Chapter 6: Discussion and Interpretation of Results

6.1 Introduction

The development of objective procedures in audiology came a long way since the 1920s. With the aid of modern technology, audiologists can measure the exact degree, configuration, and site of hearing loss in adults and confirm these findings with a series of objective electrophysiologic and physiologic procedures, such as tympanometry, the acoustic reflex, ABR, and otoacoustic emissions (Northern, 1991). From the overview of the development of objective procedures in audiology in Chapter 2, however, it is evident that there are some weaknesses in current objective diagnostic procedures when it comes to the evaluation of special populations. In the evaluation of neonates from birth to 6 months, the crucially ill, and malingerers, audiologists often have to rely heavily on objective electrophysiologic procedures to determine hearing ability. To determine hearing thresholds with electrophysiologic procedures is often costly, requires a large amount of time as well as highly trained and specialized personnel, and may require sedation. Above all, current objective physiologic procedures, such as ABR, have a limited frequency area in which hearing ability can be determined accurately. There is therefore a definite need for an objective, reliable, rapid, and economic test of hearing that evaluates hearing ability across a range of frequencies to aid in the assessment of difficult-to-test populations.

The distortion product otoacoustic emission has been intensely investigated as a possible new test of hearing (Probst & Hauser, 1990; Gorga et al. 1993; Moulin et al. 1994; Kimberley et al. 1994a; Kimberley et al. 1994b; Stover et al. 1996a). Many studies found a strong correlation between DPOAEs and PTTs despite the fact that results obtained from conventional pure tone audiometry involves an evaluation of the entire auditory system opposed to DPOAE results that involve only an evaluation of cochlear functioning (Gaskill & Brown, 1990; Lee et al. 1993; Vinck et al. 1996; Kummer et al. 1998). Some studies attempted to predict normal hearing with DPOAEs and conventional statistical methods (Kimberley et al. 1994b; Moulin et al. 1994) and artificial neural networks (Kimberley et al. 1994a; De Waal, 1998). Researchers however, experienced difficulty to predict hearing status at low frequencies due to problems with the rising noise floor and the inability of conventional statistics to deal with nonlinear and noisy data sets (Durrant, 1992; Gorga et al. 1996; Gorga et al. 1993; Kimberley et al. 1994b; Stover et al. 1996). This could possibly be attributed to the fact that most studies attempted to correlate normal functioning of a single area on the basilar membrane, such as the f2 frequency place (Durrant, 1992; Harris et al. 1989; Kimberley et al. 1994b), the GM frequency place (Martin et al. 1990b; Lonsbury-Martin & Martin, 1990; Bonfils et al. 1991) or the 2f1-f2 place (Smurzynski et al. 1990) with the corresponding pure tone frequency and used the threshold or amplitude of a single DPOAE to correlate with a pure tone threshold.

The current study investigated the correlation between DPOAEs and PTTs with artificial neural networks, a data processing technique proven superior to conventional statistics when it comes to dealing with noisy non-linear data sets (Kimberley et al. 1994a; Raghupathi et al. 1993; Hann & Steurer, 1996). This study also used the presence or absence and amplitudes of all 11 DPOAE responses to predict hearing ability at a single PTT frequency. The fact that pure tone thresholds are interrelated made it possible to predict hearing status at low frequencies objectively and accurately with this method, for the first time (De Waal, 1998). The present study once again showed that the distortion product otoacoustic emission has enough data at surrounding frequencies to make objective predictions at problematic frequencies such as 500 Hz with a high level of accuracy.

The aim of this chapter is to discuss all findings and interpret the significance thereof. Firstly, the correlation found between DPOAEs and PTTs will be discussed, secondly, the results obtained for the prediction accuracy at the four frequencies 500, 1000, 2000 and 4000 Hz. Thirdly, the significance of results in terms of readiness for broad clinical use will be discussed, with emphasis on the requirements for sensitivity and specificity of screening and diagnostic procedures. Then, to integrate findings of the four frequencies, a few case studies will be discussed where the network made accurate predictions for all frequencies, and also case studies predicted inaccurately with possible reasons for inaccurate predictions. Lastly, all the DPOAE and ANN variables that were experimented with to determine their effect on prediction accuracy will be discussed and interpreted.

