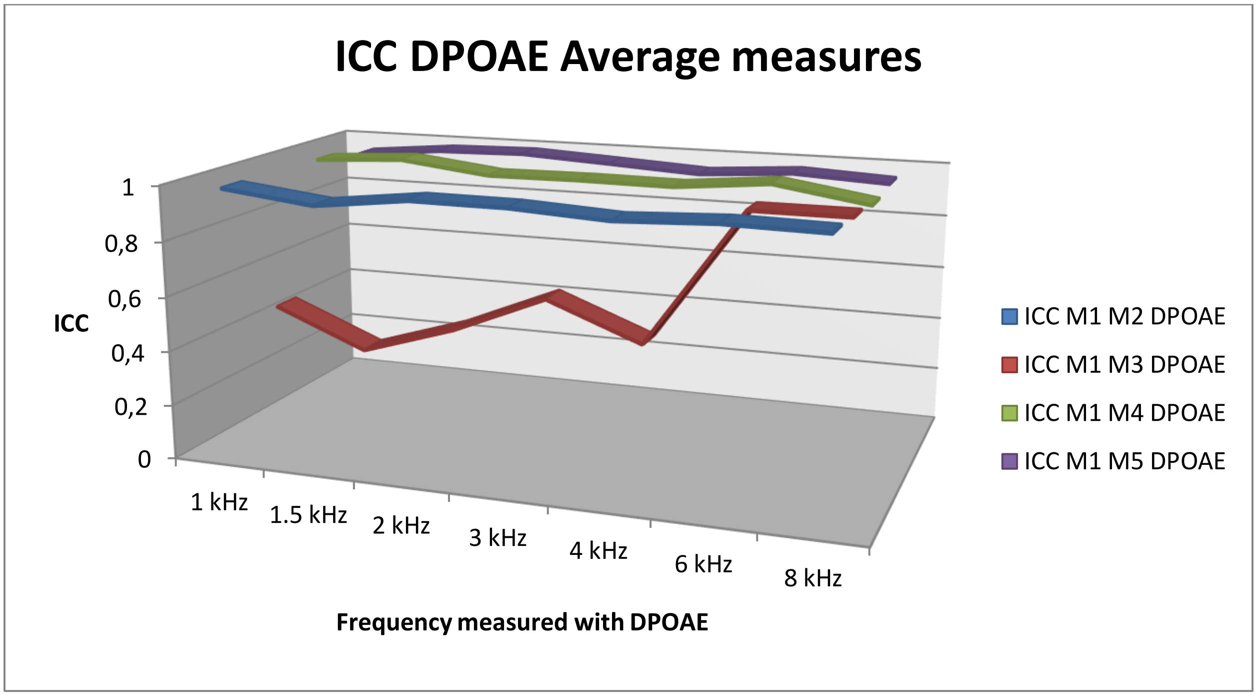


Figure 9: ICC DPOAE Average measures

Figure 9 illustrates the comparison of the average measures of ICC of DPOAEs at different frequencies at all of the different time intervals. The different DPOAEs are all compared to the initial measured (M1) DPOAE to create an ICC. ICC average measures results refer to the calculation of ICC for the group of participants at each point the DPOAEs were measured. This means that the test-retest reliability of measures within the group as a whole is estimated by these results. This gives an accurate indication of the reliability of data collected during this study for the measures collected for the group as a whole at different time intervals. The ICC is used to quantify the degree to which the group of participants' DPOAE results resemble the results measured at other points in time.

An ICC close to 1 indicates high correlation between different results. As can be seen in Figure 9, the ICC indicated a high reliability between tests M1 and M2, M1 and M4, and M1 and M5 at low and high frequencies measured. The ICC obtained for M1 compared to M3 did not indicate the same high reliability at lower frequencies as the other measures. However, the ICC remained consistently high for high frequency measures of 6 kHz and above. This result could not be explained by a specific variable factor present at M3 (one hour after initial measure). The results of the average measures ICC correlate well with the results of single measures ICC, which again emphasises that any single test should not be viewed in isolation and should be viewed in conjunction with other test results, and all results should be confirmed with other measures at different points in time. As the test-retest reliability of DPOAEs have been proven over all the different time periods when tests were conducted, the same statistical principles were used to further analyse the TEOAE results collected.

According to previous studies (Sockalingham, Kei, & Ho, 2007) analysing test-retest reliability using the ICC DPOAE two recordings were performed on the same subjects in the same location using the same equipment. The second recordings were made 13 to 15 days after the first recording. The DPOAE level recorded in the subjects ranged between 13.10 and 20.20 dB for all the five frequencies. The variation in DPOAE level was greater at 10 028 Hz than at other frequencies. The mean difference between the test and retest recordings was 0.52 ±2.87, -1.57± 4.62, 0.01 ±3.38, -0.55 ±2.85, and -0.56 ± 5.57dB at 2530, 3561, 5014, 7029, and 10 028 Hz, respectively. The intra-correlation coefficients for DPOAE level at each of the five (F2) frequencies were 0.85, 0.68, 0.62, 0.89, and 0.64 respectively. Calculations of mean-2SD showed that retest recordings greater than 6.26, 7.67, 6.81, 5.15, and 10.58 dB SPL at 2530, 3561, 5014, 7029, and 10 028 Hz respectively could possibly be interpreted as a significant change in status of the ear. The test-retest reliability illustrated by Figure 8 and Figure 10 confirm that high test-retest reliability was confirmed for all the test frequencies at M2, M4 and M5 when compared to M1 and also for the high frequencies (6 kHz and 8 kHz) when M3 is compared to M1 for single and average measures. TEOAE results are presented and analysed statistically to further explore the test-retest reliability of these measures.

4.3.2 TEOAE Descriptive Statistics

The descriptive statistics for the TEOAE tests performed at all the different time intervals are presented and discussed in this section. The descriptive statistics provide a summary of the results for all participants with all the results recorded for each participant.

Table 23: TEOAE Descriptive statistics

TEOAE Descriptives

Frequency	Measure	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for mean		Minimum	Maximum
						Lower Bound	Upper bound		
1 kHz	M1	31	7.710	5.3366	0.9585	5.752	9.667	-2.5	16.6
	M2	30	8.170	5.2473	0.9580	6.211	10.129	-3.2	17.0
	M3	30	8.313	5.2985	0.9674	6.335	10.292	-3.4	17.3
	M4	12	8.158	5.8165	1.6791	4.463	11.854	-3.6	15.6
	M5	19	7.700	4.7939	1.0998	5.389	10.011	-3.8	17.3
	Total	122	8.014	5.1915	0.4700	7.083	8.944	-3.8	17.3
1.5 kHz	M1	31	9.906	5.7750	1.0372	7.788	12.025	-4.2	17.5
	M2	30	10.517	5.9003	1.0772	8.313	12.720	-4.8	18.6
	M3	30	10.340	5.9836	1.0925	8.106	12.574	-4.8	20.4
	M4	12	8.300	6.4892	1.8733	4.177	12.423	-5.4	16.0
	M5	19	9.884	5.1147	1.1734	7.419	12.349	-1.3	17.8
	Total	122	10.002	5.7705	0.5224	8.967	11.036	-5.4	20.4
2 kHz	M1	31	7.929	5.6824	1.0204	5.845	10.013	-5.0	18.7
	M2	30	8.917	5.0811	0.9277	7.019	10.814	-2.9	18.9
	M3	30	8.553	5.1122	0.9334	6.644	10.462	-3.0	18.0
	M4	12	9.242	5.0506	1.4580	6.033	12.451	2.5	18.1
	M5	19	8.137	5.2648	1.2078	5.599	10.674	-1.4	17.6
	Total	122	8.487	5.2059	0.4713	7.554	9.420	-5.0	18.9
3 kHz	M1	31	5.850	5.2830	0.9490	3.910	7.790	-3.0	18.0
	M2	30	6.500	4.8580	0.8870	4.680	8.310	-2.0	19.0
	M3	30	6.110	5.0990	0.9310	4.210	8.010	-2.0	18.0
	M4	12	7.480	5.4100	1.5620	4.050	10.920	0.0	17.0
	M5	19	6.160	5.0460	1.1580	3.730	8.590	-2.0	16.0
	Total	122	6.280	5.0470	0.4570	5.380	7.190	-3.0	19.0
4 kHz	M1	31	3.703	6.0588	1.0882	1.481	5.926	-5.8	15.7
	M2	30	3.957	5.7731	1.0540	1.801	6.112	-6.8	15.8
	M3	30	4.603	5.8891	1.0770	2.401	6.806	-4.3	15.5
	M4	12	4.583	5.5475	1.6014	1.059	8.108	-4.4	15.3
	M5	19	3.763	6.1905	1.4202	0.779	6.747	-4.5	14.5
	Total	122	4.083	5.8375	0.5285	3.036	5.129	-6.8	15.8

In Table 23 the TEOAE descriptive statistics are presented in tabular format. The mean, SD, SEM, 95% confidence interval for the mean with upper- and lower-bound values as well as minimum and maximum values are presented according to the different tests at all the different frequencies measured for TEOAEs. This summary of the descriptive statistics with all TEOAE tests included in this study (immediate re-testing, re-testing after one hour, re-testing after one week and re-testing after one month) proves the test-retest reliability between the different tests conducted at all the different frequencies. The 95% confidence interval for the mean is indicative of the test-retest reliability. As can be seen in Table 23, the variability at all frequencies and at all different time intervals remained in a narrow band between the lower bound and upper-bound 95% confidence interval for the mean. This further indicates the high test-retest reliability with statistical analysis.

The statistical results obtained with this study indicated that the SD in dB calculated for M1 was 5.33 at 1 kHz, 5.77 at 1.5 kHz, 5.68 at 2 kHz, 5.28 at 3 kHz and 6.05 at 4 kHz. The SEM in dB calculated for M1 was 0.95 at 1 kHz, 1.03 at 1.5 kHz, 1.02 at 2 kHz, 0.94 at 3 kHz and 1.08 at 4 kHz. The statistical results obtained with this study indicated that the SD in dB calculated for M2 was 5.24 at 1 kHz, 5.9 at 1.5 kHz, 5.08 at 2 kHz, 4.85 at 3 kHz and 5.7 at 4 kHz. The SEM in dB calculated for M1 was 0.95 at 1 kHz, 1.07 at 1.5 kHz, 0.92 at 2 kHz, 0.88 at 3 kHz and 1.05 at 4 kHz. The statistical results obtained with this study indicated that the SD in dB calculated for M3 was 5.29 at 1 kHz, 5.9 at 1.5 kHz, 5.1 at 2 kHz, 5.09 at 3 kHz and 5.8 at 4 kHz. The SEM in dB calculated for M1 was 0.96 at 1 kHz, 1.09 at 1.5 kHz, 0.93 at 2 kHz, 0.93 at 3 kHz and 1.07 at 4 kHz. The statistical results obtained with this study indicated that the SD in dB calculated for M4 was 5.81 at 1 kHz, 6.48 at 1.5 kHz, 5.05 at 2 kHz, 5.41 at 3 kHz and 5.54 at 4 kHz. The SEM in dB calculated for M1 was 1.67 at 1 kHz, 1.879 at 1.5 kHz, 1.45 at 2 kHz, 1.56 at 3 kHz and 1.6

at 4 kHz. The statistical results obtained with this study indicated that the SD in dB calculated for M5 was 4.79 at 1 kHz, 5.1 at 1.5 kHz, 5.2 at 2 kHz, 5.04 at 3 kHz and 6.19 at 4 kHz. The SEM in dB calculated for M1 was 1 at 1 kHz, 1.17 at 1.5 kHz, 1.20 at 2 kHz, 1.15 at 3 kHz and 1.42 at 4 kHz.

In previous studies (Vedantan & Musiek, 1991) the TEOAE test-retest reliability was measured and quantified according to the amplitude of the TEOAE echo obtained. The echo amplitude showed wide variation across ears and had a range of 20.9 dB, with the mean TEOAE 13dB. These results indicated reliable test-retest measures. With the amplitude measured in the current study, less variation with amplitude was calculated statistically (amplitude for this study range between 4.7 dB to 6.4 dB) as can be seen in Table 23, which confirms higher test-retest reliability for TEOAEs for this study than for the previous study (Vedantan & Musiek, 1991) conducted.

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also correlated with the high test-retest reliability at all measurements during this study, which confirms the test-retest reliability. Differences in SD in TEOAEs across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) which was even smaller in this study, which did not influence overall test-retest reliability. Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and more recent studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. The results of this study showed that the amplitudes were not reduced at the higher frequencies over longer periods of time indicating even higher test-retest reliability than previous studies conducted. Even with these statistical results indicating lower amplitudes the test-retest reliability was confirmed with previous studies, which proves that with more encouraging statistical results obtained in this study, confidence in TEOAEs across all the different time frames when TEOAEs were recorded, is confirmed. This test-retest reliability of TEOAEs can further be demonstrated by the visual representation of the TEOAE results for each participant at all different times measured in Figure 11.

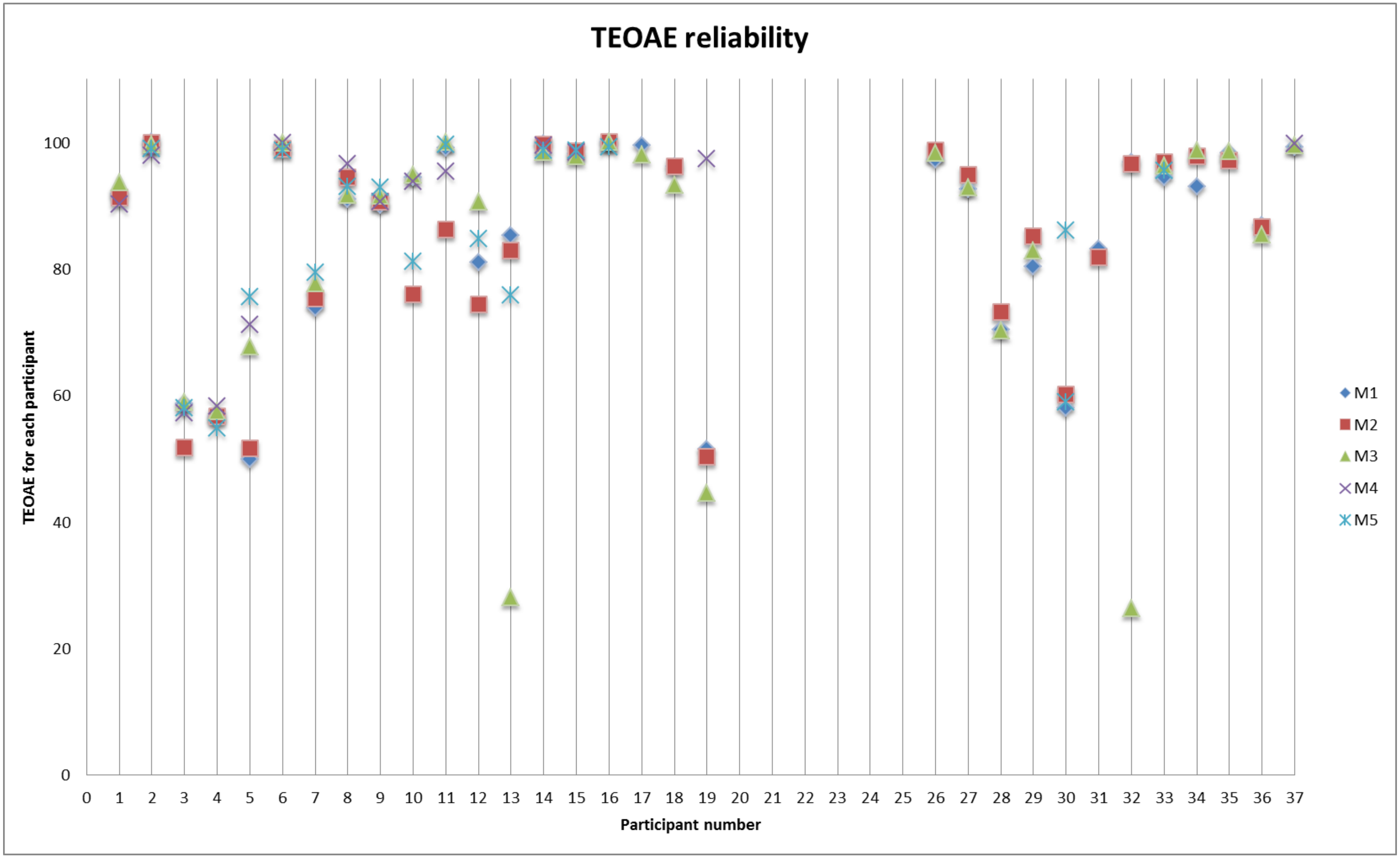


Figure 11: Test-retest reliability of TEOAEs

Figure 11 illustrates the test-retest reliability of TEOAEs. The figure indicates the average TEOAE response in dB for each participant at all of the different time intervals measured. The deductions made from this figure confirm the limited variance in TEOAE response at the different time intervals for each participant explained by Table 23 indicating high test-retest reliability. This highlights the high test-retest reliability of TEOAEs at the different time intervals, because a minimal change over time is observed between the different measures for each participant. Outliers illustrated by this graph are limited to three participants (number 13, 19, and 32). These participants showed significant variance in one of the tests measured (respectively M3, M4, and M3). A single measure that differed from other measures, even though significant variance was observed, did not affect the overall test-retest reliability of the measures. Test-retest reliability was not only measured by variance for each participant, but also between the results classified as a single group per time interval and measured with ANOVA as can be seen by the p-value (Sig) also calculated in Table 24. The mean results at different time intervals are further used to illustrate the test-retest reliability in Figure 12.

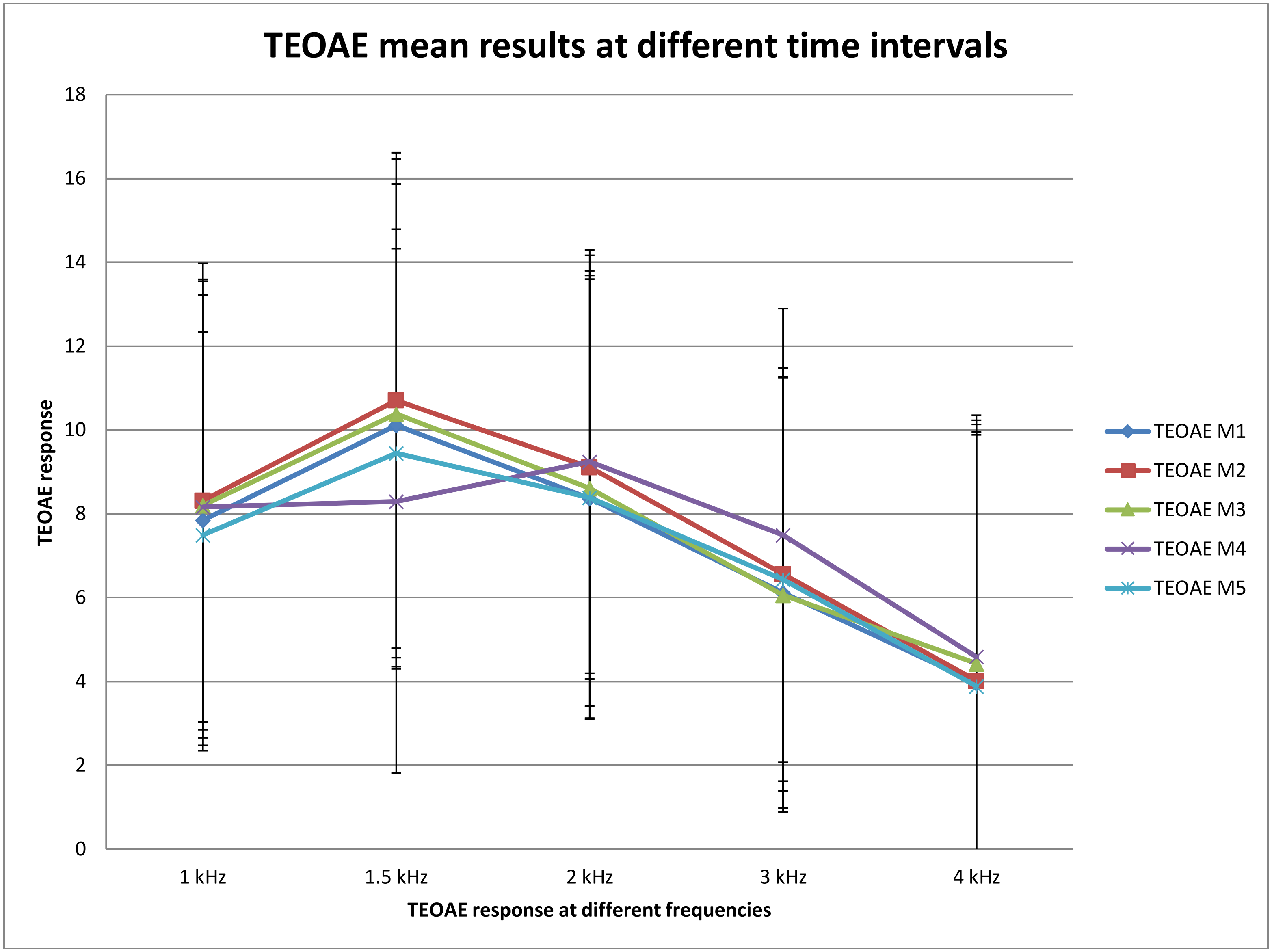


Figure 12: TEOAE mean results at different time intervals

Figure 12 illustrates the mean results (with SD shown with error bars) for TEOAE measures collected at the different time intervals. This figure illustrates the high correlation between test results obtained at different intervals in time for the group as a whole at different frequencies measured. The highest correlation (points grouped closest together) can be observed at 4 kHz and 1 k Hz. The lowest correlation occurs at 1.5 kHz with the measure at M4 being classified as an outlier on this scale. Even though lower correlation can be observed at this single frequency, it does not have an effect on the statistical reliability as the difference is not significant ($p<0.05$). Thus overall test-retest reliability can be illustrated through the correlation of the mean results at different times of TEOAE measurements at all frequencies measured. The ICC can also be visually represented to illustrate the test-retest reliability of TEOAEs made at all different time intervals as shown in Figure 13 and Figure 14.

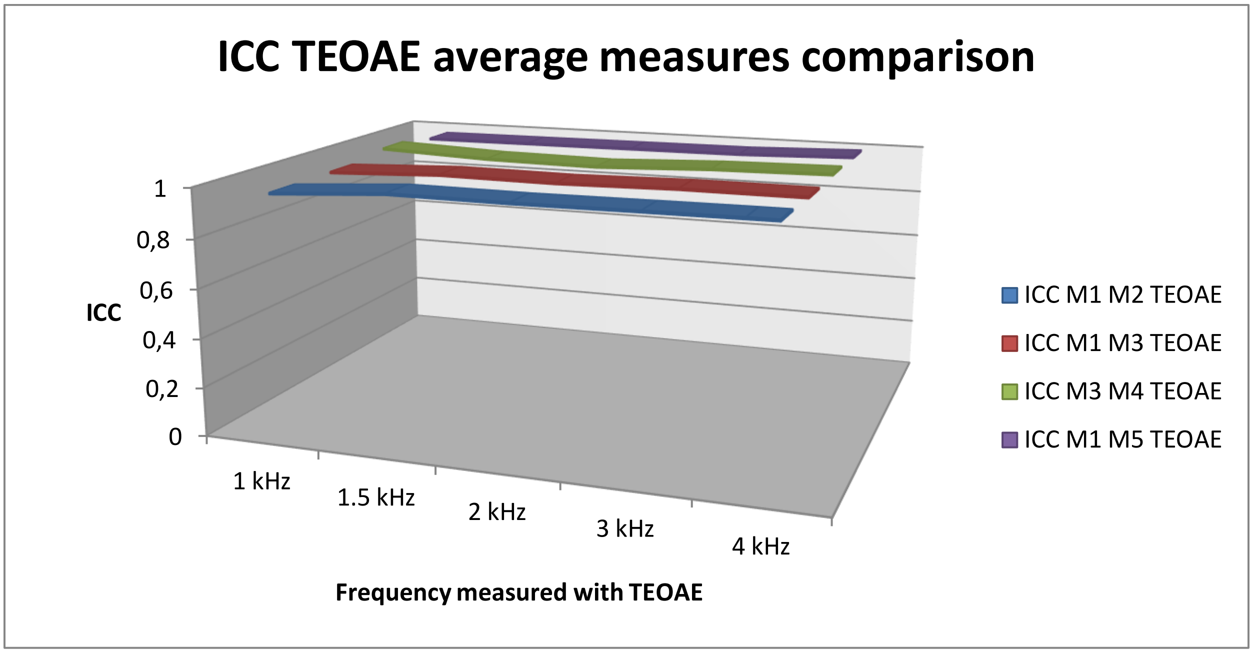


Figure 13: ICC TEOAE average measures comparison

Figure 13 illustrates the comparison of the average measures of ICC of TEOAEs at different frequencies at all of the different time intervals. The different TEOAEs are all compared to the initial measured (M1) TEOAE to create an ICC. ICC average measures results refer to the calculation of ICC for the group of participants at each point the TEOAEs were measured. This means that the test-retest reliability of measures within the group as a whole is estimated by these results. This gives an accurate indication of the reliability of data collected during this study for the measures collected for the group as a whole at different time intervals.