6.2 Indication of a Correlation between DPOAE Measurements and Pure Tone Thresholds

Many other studies used statistical techniques to determine the correlation between DPOAEs and pure tone thresholds or described case studies that demonstrated a close relationship between DPOAEs and pure tone thresholds (Gaskill & Brown, 1990; Gorga et al. 1993; Kummer et al. 1998; Lee et al. 1993; Vinck et al. 1996). In the case of statistical methods, a correlation coefficient can be determined and that serves as an indication of the correlation found and its significance. In this study, however, artificial neural networks were used to predict pure tone thresholds. The network extracts necessary information from input stimuli and then forms an internal representation of relations between different data sets by adjustment of the weights of the middle neurons. The neural network then uses the learned representations to make predictions. One of the limitations of a neural network is that it cannot justify the learned relationships and specify exact correlations in terms of strength or significance. By analyzing the accuracy of the predictions, one can make assumptions about the correlation between DPOAEs and pure tone thresholds, but one cannot dissect a neural network to find precise reasons and relationships for accurate predictions. With this aspect in mind, the implied correlation between DPOAEs and pure tone thresholds will be discussed briefly.

If there were no correlation between DPOAEs and pure tone thresholds (PTTs), then the neural network would not have been able to make accurate predictions of hearing ability with more than 50% accuracy. Correct predictions of hearing ability would have been mere chance or at random. If a histogram was drawn to illustrate the prediction accuracy of a data set that is not correlated with another data set at all, one can expect to see an equal number of predictions in every one of the domain values. This would result in a "flat" histogram or a histogram representing random predictions at the various domain values. It would definitely not result in a histogram depicting a normal curve distribution.

The prediction accuracy of this neural network is illustrated in the histograms in Figure 6.1.