The ICC is used to quantify the degree to which the group of participants' TEOAE results resemble the results measured at other points in time. An ICC close to 1 indicates high correlation between different results. As can be seen in Figure 13 the ICC indicated a high reliability between TEOAEs conducted at all different time intervals at all frequencies. The results indicate overall high test-retest reliability of TEOAEs in this study. The ICC was also calculated for the single measures to compare over all different time intervals when TEOAEs were conducted.

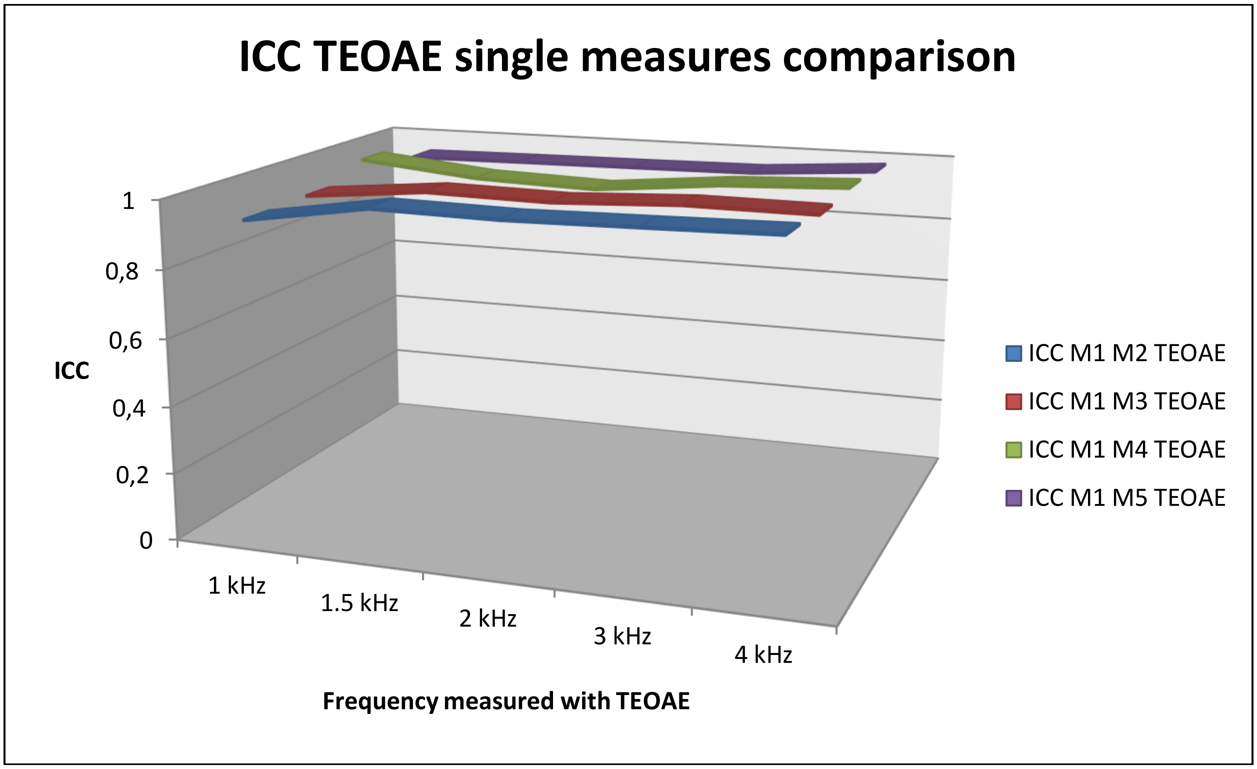


Figure 14: ICC TEOAE single measures comparison

In Figure 14 the ICC single measures are compared over all the different time intervals the TEOAEs were measured. The different TEOAEs are all compared to the initial measured (M1) TEOAE to create an ICC. ICC single measures results refer to the calculation of ICC for each individual participant. This means that the test-retest reliability of measures per single participant is estimated by these results. It describes how strongly units in the same group resemble each other. This gives an accurate indication of the reliability of data collected during this study for each participant at different time intervals.

The ICC is used to quantify the degree to which each participant’s individual TEOAE results resemble the results of the same participant measured at other points in time. An ICC close to 1 indicates high correlation between different results. As can be seen in Figure 14, the ICC indicated a high reliability between all the different tests conducted at all frequencies measured. Overall ICC for single measures indicates high test-retest reliability for TEOAEs between participants at different points of measurement. Inferential statistics also included ANOVA analysis. These results are also further analysed and discussed for the TEOAE measures as a whole to measure test-retest reliability.

Table 24: TEOAE ANOVA

TEOAE ANOVA						
		Sum of Squares	df	Mean Square	F	p
1 kHz	Between Groups	8.412	4	2.103	.076	.990
	Within Groups	3252.794	117	27.802		
	Total	3261.206	121			
1.5 kHz	Between Groups	46.682	4	11.671	.343	.849
	Within Groups	3982.498	117	34.038		
	Total	4029.180	121			
2 kHz	Between Groups	24.485	4	6.121	.220	.927
	Within Groups	3254.834	117	27.819		
	Total	3279.319	121			
3 kHz	Between Groups	25.710	4	6.427	.246	.912
	Within Groups	3056.117	117	26.121		
	Total	3081.827	121			
4 kHz	Between Groups	18.020	4	4.505	.128	.972
	Within Groups	4105.294	117	35.088		
	Total	4123.314	121			

Inferential statistics for the TEOAE results included one-way repeated measures Analysis of Variance (ANOVA) as seen in Table 24. ANOVA was used to determine the changes in TEOAE results between the measures M1 to M5. For TEOAE results the mean signals were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Further descriptive statistical calculations were made for each of the tests to provide a clear picture of test-retest reliability. For each frequency, the different measurements at the time intervals specified were compared between groups and within groups, according to the sum of squares, mean square, F value calculated, and statistical significance (Sig.). The null hypothesis states that the means do not differ across M1 to M5, while the alternative hypothesis states that there is indeed a significant difference between at least two of the time measurements. The p values (Sig) are 0.900 at 1 kHz, 0.849 at 1.5 kHz, 0.927 at 2 kHz, 0.912 at 3 kHz and 0.972 at 4 kHz. Since the p-values (Sig) are greater than 0.05 at all the frequencies, it can be concluded that no statistically significant differences exist in the mean signals across all time measurements (Field, 2009). One-way repeated measures ANOVA did not show a significant change in emissions between TEOAE measurements across frequencies. With all TEOAE tests included in this study (immediate re-testing, re-testing after one hour, re-testing after one week and re-testing after one month) there was no significant difference in the time lapse to influence the test-retest reliability. The different tests over time showed no significant difference as the significance value (Sig) was in every case > 0.05 indicating no statistical significance in the average results calculated for all tests (Laerd, 2015; Laerd, 2015; Dallal, 2013). This high value of Sig calculated between groups at all different frequencies tested confirms the high test-retest reliability of TEOAEs at different time intervals. The null hypothesis states that the means do not differ across M1 to M5, while the alternative hypothesis states that there is indeed a significant difference between at least two of the time measurements. Since the p-values have been proven to be greater than 0.05, at all the frequencies, it can be concluded that there are no significant differences in the mean results of TEOAE measures across all time intervals (Field, 2009). The OAE-gram results were analysed statistically to either confirm or reject test-retest reliability of the results.

4.3.3 OAE-gram Descriptive Statistics

The descriptive statistics for the OAE-gram tests performed at all the different time intervals are presented and discussed in this section. The descriptive statistics provide a summary of the results for all participants with all the results recorded for each participant.

Table 25: OAE-gram Descriptive statistics

OAE-gram Descriptive statistics

Frequency	**Measure**	**N**	**Mean**	**Std. Deviation**	**Std. Error**	**95% Confidence Interval for mean**		**Minimum**	**Maximum**
						Lower Bound	**Upper bound**		
1 kHz	M1	60	87.34	17.88	2.31	82.73	91.96	34.53	100
	M2	58	89.22	14.64	1.92	85.37	93.07	47.51	100
	M3	58	82.88	27.38	3.59	75.68	90.07	0.00	100
	M4	24	89.08	17.11	3.49	81.85	96.30	42.17	100
	M5	36	89.81	14.20	2.37	85.01	94.61	50.47	100
	Total	236	87.26	19.49	1.27	84.76	89.76	0.00	100
1.5 kHz	M1	60	96.21	10.38	1.34	93.53	98.89	51.32	100
	M2	58	93.69	13.72	1.80	90.08	97.30	49.84	100
	M3	58	89.18	27.56	3.62	81.93	96.42	0.00	100
	M4	24	94.27	12.38	2.53	89.04	99.49	62.14	100
	M5	36	99.12	2.51	0.42	98.27	99.97	90.24	100
	Total	236	94.11	16.85	1.10	91.95	96.27	0.00	100
2 kHz	M1	60	93.08	15.83	2.04	88.99	97.17	29.32	100
	M2	58	93.97	14.57	1.91	90.14	97.81	33.93	100
	M3	58	87.63	28.11	3.69	80.24	95.02	0.96	100
	M4	24	95.57	8.50	1.74	91.98	99.16	70.05	100
	M5	36	95.88	7.18	1.20	93.45	98.31	75.56	100
	Total	236	92.64	18.16	1.18	90.31	94.97	0.96	100
2.5 kHz	M1	60	92.48	19.31	2.49	87.86	97.83	30.21	100
	M2	58	95.05	15.05	1.98	91.09	99.01	32.10	100
	M3	58	86.93	27.43	3.60	79.72	94.15	0.93	100
	M4	24	94.10	18.66	3.81	86.22	101.98	33.58	100
	M5	36	91.61	20.60	3.43	84.64	98.58	25.09	100
	Total	236	91.87	20.92	1.36	89.19	94.56	0.93	100
3 kHz	M1	60	93.11	18.43	2.38	88.34	97.87	27.54	100

	M2	58	93.36	17.13	2.25	88.85	97.86	29.01	100
	M3	58	88.13	24.58	3.23	81.66	94.59	0.77	100
	M4	24	91.35	17.81	3.64	83.83	98.87	42.01	100
	M5	36	91.23	19.69	3.28	84.56	97.89	30.93	100
	Total	236	91.48	19.91	1.30	88.93	94.03	0.77	100
3.5 kHz	M1	60	91.29	16.95	2.19	85.91	95.67	34.56	100
	M2	58	88.75	20.27	2.66	83.42	94.07	29.97	100
	M3	58	84.65	27.63	3.63	77.38	91.91	0.00	100
	M4	24	90.92	14.32	2.92	84.88	96.97	53.86	100
	M5	36	89.65	16.95	2.83	83.91	95.38	41.11	100
	Total	236	88.74	20.64	1.34	86.10	91.39	0.00	100
4 kHz	M1	60	83.56	22.85	2.95	77.65	89.46	22.47	100
	M2	58	79.30	24.81	3.26	72.78	85.83	16.54	100
	M3	58	78.52	29.41	3.86	70.79	86.26	0.09	100
	M4	24	84.40	23.47	4.79	74.49	94.31	21.99	100
	M5	36	79.07	26.46	4.41	70.12	88.02	22.23	100
	Total	236	80.38	25.57	1.66	77.40	83.95	7.27	100
4.5 kHz	M1	60	79.70	24.90	3.21	73.27	86.13	21.99	100
	M2	58	76.49	26.00	3.41	69.66	83.33	22.23	100
	M3	58	79.70	25.77	3.38	72.93	86.48	7.27	100
	M4	24	80.76	28.68	5.85	68.65	92.87	20.17	100
	M5	36	73.30	28.13	4.69	63.78	82.82	10.90	100
	Total	236	78.04	26.18	1.70	74.69	81.40	7.27	100
5 kHz	M1	60	79.11	25.09	3.24	72.63	85.59	17.25	100
	M2	58	74.30	27.32	3.59	67.11	81.48	17.65	100
	M3	58	78.24	28.54	3.75	70.74	85.75	0.00	100
	M4	24	81.60	29.00	6.12	68.93	94.26	11.09	100
	M5	36	76.63	30.20	5.03	66.41	86.85	7.24	100
	Total	236	77.60	27.67	1.80	74.04	81.14	0.00	100
5.5 kHz	M1	60	82.43	26.72	3.45	75.53	89.34	14.31	100
	M2	58	75.40	29.59	3.89	67.62	83.18	14.68	100
	M3	58	82.61	26.04	3.42	75.76	89.46	15.57	100
	M4	24	81.96	31.23	6.37	68.78	95.15	15.02	100
	M5	36	75.61	32.16	5.36	64.73	86.49	3.32	100
	Total	236	79.66	28.59	1.86	75.99	83.33	3.32	100
6 kHz	M1	60	82.06	28.79	3.72	74.63	84.50	15.40	100
	M2	58	76.88	29.91	3.93	69.02	84.74	23.58	100
	M3	58	84.89	27.37	3.59	77.69	92.08	16.91	100

	M4	24	82.17	27.52	5.62	70.55	93.79	26.59	100
	M5	36	78.29	30.16	5.03	68.08	88.49	22.96	100
	Total	236	80.92	28.74	1.87	77.23	84.60	15.40	100
6.5 kHz	M1	60	83.06	28.67	3.70	75.65	90.47	0.00	100
	M2	58	83.92	22.67	2.98	77.96	89.88	32.55	100
	M3	58	83.66	31.04	4.08	75.50	91.82	0.00	100
	M4	24	82.88	24.29	4.96	72.62	93.14	23.04	100
	M5	36	88.68	19.91	3.32	81.94	95.41	25.45	100
	Total	236	84.26	26.18	1.70	80.90	87.61	0.00	100

The OAE-gram descriptive statistics are presented in Table 25. The mean, SD, SEM, 95% confidence interval for the mean with upper- and lower-bound values as well as minimum and maximum values are presented according to the different tests at all the different frequencies measured for OAE-grams. This summary of the descriptive statistics with all OAE-gram tests included in this study namely immediate re-testing, re-testing after one hour, re-testing after one week and re-testing after one month, proves the test-retest reliability between the different tests conducted at all the different frequencies. The 95% confidence interval for mean is indicative of the test-retest reliability. As demonstrated in Table 25, the variability at all frequencies and at all different time intervals tests were conducted remained in a narrow band between the lower-bound and upper-bound 95% confidence interval for the mean. This further indicates the high test-retest reliability with statistical analysis.

The statistical results obtained with this study indicated that the SD in percentage hair cell damage calculated for M1 was 5.24 at 1 kHz, 3.93 at 1.5 kHz, 4.24 at 2 kHz, 4.66 at 3 kHz, 4.12 at 4 kHz, 5.24 at 6 kHz and 7.92 at 8 kHz. The SEM in percentage hair cell damage calculated for M1 was 0.96 at 1 kHz, 0.72 at 1.5 kHz, 0.77 at 2 kHz, 0.85 at 3 kHz, 0.75 at 4 kHz, 0.96 at 6 kHz and 1.45 at 8 kHz.

The statistical results obtained with this study indicated that the SD in percentage hair cell damage calculated for M2 was 5.86 at 1 kHz, 4.10 at 1.5 kHz, 4.50 at 2 kHz, 4.26 at 3 kHz, 4.23 at 4 kHz, 5.70 at 6 kHz and 8.14 at 8 kHz. The SEM in percentage hair cell damage calculated for M2 was 1.09 at 1 kHz, 0.76 at 1.5 kHz, 0.83 at 2 kHz, 0.79 at 3 kHz, 0.79 at 4 kHz, 1.06 at 6 kHz and 1.51 at 8 kHz.

The statistical results obtained with this study indicated that the SD in percentage hair cell damage calculated for M3 was 12.16 at 1 kHz, 12.68 at 1.5 kHz, 12.77 at 2 kHz, 9.54 at 3 kHz, 9.59 at 4 kHz, 4.96 at 6 kHz and 7.45 at 8 kHz. The SEM in percentage hair cell damage calculated for M3 was 12.16 at 1 kHz, 12.68 at 1.5 kHz, 12.77 at 2 kHz, 9.54 at 3 kHz, 9.59 at 4 kHz, 4.96 at 6 kHz and 7.45 at 8 kHz.

The statistical results obtained with this study indicated that the SD in percentage hair cell damage calculated for M4 was 4.83 at 1 kHz, 4.06 at 1.5 kHz, 4.634 at 2 kHz, 4.85 at 3 kHz, 4.73 at 4 kHz, 6.39 at 6 kHz and 7.02 at 8 kHz. The SEM in percentage hair cell damage calculated for M4 was 4.83 at 1 kHz, 4.06 at 1.5 kHz, 4.63 at 2 kHz, 4.85 at 3 kHz, 4.73 at 4 kHz, 6.39 at 6 kHz and 7.02 at 8 kHz.

The statistical results obtained with this study indicated that the SD in percentage hair cell damage calculated for M5 was 4.42 at 1 kHz, 3.32 at 1.5 kHz, 3.46 at 2 kHz, 5.07 at 3 kHz, 4.37 at 4 kHz, 6.64 at 6 kHz and 7.47 at 8 kHz. The SEM in percentage hair cell damage calculated for M5 was 1.04 at 1 kHz, 0.78 at 1.5 kHz, 0.82 at 2 kHz, 1.20 at 3 kHz, 1.03 at 4 kHz, 1.56 at 6 kHz and 1.76 at 8 kHz.

No previous studies were found where the OAE-gram as a combination of the DPOAE and TEOAE results were combined and analysed statistically. However, the DPOAEs and TEOAEs were analysed separately and high test-retest reliability was proven with the descriptive statistical results found. The following conclusions were made based on the results of the DPOAE and TEAOE results:

According to previous DPOAE studies (Beattie, Kenworthy, & Luna, 2003) the SEM at 550 Hz (~4.6dB) were nearly twice as large as those found for 1000Hz, 2000Hz and 4000Hz (~2.5dB). The short-term test-retest data suggest that there is a 95% probability that an individual’s true DPOAE will fall within 5 dB of the obtained distortion product at 1000-4000Hz and within 10dB at 550Hz. The SEM of the difference was calculated to assess whether two or more DPOAE measurements are significantly different. The data revealed that short-term differences (probe removed and subject retested on the same day or on different days) between two DPOAEs must exceed approximately 14 dB at 550 Hz and 7 dB at 1000-4000Hz to be statistically significant at the 0.05 level of confidence. According to these results, this study also confirms and indicates a

95% probability that an individual's true DPOAE will fall within 5dB of the distortion product at the frequencies measured between 1 kHz and 8 kHz.

Other previous studies examining DPOAEs have found greater variability at the higher frequencies (>8 kHz) for DPOAE level measurements (Dreisbach, Long, & Lees, 2006). The average DPOAE level differences between trials for the higher and lower frequencies for the four different stimulus level conditions was 5.15 (SD = 4.40 dB) and 2.80 (SD = 2.70 dB) dB, respectively. Individual subject analysis revealed that high frequency DPOAE levels varied no more than 10 dB for 87.5 and 83.1% of young adult subjects for the 70/75 and 60/50 dB SPL stimulus level conditions, respectively. These results were tested at the highest DPOAE results measured in this study with all (100%) of results obtained in this study exhibiting DPOAE levels that varied no more than 10dB with 70/75 dB SPL stimulus level conditions, confirming high test-retest reliability.

Previous studies analysing the DPOAE amplitudes in the default frequency range (1 to 6 kHz) between test and short-term retest were highly correlated (Ng & MacPherson, 2005), with average r=0.81. SEM was found to average 2.64 dB for DPOAE amplitude in the default frequency range. DPOAE amplitudes in the higher frequency range of 6.5 to 7 kHz were also significantly correlated between test and short-term retest; r ranged from 0.80 to 0.82 and SEM ranged from 2.59 to 3.04 dB. Repeatability of inhibition of DPOAE magnitudes was evaluated by Cronbach's alpha, ICC, SEM, and its 95% confidence interval and smallest detectable difference (Kumar, Methi, & Avinash, 2013), and also found to have high test-retest reliability with immediate remeasurements. The various resolution and retest conditions showed no significant differences in reliability. DPOAE measurement in the 6.5 to 7 kHz range can be considered a reliable tool for monitoring cochlear function in cases such as exposure to ototoxic medication or noise. According to the results obtained with the study in 2005, this study conducted indicated significant correlation between short and longer term test-retest reliability (up to one month after initial measures) as SEM ranges confirmed less variance than any of the previous studies conducted (SEM ranged from 0.51 to 2.37).

The SD and SEM calculated with the DPOAE results closely resemble the results obtained with previous studies indicating that the results fall within the predictable range expected of DPOAE amplitudes with high test-retest reliability. In the previous section the test-retest reliability of DPOAEs were proven for each separate time frame when DPOAEs were conducted. The descriptive statistics confirm the test-retest reliability of the DPOAEs for all participants with all the different measures calculated together. Therefore, high test-retest reliability for DPOAEs can be statistically confirmed.

In previous TEOAE studies (Vedantan & Musiek, 1991) the TEOAE test-retest reliability was measured and quantified according to the amplitude of the TEOAE echo obtained. The echo amplitude showed wide variation across ears and had a range of 20.9 dB, with the mean TEOAE 13dB. These results indicated reliable test-retest measures. With the amplitude measured in the current study, less variation with amplitude was calculated statistically (amplitude for this study range between 4.7 dB to 6.4 dB) as can be seen in Table 23, which confirms higher test-retest reliability for TEOAEs for this study than for the previous study (Vedantan & Musiek, 1991) conducted.

The SEM used by other authors (Beattie, Kenworthy, & Luna, 2003) for immediate retest measurements also correlated with the high test-retest reliability at all measurements during this study, which confirms the test-retest reliability. Differences in SD in TEOAEs across different participants were observed in previous studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) which was even smaller in this study, which did not influence overall test-retest reliability. Older studies (Franklin, McCoy, Martin, & Lonsbury-Martin, 1992) and more recent studies (Chan & McPherson, 2000) also found that the amplitude for TEOAEs at the high frequencies was reduced when measured over longer periods of time. The results of this study showed that the amplitudes were not reduced at the higher frequencies over longer periods of time indicating even higher test-retest reliability than previous studies conducted. Even with these statistical results indicating lower amplitudes the test-retest reliability was confirmed with previous studies, which proves that with more encouraging statistical results obtained in this study, confidence in TEOAEs across all the different time frames when TEOAEs were recorded, is confirmed.

Thus, the test-retest reliability of DPOAEs and TEOAEs were confirmed statistically according to the different sub-aims comprising of the different time-frames of measures where OAE-gram results closely resembled DPOAE and TEOAE test results. The test-retest reliability was further proven of DPOAEs and TEOAEs by descriptive statistics and analysing these results in detail and comparing them to previous studies using the same methods to quantify test-retest reliability.

It follows logically that if these results indicate statistically that the test-retest reliability is confirmed, that an algorhythm utilising the two different (highly reliable) test results should also show similar test-retest reliability. The test-retest reliability of the OAE-gram is explained by Table 25. The SEM and SD are similar to the results obtained with the DPOAE and TEOAE measures, though the unit of measurement for these OAE-gram results is the percentage of hair cell damage calculated by the OAE-gram algorhythm. These results, even though they are in a unit of measurement not used in any previous studies, show sufficient statistical significance (Field, 2009) to be confirmed as indicating high test-retest reliability for the OAE-gram. These results are further illustrated by the visual representation in Figure 15.