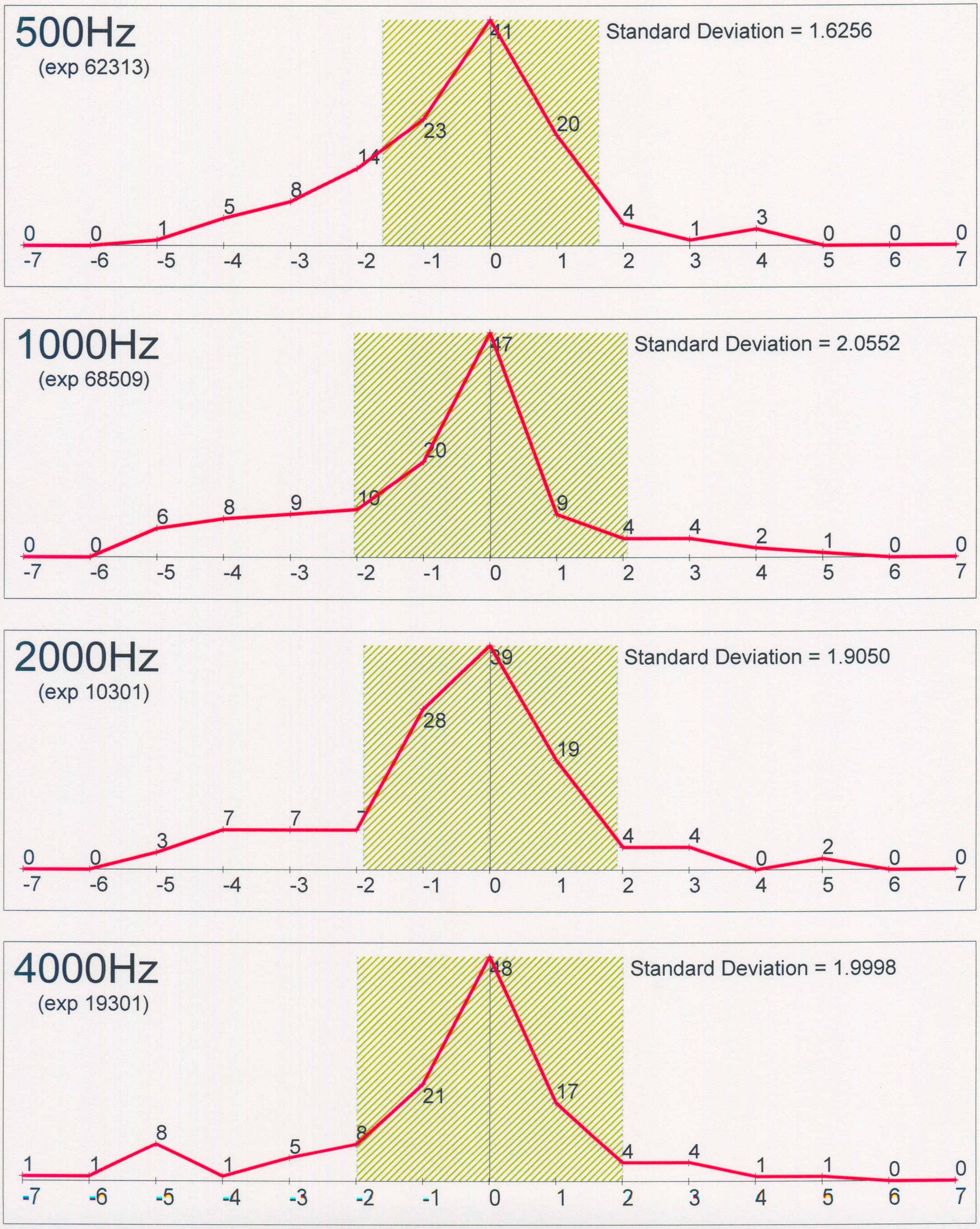

Figure 6.1: Prediction accuracy of PTTs with DPOAEs and ANNs

At first glance the presence of a normal curve distribution can be seen in all these histograms in Figure 6.1. Most ears were predicted accurately within the same class (these ears are indicated at the zero (0) place on the histogram) or within one category of hearing impairment. This is clearly an indication that the neural network found a correlation between DPOAEs and PTTs and used that correlation to make the predictions.

This study therefore confirms the results of many other researchers that the distortion product is strongly correlated with pure tone thresholds in normal hearing and hearing-impaired ears (Gaskill & Brown, 1990; Gorga et al. 1993; Kimberley et al. 1994; Kummer et al. 1998; Lee et al. 1993; Moulin et al. 1994; Vinck et al. 1996).

6.3 Prediction of 500 Hz

The prediction of 500 Hz with distortion product otoacoustic emissions has been problematic for many authors (Gorga et al. 1993; Moulin et al. 1994; Probst & Hauser, 1990; Stover et al. 1996a). Regardless of the loudness level of the primaries, the chosen frequency ratios, loudness level ratios or any other variables that could influence the study, the rising noise floor below 1000 Hz limited the measurement of clear responses at f2 = 500 Hz (Durrant, 1992). Probst and Hauser (1990) attempted to predict hearing ability as normal or impaired in the geometric mean frequency range of 500 Hz to 8000 Hz. Their findings indicated that the majority of normal and near-normal ears had no or small DPOAE amplitudes at 500 Hz and 8000 Hz. No correlations with hearing threshold could be established at these two frequencies. In a study by Stover et al. (1996a), the noise floor for lower frequencies (500 Hz and 707 Hz) was so high, that data at these two frequencies were interpreted as unknown or absent and coded as missing for data analysis. At 500 Hz, the prediction of normal hearing was no better than chance. Gorga et al. (1993) and Moulin et al. (1994) experienced similar problems with very high noise measurements at low frequencies and could also not predict normal hearing at 500 Hz.