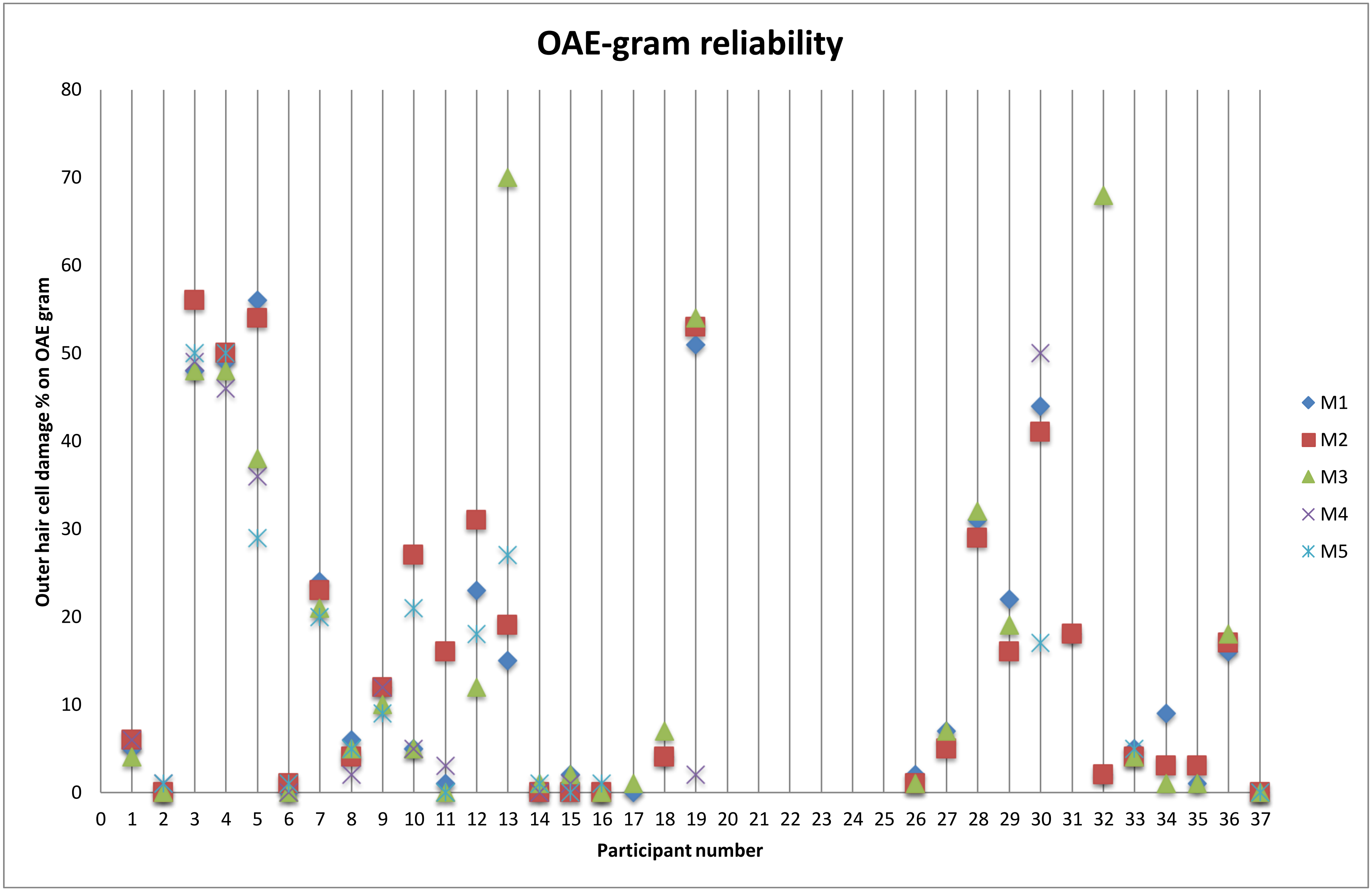


Figure 15: Test-retest reliability of OAE-gram

Inferential statistics for the OAE-gram results included one-way repeated measures Analysis of Variance (ANOVA) as seen in Figure 15, which illustrates the test-retest reliability of the OAE-gram. The graph indicates the average OAE-gram response in percentage of outer hair cell damage for each participant at all of the different time intervals measured. The deductions made from this graph confirm the limited variance in OAE-gram response at the different time intervals for each participant. This highlights the high test-retest reliability of OAE-gram at the different time intervals because a minimal change over time is observed between the different measures for each participant. Outliers illustrated by this graph are limited to three participants (number 13, 19, and 32). These participants had significant variance in one of the tests measured (respectively M3, M4, and M3). A single measure that differed from other measures, even though the variance observed was significant, did not affect the overall reliability of the measures (Field, 2009). Test-retest reliability was not only measured by variance between each participant, but also between the results classified as a single group per time interval and measured with ANOVA as can be seen by the p-value (Sig) also calculated in Table 26.

Table 26: OAE-gram ANOVA statistical analysis

OAE-gram ANOVA

		Sum of Squares	**df**	**Mean Square**	**F**	**p**
1.0 kHz	Between Groups	1651.207	4	412.802	1.089	.363
	Within Groups	87587.767	231	379.168		
	Total	89238.974	235			
1.5 kHz	Between Groups	2593.841	4	648.460	2.336	.056
	Within Groups	64124.418	231	277.595		
	Total	66718.260	235			
2.0 kHz	Between Groups	2155.504	4	538.876	1.651	.162
	Within Groups	75380.263	231	326.321		
	Total	77535.767	235			
2.5 kHz	Between Groups	2177.484	4	544.371	1.249	.291
	Within Groups	100679.574	231	435.842		
	Total	102857.058	235			
3.0 kHz	Between Groups	1016.918	4	254.229	.638	.636
	Within Groups	92092.946	231	398.671		
	Total	93109.864	235			
3.5 kHz	Between Groups	1505.151	4	376.288	.881	.476
	Within Groups	98648.660	231	427.050		
	Total	100153.810	235			
4.0 kHz	Between Groups	1301.203	4	325.301	.493	.741
	Within Groups	152367.608	231	659.600		
	Total	153668.811	235			
4.5 kHz	Between Groups	1450.862	4	362.715	.525	.717
	Within Groups	159582.241	231	690.832		
	Total	161033.102	235			
5.0 kHz	Between Groups	1210.446	4	302.612	.391	.815
	Within Groups	178745.364	231	773.789		
	Total	179955.810	235			
5.5 kHz	Between Groups	2734.266	4	683.567	.834	.505
	Within Groups	189328.569	231	819.604		
	Total	192062.836	235			
6.0 kHz	Between Groups	2224.305	4	556.076	.670	.614
	Within Groups	191862.387	231	830.573		
	Total	194086.692	235			
6.5 kHz	Between Groups	861.476	4	215.369	.311	.871
	Within Groups	160164.740	231	693.354		
	Total	161026.216	235			

In Table 26 ANOVA was used to determine the changes in OAE-gram results between the measures M1 to M5. For OAE-gram results the mean signals were compared across time by performing a one-way repeated ANOVA at each of the frequencies measured. Further descriptive statistical calculations were made for each of the tests to give a clear picture of test-retest reliability. For each frequency, the different measurements at the time intervals specified were compared between groups and within groups, according to the sum of squares, mean square, F value calculated, and statistical significance (Sig.). The null hypothesis states that the means do not differ across the different time intervals, while the alternative hypothesis states that there is indeed a significant difference between at least two of the time measurements. The p values (Sig) are 0.363 at 1 kHz, 056 at 1.5 kHz, 0.162 at 2 kHz, 0.291 at 2.5 kHz, 0.636 at 3 kHz, 0.476 at 3.5 kHz, 0.741 at 4 kHz, 0.717 at 4.5 kHz, 0.815 at 5 kHz, 0.505 at 5.5 kHz, 0.614 at 6 kHz and 0.871 at 6.5 kHz. Since all the p-values (Sig) are greater than 0.05 at all the frequencies measured, it can be concluded that there are no statistical significant differences in the mean signals across all different time intervals (Field, 2009). One-way repeated measures ANOVA did not show a significant change in emissions between OAE-gram measurements across frequencies. With all OAE-gram tests included in this study (immediate re-testing, re-testing after one hour, re-testing after one week and re-testing after one month) there was no significant difference in the time lapse to influence the test-retest reliability. The different tests over time showed no significant difference as the significance value (Sig) calculated was in every case larger than 0.05, indicating no statistical significance in the average results calculated for all tests (Laerd, 2015; Dallal, 2013). This high value calculated between groups at all different frequencies tested confirms the high test-retest reliability of OAE-grams at different time intervals. The null hypothesis states that the means do not differ across M1 to M5, while the alternative hypothesis states that there is indeed a significant difference between at least two of the time measurements. Since the p-values have been found to be greater than 0.05 at all the frequencies, it can be concluded that there are no significant differences in the mean results of OAE-gram measures across M1 – M5. The ICC was explored further for the OAE-gram for single measures and for average measures.

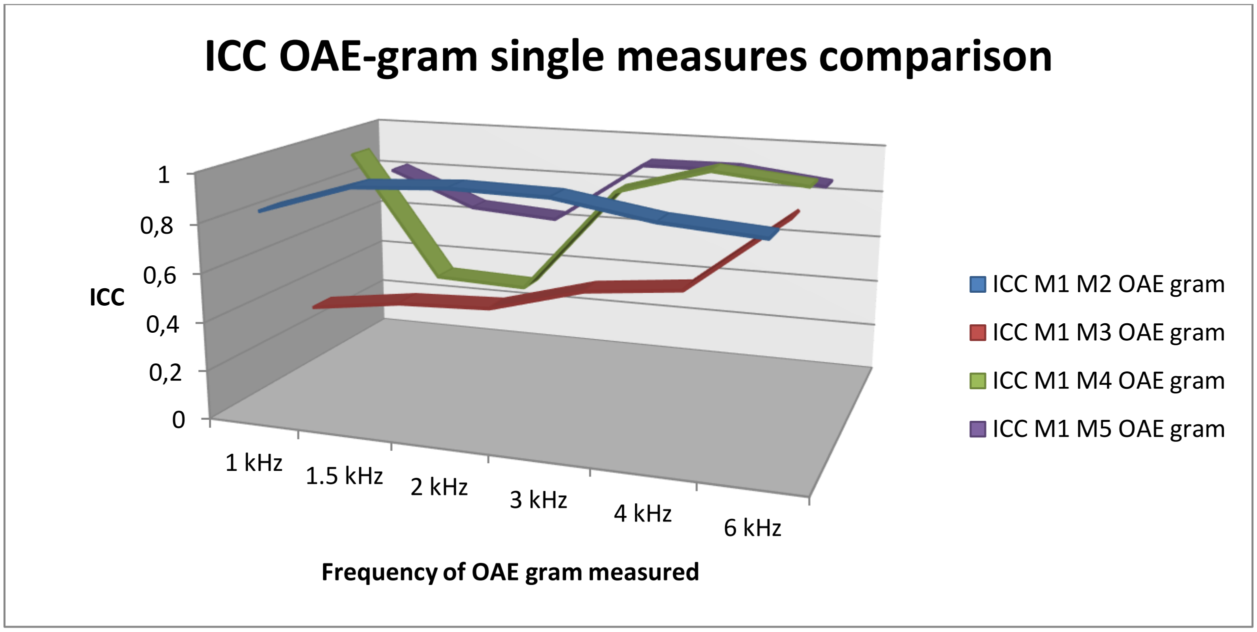


Figure 16: ICC OAE-gram single measures comparison

In Figure 16 the ICC single measures are compared over all the different time intervals for OAE-grams. The different OAE-grams are all compared to the initial measured (M1) OAE-gram to create an ICC. ICC single measures results refer to the calculation of ICC for each individual participant. This means that the test-retest reliability of measures per single participant is estimated by these results. It describes how strongly units in the same group resemble each other.

This gives an accurate indication of the reliability of data collected during this study for each participant at different time intervals.

The ICC is used to quantify the degree to which each participant's individual OAE-gram results resemble the results of the same participant measured at other points in time.

An ICC close to 1 indicates high correlation between different results. As can be seen in Figure 16, the ICC indicated a high reliability between tests M1 and M2, and M1 and M5 at low and high frequencies measured. The ICC obtained for M1 compared to M3 did not indicate the same high reliability at lower frequencies as the other measures. However, ICC remained consistently high for high frequency measures of 6 kHz and above. The ICC at 1.5 kHz and at 2 kHz for M1 compared to M4 also indicated lower reliability. These results could not be explained by specific variable factors present. Overall ICC still indicates high test-retest reliability with OAE-grams, but serves to emphasise that any single test should not be viewed in isolation and should be correlated, and results confirmed with other measures at different points in time. The high ICC at all other test points and specifically at the high frequencies confirms the high test-retest reliability.