The prediction accuracy for hearing ability at 500 Hz for this study yielded promising and interesting results. Normal hearing (0-25dB HL) at 500 Hz was predicted as normal 94% of the time. The improvement in prediction ability of low frequencies in this study can be attributed to two reasons. The first reason includes the different data processing procedure, an artificial neural network with excellent correlation finding and prediction capabilities. The second reason is that pure tone threshold at 500 Hz was not predicted with DPOAE amplitude or threshold at low frequencies, but with a pattern of all present and absent DPOAE responses with their amplitudes across all 11 DPOAE frequencies and all eight DP Grams. Previous studies attempted to correlate the pure tone threshold with either the threshold (or amplitude) of the f2 frequency (Harris et al. 1989; Kimberley et al. 1994), the geometric mean frequency (Bonfils et al. 1991; Lonsbury-Martin & Martin, 1990) or the distortion product frequency (Smurzynski et al. 1990). This study did not use a single point DPOAE measurement to predict a single point pure tone threshold, but used the whole spectrum of emissions to predict a single pure tone threshold. The artificial neural network was able to gain enough information from the whole spectrum of absent and present responses to predict normal hearing ability at 500 Hz correctly 94% of the time.

Another aspect of the prediction of hearing ability at 500 Hz that should be investigated, is the number of false positive and false negative predictions the neural network made. Category one (0 + 5dB) had 0% false positive responses, in other words, subjects with normal hearing were never predicted as having a hearing loss. Category two (10 + 15dB) had 2% false positive responses. Category three (20 + 25dB) had 1% false positive responses. If the three categories are combined to represent normal hearing (0-25dB), the false positive rate is only 3%. This procedure was therefore very specific and almost never predicted a subject with normal hearing as hearing impaired at 500 Hz.

The sensitivity of the procedure (the percentage of hearing impaired ears predicted as normal) is influenced by the number of false negative responses. The false negative rates at 500 Hz for category four (30 + 35dB) was 6%, category five (40 + 45dB) was 5%, category six (50 + 55dB) was 4% and category seven (60 + 65dB) was 0%. The percentage of ears predicted as normal decreased as the category's hearing loss increased. This finding supports results from the study of Probst and Hauser (1990) in which they found that mild hearing losses are often predicted as normal.

The total false negative rate for the combined categories spanning hearing loss was 15%. This value is much lower than reported elsewhere: no study predicted normal hearing at a frequency as low as 500 Hz and reported that false negatives were so high at low frequencies that performance was no better than chance (Gorga et al. 1993; Moulin et al. 1994; Probst & Hauser, 1990; Stover et al. 1996a). Unfortunately, this value might still be considered too high and raises questions regarding the sensitivity of this procedure for screening purposes. It could also be argued that category four (30 + 35dB) represents a mild hearing loss, and that the sensitivity of this procedure to identify moderate and severe hearing losses could be determined by the false negative rates for categories five to seven, which would bring the number down to 9%. Even though this procedure does not perform perfect classification of normal and impaired hearing, the results obtained in this study at 500 Hz clearly shows that DPOAEs and ANNs can predict hearing status in populations with varying degrees of hearing loss objectively with high degrees of accuracy, sensitivity and specificity. (The false positive and false negative rates for all four frequencies in terms of suitability for use in screening and diagnostic audiology tests are further discussed in 6.8 "Results of the present study in perspective: Readiness for broad clinical use".)

Bonfils et al. (1991) investigated objective low-frequency audiometry by distortion product otoacoustic emissions and found that active emissions could be measured as low as 2f1-f2 = 512 Hz. The authors concluded that DPOAEs could be used as an objective low-frequency test of auditory functioning. This study confirms the results of Bonfils et al. (1991). **DPOAEs and ANNs can accurately categorize hearing ability at 500 Hz as normal, 94% of the time.**

It was also attempted to predict impaired hearing at 500 Hz into 10 dB categories. The prediction of impaired hearing at 500 Hz, but also at all the other frequencies (1000 Hz, 2000 Hz and 4000 Hz) was very unsatisfactory and prediction percentages for categories spanning hearing loss indicated poor prediction capabilities of the neural network. Many possible reasons come to mind for this incapability of hearing impaired prediction. The reasons apply to all the frequencies and these reasons will be discussed in the next section.

6.4 Possible Reasons for Poor Prediction of Categories Representing Hearing Loss.

The first reason that can be suggested for the poor prediction of hearing impaired categories is that hearing impaired subjects might demonstrate less clear DPOAE responses that might influence the correlation between DPOAEs and PTTs. This possibility was seen in the DPOAE evaluation where it took the DPOAE equipment longer to test a hearing impaired subject in order to get enough frames of data to regard a test as "accepted" (the criterion for a test to be accepted for the GSI-60 DPOAE system for a screening test is that the DPOAE amplitude had to be 10dB above the noise floor or the cumulative noise level had to be –18dB SPL. The maximum number of frames tested in a screening procedure is 400, and if no clear response is measured in that time, the test is scored "timed out" which means that no response was obtained. This was described in "4.6.5.2 Determination of optimal stimulus parameters"). It is possible that more responses could be obtained from hearing-impaired subjects if the criterion for test acceptance is lowered to 5 dB for example. The lowered criterion for the acceptance of a test as "accepted" could possibly enhance the number of useable responses from hearing-impaired subjects and might therefore enhance prediction accuracy of categories spanning impaired hearing. This aspect should be further investigated.

Another reason for poor prediction accuracy of hearing impaired categories might be that the optimal procedure for data analysis has not yet been identified. I can be hypothesized that another type of neural network with different topologies, learning rules and error tolerances would be able to make more accurate predictions. Or could a complete new form of data processing, such as "genetic programming", inspired by

Darwinian invention and problem solving that "progressively breeds a population of computer programs over a series of generations" (Koza, Bennett, Andre & Keane, 1999:3) to find an optimal solution to a problem, be able to make more accurate predictions? These possibilities have yet to be investigated.

One definite aspect however, that seems to have influenced the prediction accuracy of hearing ability in categories spanning hearing impairment more than neural network capabilities and the underlying correlation between the two data sets, was the number of ears in every category that the neural network had to train on.

Neural networks need enough examples in every category to form representations of how a specific ear's DPOAE type relate to its PTT type in order to make accurate predictions. Even though it was attempted to categorize all audiograms in this study into three groups to ensure that hearing impairment was as well represented as normal hearing, the nature of sensorineural hearing impairment tends to affect certain frequencies more than others, and low frequencies are often normal. In the case of 500 Hz, this leads to the uneven distribution of many ears in categories representing normal hearing and few ears in hearing loss categories. At 4000 Hz however, category six (50 + 55dB) had more ears (a total of 20) and was predicted accurately 45% of the time and within 10dB 30% of the time. Hearing loss at 4000Hz in this category was wrongly predicted only 25% of the time with a false negative rate of only 2%. The same category at 500 Hz had only 7 ears to train on and prediction accuracy was never correct and within 10dB only 14% of the time. If the network had more ears in every category to train on, prediction accuracy might have been considerably better.

To illustrate this concept, figure 6.2 plots the number of ears in every category against prediction accuracy. It clearly demonstrates that the number of ears in every category had an enormous effect on prediction accuracy.