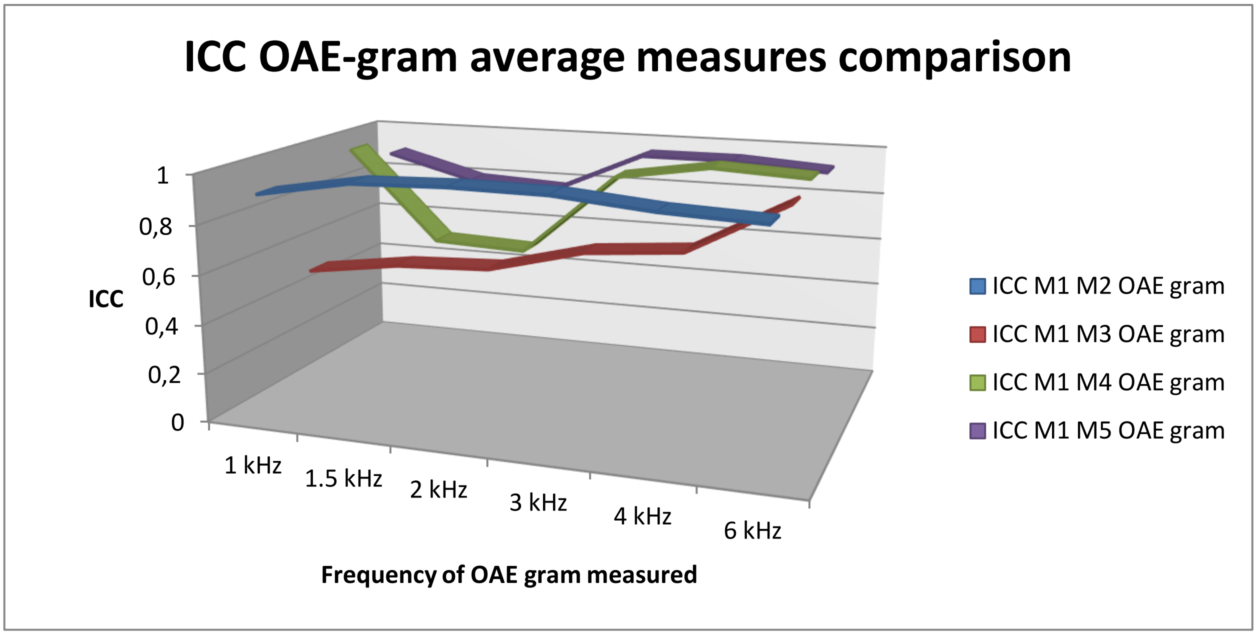


Figure 17: ICC OAE-gram average measures comparison

Figure 17 illustrates the comparison of the average measures of ICC of OAE-grams at different frequencies at all of the different time intervals. . The different OAE-grams are all compared to the initial measured (M1) OAE-gram to create an ICC. ICC average measures results refer to the calculation of ICC for the group of participants at each point the OAE-grams were measured. This means that the test-retest reliability of measures within the group as a whole is estimated by these results. This gives an accurate indication of the reliability of data collected during this study for the measures collected for the group as a whole at different time intervals.

The ICC is used to quantify the degree to which the group of participants' OAE-gram results resemble the results measured at other points in time. An ICC close to 1 indicates high correlation between different results. As can be seen in Figure 17, the ICC indicated a high reliability between tests M1 and M2, and M1 and M5 at low and high frequencies measured.

The ICC obtained for M1 compared to M3 did not indicate the same high reliability at lower frequencies as the other measures. However, ICC remained consistently high for high frequency measures of 6 kHz and above. The ICC at 1.5 kHz and at 2 kHz for M1 compared to M4 also indicated lower reliability. The results of the average measures ICC correlate well with the results of single measures ICC, which again emphasises that any single test should not be viewed in isolation and should be viewed in conjunction with other test results, and all results should be confirmed with other measures at different points in time.

4.3.4 Reliability of results

After statistical analyses were conducted some discrepancies were noticed in the data. These discrepancies could possibly explain some test-retest reliability results and further enhance the credibility of results. Three different participants had some outliers in results measured. These participants were 13, 19, and 32 at their measures of M3, M4, and M3 respectively. Upon further inspection of test results, some details were noted which could explain the variance observed between their different tests, and also the overall test results as these test results were not excluded. Participant 13 at M3 had increased noise levels at the lower frequencies when the DPOAE and TEOAE tests were conducted. No indication was given of high noise levels at the time of testing, and results were deemed reliable at the time. For participants 19 and 32 no specific explanation could be found, although emissions were reduced for both DPOAE and TEOAE results at the lower frequencies at M4 and M3. These three participants were not tested on the same day, thus results were probably not influenced by the same external variable factors present at a specific time. These three results could possibly provide an explanation for the ICC variability present at M3 and M4 for the DPOAEs, TEOAEs and OAE-gram indicating somewhat reduced correlation as discussed extensively above. These discrepancies are indicative of a possible real-life testing environment where not all variables can be controlled. Thus, despite the proven test-retest reliability, these observations emphasise that OAE results must be interpreted in correlation with other results and not in isolation as absolute measures of cochlear function.

4.4 Summary

All of the tests conducted were aimed at answering the research question. In this research study the test-retest reliability of DPOAEs and TEOAEs had to be determined. From the results obtained it can be concluded that if all relevant variable factors are excluded (mainly noise), DPOAEs and TEOAEs are reliable and valid over test-retest intervals of time ranging from a few minutes up to one month.

With all tests included in this study namely immediate re-testing, re-testing after one hour, re-testing after one week, and re-testing after one month, there was no significant difference in the test-retest reliability. In fact these different tests showed no significant difference over time as the F value calculated was in every case larger than 0.05, indicating that there was no statistical significance in the average results calculated for all tests (Laerd, 2015; Laerd, 2015; Dallal, 2013).

The DPOAE and TEOAE measures taken during this research study in low noise environment indicated high test-retest reliability and a high correlation between test results over different periods of time.

The ICC for DPOAEs, TEOAEs and OAE-gram indicates little variability between different tests, as well as the SD and ANOVA statistical analysis. These results indicated high test-retest reliability with small deviations between the different tests executed.

5. CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

The main aim of this study was to investigate and determine the test-retest reliability of DPOAEs and TEOAEs and the OAE-gram generated from a combination of the DPOAE and TEOAE results. The reliability was determined by data collection of DPOAEs and TEOAEs, and subsequent generation of the OAE-gram of participants with normal hearing over different periods of time, namely within five minutes of initial testing, within one hour, within one week and within one month of the initial testing taking place. The implications of this research study are presented as well as making recommendations for future research.

5.1 Summary of research findings

The research process involving data collection was successful in answering the research question and obtaining results to get a conclusion on the main aim of this study. The goal of this study was to determine the test-retest reliability of DPOAEs and TEOAEs – this reliability was determined by different OAE measures as well as the generated OAE-gram in normal hearing participants over a set period of time and comparing these different measures to determine whether a correlation exists or not. The time frame pre-determined for this study was:

M1 Initial OAE measures,

M2 OAE measures with immediate re-fitting of the probe (within five minutes of initial test)

M3 OAE measures one hour after the initial test

M4 OAE measures one week after the initial test

M5 OAE measures one month after the initial test.

All of these results obtained from all the participants over the five different measures were then analysed statistically in order to make inter-participant and intra-participant conclusions.

This research study revealed the following results pertaining to test-retest reliability of DPOAEs and TEOAEs:

- There was no statistically significant difference in results when the test-retest reliability was compared over the five different periods in time
- TEOAEs had statistically significant higher test-retest reliability when compared to DPOAEs
- When considering all research participants, the test-retest reliability of OAE results is highest for tests repeated within five minutes of the initial test (M1 compared to M2), and lowest for tests repeated one hour after initial measure (M1 compared to M3).
- Statistical differences were noted with regards to the test-retest reliability of DPOAEs when comparing M1 to M3 at the low frequencies. This important difference could possibly be explained by the identification of some outliers after statistical analysis were conducted. These single statistical differences, did not, however, influence the overall test-retest reliability as calculated through several different tests. No statistical significant difference was evident between the other tests conducted at different time intervals indicating high test-retest reliability.

- A hypothesis test was selected, which revealed that there is not a statistically significant difference between extreme OAE measurements for the 30 participants included in this study. This proves that OAE measurements and the generated OAE-gram are reliable and not merely by chance.

5.2 Implications of study

There are continuous advances in the field of audiology and specifically in OAE measures. These new measures create a continuous flow of new information on OAEs and many new research questions arise from these developments. Numerous studies investigated the further development of technology and OAEs. One of the main goals of this study was to determine the test-retest reliability of DPOAEs and TEOAEs and using the algorhythm of the OAE gram with the combined measurements within a group of normal hearing participants. This goal was selected as the advances in technology create vast new fields and researchers move from one subject to another, but for clinical relevance the reliability of any measure needs to be proven.

Therefore, this study investigated the test-retest reliability of DPOAE and TEOAE measures as well as the resulting OAE-gram over time. The results of this study indicated that these measures are valid and have shown that there were no significant differences and therefore, the test procedure and the results are reliable.

The current study therefore, provided much-needed information about the reliability of these measures. The confirmation of the test-retest reliability over any period of time improves the clinical relevance of these tests and should result in an increased confidence in DPOAEs, TEOAEs and the OAE-gram among clinicians as valuable source of information on functioning of the outer hair cells of the cochlea.

5.3 Analysis of strengths of study

An analysis of the strengths of the study provides information about the positives encountered while this study was conducted, and could provide guidance for future studies to further make use of these strengths identified.

- This study was performed with a longer period of time between the different measures than previous studies. Previous studies examined the test-retest reliability up to a week after initial measures were performed, but this study illustrated that test-retest reliability remained constant over a longer period of time between different measures.
- Use of a single DPOAE and TEOAE system ensured that external variables with possible effects on results are minimized. As a single OAE measurement system was used during this study, results were statistically quantifiable and could be compared over time. Different protocols or measurement environments were minimized by using a single system for all measures.
- Statistical comparison between different time intervals proved significance of test-retest reliability as it might be used clinically as monitoring device, and this relates well with real-life application of DPOAEs, TEOAEs and the OAE-gram.
- Similarity between the research design and statistical analysis between this study and previous research projects conducted facilitates comparison with existing literature.

- Use of a single tester yielded reliable results as interrater reliability played no role in the results obtained because a single researcher was responsible for all measurements made.

5.4 Analysis of weaknesses of study

Analysis of weaknesses of the study provides insight into some of the weaknesses encountered with the study performed, and could increase the value of results if future studies take these weaknesses into consideration.

- Generalization of the results to the general population may be limited as only normal hearing adults between 18 to 25 years of age were used in the study.
- Researcher or clinician bias as only one researcher was used in this study and any differences in testing methods that the single researcher may not be aware of with different clinicians have not been accounted for.
- Test subjects failing to arrive for follow-up tests: Some participants did not arrive for all or several of their follow-up tests producing limited sets of data.
- Limited number of participants was used and the study could be expanded to include a larger participation basis.

5.5 Recommendations for future research

The data collection for this project provided further insight into the test-retest reliability of DPOAEs and TEOAEs. Studying the results from different participants and analysing the data quantitatively and statistically have provided sufficient information to draw the conclusions that these measures are reliable and valid in test-retest circumstances even over a prolonged period of time. However, there are still areas that can be explored. These areas are summarized as follows:

In the methodology, the following recommendations can be made:

- Change the clinician conducting the research to determine whether the clinician and individual probe fitting might have an influence on OAE results. In the current study only one clinician was used due to lack of manpower and as this was not the main aim of the study. Using only one clinician might give one-sided results. It should be determined whether the OAE results can be influenced by changing the clinician conducting the test.
- A sample including male and female participants could be used in future studies for comparison to the general population.
- Using both ears of each participant to increase the amount of results obtained and to observe whether significant differences can be seen between the two ears of participants.

New research questions can arise from this study that need to be answered. The following new research studies are proposed:

- Include participants with a hearing loss in the study. This would further determine the test-retest reliability in a population with hearing loss in comparison with a normal hearing population. The participants can further be grouped according to degree of hearing loss to compare the percentage of outer hair cell damage. In this study the main aim was to determine the test-retest reliability on a normal hearing population as this is the most homogenous population for initial testing, but this study can be further expanded in future to include participants with hearing loss.

- Complete a longitudinal study with the same participants as this study in a few years' time to determine whether participants with higher percentage of hair cell damage are more prone to hearing loss later in life. The main aim of this research study was to determine test-retest reliability amongst a normal hearing population. There was much inter-participant variance within this homogenous group which can not be explained by pure tone thresholds. In a longitudinal study using the same participants it can be determined whether lowered OAE amplitudes early in life could be a predictor of hearing loss later in life.

- To be used for monitoring hearing loss in different populations, such as ototoxicity monitoring and monitoring of NIHL in specific populations. These tests can be applied in a population where longitudinal comparison over time may indicate deterioration of inner ear function and to monitor and minimize this deterioration over a period of time.

5.6 Final conclusion

This research project has successfully answered the research question raised, namely:

"What is the test-retest reliability of DPOAEs and TEOAEs in a normal hearing population?"

The main aim and sub aims selected in the first and second chapters have been reached during the course of this study. The test-retest reliability of DPOAEs, TEOAEs and the OAE-gram were studied and analysed through quantitative and statistical procedures while still considering qualitative aspects and noting detail of each participant in the study. The study generated new results with regards to high DPOAE and TEOAE test-retest reliability. The proven reliability of these tests should increase the usage of these measures in clinical audiology practice, and these results provide support for a wider application of these measures in audiology in general.