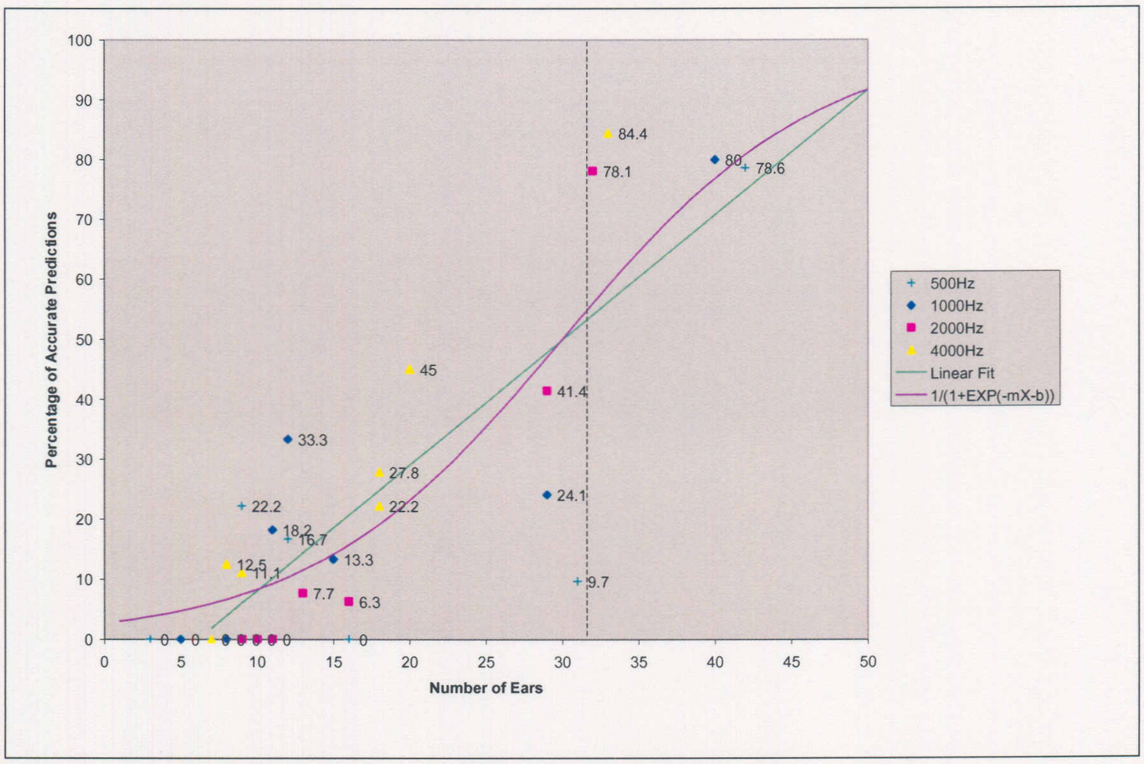

Figure 6.2: Prediction accuracy and ear count correlation.

The aim of this study is not to describe the relationship between the number of ears per category and prediction accuracy thereof. However, Figure 6.2 shows a number of possible relationships merely to illustrate the notion of "more-is-better". The linear fit (in green) shows that higher number of ears lies significantly above expectation. It is worthy to note that the relationship cannot be linear, since any line with a slope larger than zero will have to cross the 100% accuracy limit at some point. This is of course not possible.

The figure indicates an alternative fit (in purple) of the form $1 / (1 + e^{-mx-b})$, that seems better and more intuitive since it starts out low, just like the experimental data, but also asymptotically approach 100% for large numbers of ears. It is also not established if this function has any correlation with the sigmoid function (Blum, 1992:39) $(1 / (1 + e^{-x})$ that was used to normalize output of the separate ANN layers but it seems to be a more likely option than a linear fit.

For this data set there is a fairly clear threshold at 32 ears per category where prediction accuracy suddenly surges into the 75% and higher region.

Should any of the example relationships hold, it is expected that a 95% or higher accurate predictions could potentially be made if an ANN receives 80 or so ears per category.

The shortage of data in certain categories is probably the main reason for this study's poor prediction capabilities for ears demonstrating hearing impairment and not poor correlation between DPOAEs and PTTs of hearing-impaired ears or incapability of the neural network to deal with this data set.

6.5 Prediction of 1000 Hz

Researchers attempting to predict normal hearing at 1000 Hz with DPOAEs performed better than at 500 Hz, but still reported very high false negative rates and influence of low frequency noise interfering with test measurements (Gorga et al. 1993; Kimberley et al. 1994b; Moulin et al. 1994; Probst & Hauser, 1990). In a study by Kimberley et al. (1994b), hearing ability at 1025 Hz could be accurately predicted as normal 90% of the time. Kimberley et al. (1994b) did not state the false negative rate for 1025 Hz specifically, but an average false negative rate (where hearing-impaired ears were predicted as normal) across the frequency range f2 = 1025 Hz to 5712 Hz of 22%. Moulin et al. (1994) accurately predicted normal hearing at 1000 Hz 73% of the time. False negative responses in this study varied between 12% and 17% with an average of 15%. In the study by Gorga et al. (1993), false negative rates ranged from about 25% to over 60% depending on the hit rate that was selected.