6. REFERENCES

Babbie, E. R. (1979). *The practice of social research* (2nd ed.). Belmont: Wadsworth.

Backus, C. B. (2007, March). Bias due to noise in otoacoustic emission measurements. *Journal of Acoustical Society of America*, *121*(3), 1588-1603.

Beattie, R. C., Kenworthy, O. T., & Luna, C. A. (2003). Immediate and short-term reliability of distortion product otoacoustic emissions. *International Journal of Audiology*, *42*(6), 348 - 354.

Bekesy, G. (1960). *Experiments in hearing*. New York: McGraw-Hill.

Berger, E. H., Royster, L. H., Royster, J. D., Driscoll, D. P., & Layne, M. (Eds.). (2003). *The Noise Manual* (5th ed.). California: AIHA.

Bleech, J., & Beattie, R. C. (2011). Effects of Sample Size on the Reliability of Noise Floor and DPOAE. *British Journal of Audiology*, 305-309.

Bless, C., & Higson-Smith, C. (2004). *Fundamentals of Social Research Methods*. Cape Town, South Africa: Juta.

Bockstael, A., Keppler, H., Dhooge, I., D'haenans, W., Maes, L., Philips, B., & Vinck, B. (2008). Effectiveness of hearing protector devices in impulse noise verified with transiently evoekd and distortion product otoacoustic emissions. *International Journal of Audiology*(47), 119-33.

Brownell, W. E. (1990). Outer hair cell electromotility and otoacoustic emissions. *Ear and Hearing*, 82-92.

Campbell, K. C. (2014, March 15). *Otoacoustic Emissions*. (A. D. Meyers, Editor) Retrieved 2016, from Medscape: http://emedicine.medscape.com/article/835943-overview

Chan, R. H., & McPherson, B. (2000). Test-retest reliability of tone-burst-evoked otoactoustic emissions. *Acta Otolaryngology*(120), 825-834.

Dallal, G. E. (2013). *The little Handbook of Statistical Practice*. Retrieved from jerrydallal.com: http://www.jerrydallal.com/LHSP/dof.htm

Dannhauer, J. L. (1997). How Otoacoustic Emissions Testing Can Change Your Audiology Practice. *The Hearing Journal, 50*(4).

D'haenans, W., Vinck, B., De Vel, E., Maes, L., Bockstael, A., Keppler, H., . . . Dhooge, I. (2008). Auditory steady state responses in normal hearing adults: A test-retest reliability study. *International Journal of Audiology*(47), 489-498.

Dhooge, I., Dhooge, C., Geukens, S., De Clerck, B., De Vel, E., & Vinck, B. (2006). Distortion product otoacoustic emissions: An objective technique for the screening of hearing loss in children treated with platin derivatives. *International Journal of Audiology*(45), 337-343.

Dreisbach, L. E., Long, K. M., & Lees, S. E. (2006). Repeatability of High-Frequency Distortion-Product Otoacoustic Emissions in Normal-Hearing Adults. *Ear and Hearing, 5*(27), 466-479.

Erasmus, M., & Grové, C. L. (2008). Gehoorkonserveringsprogram - Hearing Coach: Sensitiwiteit en spesifisiteit van TOOAE's en DPOAE's by normaalhorende volwassenes. *Unpublished undergraduate study, University of Pretoria*.

Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: SAGE Publications.

Franklin, D. J., McCoy, M. J., Martin, G. K., & Lonsbury-Martin, B. L. (1992). Test-retest reliability of distortion product and transiently evoked otactoustic emissions. *Ear and Hearing, 13*(6), 417-489.

Gelfand, S. A. (2009). *Essentials of Audiology* (3rd ed.). New York: Thieme.

Hall, J. W. (2000). *Handbook of otoacoustic emissions*. San Diego, California: Singular Publishing Group Inc.

Hall, J. W. (2015). Evidence-Based Clinical applications of OAEs in Children and Adults.

Hall, A. J., & Lutman, M. E. (1999). Methods for early identification of noise-induced hearing loss. *International Journal of Audiology*, *38*(5), 277-280.

Hall, J. W., & Mueller, H. G. (1997). *Diagnostic Audiology principles, procedures and protocols: Audiologists' Desk Reference*. San Diego, California: Singular Publishing Group, Inc.

Havelock, D., Kuwano, S., & Vorländer, M. (Eds.). (2008). *Handbook of Signal Processing Acoustics*. New York: Springer.

Hearing Coach Manual to interpret Oto-Acoustic emissions. (2007).

Hearing Coach User Manual. (2007).

Heitmann, J., Waldmann, B., & Plinkert, P. K. (1996). Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure. *European archives of Oto Rhino Lanryngology and Head & Neck*(253), 167-171.

Hoth, S. (2005). On a possible prognostic value of otoacoustic emissions: a study on patients with sudden hearing loss. *European archives of Oto Rhino Lanryngology and Head & Neck*(262), 217-224.

Hoth, S., Gudmunsdottir, K., & Plinkert, P. (2010). Age dependence of otoacoustic emissions: the loss of amplitude is primarily caused by age-related hearing loss. *European archives of Oto Rhino Lanryngology and Head & Neck*(267), 679-690.

Johnson, C. D., & Seaton, J. B. (2012). *Educational Audiology Handbook*. New York: Delmar.

Johnson, C. E., & Danhauer, J. L. (2002). *Handbook of Outcomes Measurement in Audiology*. New York: Delmar learning.

Katz, J. (2014). *Handbook of Clinical Audiology* (7th ed.). Philadelphia: Lipincott Wiliiams & Wilkins.

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America*, *64*, 1386-1391.

Keppler, H., Dhooge, I., Maes, L., D'haenans, W., Backstael, A., Philips, B., . . . Vinck, B. (2010). Transient-evoked and distortion product otoacoustic emissions: A short-term test-retest reliability study. *International Journal of Audiology*(49), 99-109.

Kumar, A. U., Methi, R., & Avinash, M. (2013). Test/retest repeatability effect of contralateral acoustic stimulation on the magnitudes of distortion product ototacoustic emissions. *Otology Neurotology*, *123*(2), 463-471.

Laerd. (2015). *How do I interpret data in SPSS for a 1 way between subjects ANOVA?* Retrieved from www.statistics.laerd.com: http://statistics-help-for-students.com/How_do_I_interpret_data_in_SPSS_for_a_1_way_between_subjects_ANOVA.htm

Laerd. (2015). *Statistics.laerd.com*. Retrieved from One way anova statistical guide: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-4.php

Leedy, P. D., & Ormrod, J. E. (2005). *Practical Research: Planning and Design*. Englewood Cliffs, New Jersey: Pearson Merrill Prentice Hall.

Lonsbury-Martin, B. L., McCoy, M. J., Whitehead, M. L., & Martin, G. K. (1993). Clinical testing of Distortion-Product Otoacoustic Emissions. *Ear and Hearing*, *1*, 11-22.

Lycke, M., Maes, E., & Michem, E. (2011). *Bruikbaarheid van OAE als screeningstechniek*. University of Gent.

Manley, G. A., Fay, R. R., & Popper, A. N. (Eds.). (2008). *Active Processes and Otoacoustic emissions*. New York: Springer.

Maxwell, D. L., & Satake, E. (2006). *Research and Statistical Methods in Communication Sciences and Disorders*. Boston: Singular Publishing.

Mc Pherson, B., Li, S. F., Shi, B. X., Tang, J. L., & Wong, B. Y. (2006, June). Neonatal hearing screening: evaluation of tone-burst and click-evoked otoacoustic emission test criteria. *Ear Hear*, *3*(27), 256-262.

McFadden, D., Martin, G. K., Stagner, B. B., & Maloney, M. M. (2009, January). Sex differences in distortion-product and transient-evoked otoacoustic emissions compared. *Journal of Acoustical Society of America*, *125*(1), 239-245.

Mouton, J. (2001). *How to succeed in your Master's and Doctoral studies*. Pretoria: Van Schaik Uitgewers.

Neuman, W. L. (1997). *Social Research Methods: Qualitative and Quantitative Approaches* (3rd ed.). Boston: Allyn and Bacon.

Ng, I. H., & MacPherson, B. (2005). Test-Retest Reliability of Distortion Product Otoacoustic Emissions in the 1 to 7 kHz Range. *Audiological Medicine*, *3*(2), 108-115.

Orlikoff, R. F., Schiavetti, N., & Metz, D. E. (2014). *Evaluating Research in Communication Disorders*. Englewood Cliffs, New Jersey: Pearson Merrill Prentice Hall: Pearson.

Pickles, J. O. (2013). *An Introduction to the Physiology of Hearing* (4th ed.). Boston: BRILL.

Prieve, B. A., & Fitzgerald, T. S. (2002). Otoacoustic emissions. In J. Katz (Ed.), *Handbook of Clinical Audiology* (5th ed., pp. 440-468). London: Lippincott Williams & Wilkens.

Probst, R., Harris, F. P., & Hauser, R. (1993). Clinical monitoring using otoacoustic emissions. *British Journal of Audiology, 27*, 236-243.

Robinette, M. S., & Glattke, T. J. (2007). *Otoacoustic emissions Clinical Applications* (3rd ed.). New York: Thieme Medical Publishers.

Robinson, D. W., & Sutton, G. J. (1979). Age effect in Hearing - A comparative analysis of published threshold data. *Audiology, 18*, 320-334.

Roester, R. J., Valente, M., & Hosford-Dunn, H. (2000). *Audiology Diagnosis*. New York: Thieme Medical Publishers, Inc.

Science, P. S. (2013). *The ANOVA Table*. Retrieved from PennState Eberly College of Science: https://onlinecourses.science.psu.edu/stat414/node/218

Shupak, A., Tal, D., Sharoni, Z., Oren, M., & Ravid. (2007). Otoacoustic emissions in early noise-induced hearing loss. *Otology & Neurotology, 27*, 745-752.

Sockalingham, R., Kei, J., & Ho, C. D. (2007). Test-retest reliability of distortion-product otoacoustic emissions in children with normal hearing: A preliminiary study. *International Journal of Audiology, 46*(7), 351 - 354.

South African National Standard [SANS] 10154-1. (2004). *Calibration of pure-tone audiometers. Part 1: Air conduction.* Pretoria: South African Bureau of Standards.

South African National Standard [SANS] 10154-2. (2004). *Calibration of pure-tone adiometers. Part 2: Bone condution.* Pretoria: Southa African Bureau of Standards.

South African National Standards [SANS] 10182. (2006). *The measurement and assessment of acoustic environments for audiometric tests.* Pretoria: South African Bureau of Standards.

Steenkamp, R. J. (2008, May). Effective hearing conservation demands new technology and re-engineering. *South African Journal of Industrial Engineering*, *19*(1), 215-229.

Struwig, F. W., & Stead, G. B. (2001). *Planning, designing and reporting research.* Cape Town, South Africa: Pearson Education South Africa.

Stuart, A., Passmore, A. L., Culbertson, D. S., & Jones, S. M. (2009). Test retest reliability of low-level evoked distortion product otoacoustic emissions. *Journal of Speech Language and Hearing Research*, *52*(3), 671-681.

Stuart, A., Passmore, A. L., Culbertson, D. S., & Jones, S. M. (2009). Test–Retest Reliability of Low-Level Evoked Distortion Product Otoacoustic Emissions. *Journal of Speech, Language, and Hearing Research*, *52*, 671-681.

Tlumak, A. I., & Kileny, P. R. (2001). Parameters that affect the measurement of otoacoustic emissions. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *9*, 279 - 283.

Vedantan, R. M., & Musiek, F. E. (1991). Click evoked oto acoustic emissions in adult subjects: standard indices and test-retest reliability. *American Journal of Otology*, *12*(6), 435-442.

Vinck, B. M., Van Cauwenberge, P. B., Corthals, P., & De Vel, E. (1998). Multi-variant Analysis of Otoacoustic Emissions and Estimation of Hearing Thresholds: Transient Evoked Ototacoustic Emissions. *International Journal of Audiology*, *37*(6), 315 - 334.

Vinck, B. M., Van Cauwenberge, P. B., Leroy, L., & Corthals, P. (1999). Sensitivity of Transient Evoked and Distortion Product Otoacoustic Emissions to the Direct Effects of Noise on the Human Cochlea. *International Journal of Audiology*, *38*(1), 44 - 52.

Vinck, B., De Vel, E., Xu, Z., & Van Cauwenberge, P. B. (1996). Distortion product otoacoustic emissions: A normative study. *Audiology*(35), 231245.

Wagner, W., & Plinkert, P. K. (1999). The relationship between auditory threshold and evoked otoacoustic emissions. *European archives of Oto Rhino Lanryngology and Head & Neck*(256), 177-188.

Wagner, W., Heppelmann, G., Vonthein, R., & Zenner, H. P. (2008). Test-Retest Repeatability of Distortion Product Otoacoustic Emissions. *Ear & Hearing*, *29*, 378 - 391.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*(19), 231-240.

www.adler.edu. (2009, October). Retrieved from (Online dictionary).

www.ec.wmich.edu/glossary. (2009).

Yost, W. A. (2006). *Fundamentals of Hearing* (5th ed.). Burlington: Elsevier.

Zurich, E. (2014). *How to calculate an ANOVA table*. Retrieved from stat.ethz.ch: https://stat.ethz.ch/education/semesters/as2010/anova/ANOVA_how_to_do.pdf

7. APPENDICES

7.1 Appendix A: Questionnaire

Appendix A

Questionnaire

A) Personal information:

Name: ____
Surname: ____
Date of birth: ____
Age: ____
Gender: Male ☐ Female ☐

B) Information about hearing:

Do you have any complaints about your hearing?
YES __________ NO__________
If YES, please specify the type of problems that you currently experience:

____
____
____

Have you ever had your hearing tested?
YES __________ NO__________
If YES, what was the reason and results of the hearing evaluation?

____
____

Do you have any family members with hearing problems?
YES __________ NO__________
If YES, please state you relation to this family member (e.g. sister, father, cousin)

____
____

Any other comments or information about your hearing:

____
____
____
____

C) Noise exposure:

Are you currently, or have you previously been, exposed to excessive noise? Please state YES/NO, when you were exposed to the particular type of noise and what the duration of the exposure was:

Type of noise exposure	YES	NO	When(and how often) are you exposed to this type of noise?	How long are you exposed to this type of noise at a time?
Firearm use				
Industrial machinery				
Loud music (e.g. clubs)				
IPod/MP3 player				
Other:				

Have you ever experienced a ringing in your ears after exposure to noise?

YES __________ NO__________

If YES, please describe the ringing (e.g. high/low, intermittent/continuous, ringing/hissing)

____

____

Do you use hearing protection when you are exposed to noise?

YES __________ NO__________

If YES, please describe the type of hearing protection, and how regularly you use it:

____

____

Have you been exposed to noise in the past 48 hours?

YES __________ NO__________

If YES, please describe the type of noise/situation:

____

____

____

____

D) Medical background:

Please indicate whether you have had any of the following illnesses, and the approximate age at which you had this illness:

Medical condition	**Yes**	**No**	**Age**
Allergies			
Meningitis			
Ventilation tubes (*grommets)*			
Mumps			
German measles			
Migraine			
Ringing in the ears (tinnitus)			
Middle ear infection & treatment			
Dizziness			
Any operations: (please specify)			

Are you currently on any medication?
YES __________ NO__________
If YES, please specify all medication you are currently using:

____
____
____

Have you ever experienced that medication has had an influence on your hearing?
YES __________ NO__________
If YES, please specify the type of medication, and length of use as well as the effect it had on your ears:

____
____
____

Thank you for your willingness to participate in this study. Be assured that it is truly appreciated!

*PLEASE NOTE:
All information in this questionnaire will only be used for research purposes, and personal information will be handled confidentially.

Appendix A

Vraelys

A) Persoonlike inligting:

Naam: ____
Van: ____
Geboortedatum: ____
Ouderdom: ____
Geslag: Manlik ☐ Vroulik ☐

B) Inligting oor gehoor:

Het u tans enige klagtes rondom u gehoor?
JA __________ NEE__________
Indien JA, noem asb. die spesifieke probleme wat u met u gehoor ondervind:
____
____
____

Het u al voorheen `n gehoortoets ondergaan?
JA __________ NEE__________
Indien JA, wat was die rede en/of uitslag van die gehoorevaluasie?
____
____

Het u `n familielid met gehoorprobleme?
JA __________ NEE__________
Indien JA, spesifiseer asb. u verwantskap met die familielid (bv. suster, vader, neef)
____
____

Enige ander kommentaar rakende u gehoor?
____
____
____
____

C) Geraasblootstelling:

Word u tans, of is u in die verlede al aan harde geraas blootgestel? Merk asb. ja/nee, wanneer u aan die geraas blootgestel is, en hoe lank hierdie blootstelling was.

Tipe geraasblootstelling	**Ja**	**Nee**	**Wanneer (en hoe gereeld) is u aan hierdie tipe geraas blootgestel?**	**Hoe lank is u op een slag aan hierdie geraas blootgestel**
Vuurwapen gebruik				
Fabrieksmasjinerie				
Harde musiek (bv. klubs)				
IPod/MP3-spelers				
Ander:				

Het u al `n suising in u ore ondervind na afloop van hierdie geraasblootstelling?
JA __________ NEE__________
Indien JA, beskryf asb. die suising (bv. hoog/laag, aanhoudend/onderbroke)
____
____

Gebruik u gehoorbeskerming wanneer u aan geraas blootgestel word?
JA __________ NEE__________
Indien JA, beskryf asb. die tipe gehoorbeskerming, en hoe gereeld u dit gebruik
____
____

Is u in die afgelope 48 uur aan geraas blootgestel?
JA __________ NEE__________
Indien JA, beskryf asb. die tipe geraas of situasie waaraan u blootgestel is:
____
____

D) Mediese agtergrond:

Dui asb. aan of u voorheen enige van die volgende siektetoestande gehad het, en die ouderdom waartydens u hierdie siektetoestande onder lede gehad het:

Mediese toestand	Ja	Nee	Ouderdom
Allergieë			
Breinvliesontsteking (meningitis)			
Ventilasiebuisies (*grommets)*			
Pampoentjies			
Duitse masels			
Migraine			
Suising in die oor (tinnitus)			
Middeloorontsteking & behandeling			
Duiseligheid			
Enige operasies: (spesifiseer)			

Gebruik u tans enige medikasie?
JA __________ NEE__________
Indien JA, spesifiseer asb. alle medikasie wat u tans gebruik:

____

____

____

____

Het u in die verlede ervaar dat enige medikasie `n invloed op u gehoor gehad het?
JA __________ NEE__________
Indien JA, spesifiseer asb. die tipe medikasie, tydperk wat u dit moes gebruik, asook die jaar en die invloed wat dit op u gehoor gehad het:

____

____

____

____

Enige ander opmerkings of inligting oor u gehoor:

____

____

____

____

___

Baie dankie vir u bereidwilligheid om aan hierdie studie deel te neem. Wees verseker dat dit opreg waardeer word!

7.2 Appendix B: Letter of informed consent

100
1908 - 2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Fakulteit Geesteswetenskappe
Departement Kommunikasiepatologie

1 Oktober 2009

Geagte Student,

Navorsingsprojek Kommunikasiepatologie

Ek is tans besig met my meestersgraad in Kommunikasiepatologie. Om aan die vereistes van die graad te voldoen word daar van my verwag om `n navorsingsprojek uit te voer. My navorsingsprojek handel oor die gebruik van oto-akoestiese emissies (OAE) by volwassenes tussen die ouderdom van 18 tot 25 jaar. Oto-akoestiese emissies is klanke wat deur die binne-oor (koglea) geproduseer word. Dit word gemeet deur `n meetproppie in die oorkanaal te plaas, en geen reaksie word van die deelnemer verwag nie, aangesien dit `n outomatiese, objektiewe elektroniese toets is.

Die doel van hierdie navorsingsprojek is om die toets-hertoets geldigheid en betroubaarheid, asook herhaalbaarheid van die resultate te ondersoek met kort en lang tydsverloop tussen toetse. OAE-toetse neem slegs 15 minute, maar om te bepaal of die kandidate geskik is vir die doel van die navorsingsprojek, moet ander oudiologiese toetse uitgevoer word, wat ongeveer `n uur per deelnemer sal neem. Die insameling van hierdie data sal by die Departement Kommunikasiepatologie op die hoofkampus van UP uitgevoer word.
Om die relevante data vir hierdie studie in te samel sal daar met jou inisiële besoek `n volledige toetsbattery uitgevoer word. Daar sal egter van jou verwag word om die departement op voorafbepaalde datums een week na die inisiële toetse, en een maand na die inisiële toetse te besoek om bykomende data in te samel. Opvolgbesoeke sal ongeveer vyftien minute duur.

As deelnemer aan hierdie studie sal u nie aan enige risiko's blootgestel word nie. Persoonlike inligting sal slegs vir navorsingsdoeleindes gebruik word, en konfidensialiteit sal deur kodering verseker word. Neem asb. daarvan kennis dat die resultate van die toetsprosedures vir `n tydperk van 15 jaar gestoor sal word, vir argivering en moontlike toekomstige studies, ooreenkomstig die eise van navorsing in hierdie veld. Deelname aan hierdie studie is vrywillig, en u kan enige tyd van die studie of toetsprosedures onttrek.

Indien u enige verdere inligting benodig kan u my kontak by 082 460 8570.

By voorbaat dank!

Carina Grové
M. Kommunikasiepatologiestudent

Dr. M. Soer
Studieleier

(wnd)
HOOF VAN DEPARTEMENT

Universiteit van Pretoria
Pretoria, 0002
Suid Afrika

Telefoon : 00 27 12 420-2357
Faksimile : 00 27 12 420-3517

brenda.louw@up.ac.za
www.up.ac.za

100
1908 - 2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Department of Communication Pathology

1 October 2009

Dear Student,

Research project Communication Pathology

I am currently busy with my Master's degree in Communication Pathology. To complete this degree I am required to complete a research project (thesis). My research project is about the use of oto-acoustic emissions (OAE) in adults between the ages of 18 to 25 years. Oto-acoustic emissions are sounds produced by the inner ear (cochlea). It evaluates the functioning of the cochlea, and no reaction is expected of the participant, because it is an automatic, objective, electronic test.

The aim of this research is to determine the test-retest reliability, validity and reproducibility of the results over a short and longer period of time. OAE-tests take only 15 minutes, but to determine whether the candidates are appropriate for the aim of this research project, other audiological tests need to be performed that would take more or less an hour for each participant. The gathering of this data will be done at the Department of Communication Pathology on the main campus of the University of Pretoria.
In order to capture the relevant data for this study the complete test procedures will be carried out on your initial visit. You will, however be required to visit the department again at intervals of one week after your initial testing, and one month after your initial testing, where short follow-up tests (about 15minutes) will be conducted to compare the original results with.

As a participant in this study, you will not be exposed to any risks. Your personal information will be used for research purposes only and confidentiality will be ensured through coding. Please take note that the results of the study will be stored for 15 years for archiving and reference for possible future studies, in accordance with the requirements for research in this field. Participation in this study is purely voluntary and if you wish to withdraw from the study or testing procedures, you may do so at any time.
If you need any other information you can contact me on 082 460 8570.

Thank you.

Carina Grové

M. Communication Pathology student

Dr. M. Soer

Research supervisor

(wnd)

HEAD OF DEPARTMENT

University of Pretoria
Pretoria, 0002
South Africa

Telephone : 00 27 12 420-2357
Facsimile : 00 27 12 420-3517

brenda.louw@up.ac.za
www.up.ac.za

7.3 Appendix C: Letter confirming ethical clearance obtained for study

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Postgraduate Committee

22 January 2010

Dear Dr Soer,

Project:	**Test-retest reliability and validity of distortion product oto-acoustic emissions and transient evoked oto-acoustic emissions in normal hearing adults**
Researcher:	**CL Grové**
Supervisor:	**Dr M Soer**
Department:	**Communication Pathology**
Reference number:	**25080271**

Thank you for the application you submitted to the Postgraduate Committee of the Faculty of Humanities.

It is my pleasure to inform you that the Postgraduate Committee formally **approved (with suggestions)** the above study on 19 January 2010. The application need not be resubmitted to this committee.

Please note that this approval is based on the assumption that the research will be carried out along the lines laid out in the proposal. Should the candidate's actual research depart significantly from the proposed research (as sometimes happens for a variety of possible reasons), it would be necessary to apply for a new research approval and ethical clearance.

The application will be submitted to the Research Ethics Committee for consideration of the ethical implications of the research. Please note that data collection may not commence prior to approval by the Research Ethics Committee.

The Committee requests that you convey this approval to Ms Grové.

We wish you success with the project.

Sincerely

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