For this study, the accurate prediction of normal hearing (0-25dB) at 1000 Hz was 88%. The false positive rate for the three categories combined is 6%. False negative responses for 1000 Hz were very high, the highest for all four frequencies tested: 2% for C4 (category four)(30 + 35dB), 10% for C5 (category 5)(40 + 45dB), 5% for C6 (50 + 55dB) and 5% for C7 (60 + 65dB). That comes to a total of 22%, which is unacceptably high to regard this procedure as sensitive enough to correctly identify hearing impairment.

Even though the predictions of normal hearing in this study for 1000 Hz were more accurate than stated elsewhere (Gorga et al. 1993; Moulin et al. 1994; Probst & Hauser, 1990), the high incidence of false negative responses influences the sensitivity of this procedure. This high incidence of false negative responses lessens the clinical applicability of this neural network run as a possible screening procedure. (The false positive and false negative rates for all four frequencies in terms of suitability for use in screening and diagnostic audiology tests are further discussed in 6.8 "Results of the present study in perspective: Readiness for broad clinical use".)

The prediction of specific categories spanning hearing impairment at 1000 Hz was again disappointing. The possible reasons were discussed in 6.4. It seems that the main reason for the network's incapability to predict hearing loss at 1000 Hz also stems from insufficient data in every category. The shortage of ears to train on might even have affected the prediction accuracy of normal hearing: C3 (category three spanning 20 + 25dB) only had 12 ears and were predicted accurately only 29% of the time versus C1(0 + 5dB) that had 40 ears and was predicted accurately 80% of the time. From Figure 6.2 it could be suggested that a study with an expansive data set of more than 80 ears in every category might be able to make predictions of hearing ability at 1000 Hz, normal or impaired, with very high levels of accuracy.

6.6 Prediction of 2000 Hz

The prediction of normal hearing at 2000 Hz in other studies yielded far more promising results than their predictions at 500 Hz and 1000 Hz. Kimberley et al. (1994b) predicted normal hearing at 2050 Hz with 84% accuracy. Mean false negative responses for f2 = 1025 Hz to 5712 Hz was 22%. Moulin et al. (1994) predicted the DPOAE frequency of 1413 Hz (closest to the GM frequency of 2000 Hz) correctly 75.6% of the time with an average false negative response of 15%.

In this study, normal hearing (0-25dB HL) at 2000 Hz could be predicted accurately 88% of the time. The false positive rate for the first three categories combined was 5%, in other words, 5% of normal ears were predicted as hearing impaired at 2000 Hz which makes this procedure quite specific in this frequency region.

False negative rates were 4% for C4 (30 + 35dB), 4% for C5, 4% for C6 and 3% for C7. This comes to a total of 15% for all categories spanning hearing loss. This number is the same as the number found by Moulin et al. (1994) and lower than in the study by Kimberley et al. (1994b) but 15% is still considered rather high and raises concern for the sensitivity of this procedure as a screening procedure to identify hearing loss at 2000 Hz. (See 6.8 "Results of the present study in perspective: Readiness for broad clinical use".)

The influence of ear count in every category on prediction accuracy of hearing impaired categories is depicted in Figure 6.2. Categories that had less than 12 ears were never predicted accurately. C3 (20 + 25dB), which forms part of the representation of normal hearing had only ten ears and was never predicted accurately. It could be speculated that normal and impaired hearing at 2000 Hz would have been predicted more accurately if there were more data in every category.

6.7 Prediction of 4000 Hz

The prediction of normal hearing at 4000 Hz has been a strong point for some of the previous studies. Moulin et al. (1994) successfully predicted the DPOAE frequency of 4000 Hz (primary frequencies are between 5000 Hz and 6000 Hz) as normal 79.4% of the time. These authors predicted the DPOAE frequency of 2826 Hz (primary frequencies between 3500 Hz and 4500 Hz) as normal 82% of the time, with an average false negative rate of 15%. Kimberley et al. (1994b) predicted normal hearing at 4052 Hz accurately 88% of the time with mean false negative rates of 22%.

The prediction of normal hearing at 4000 Hz with the use of DPOAEs and ANNs revealed that normal hearing could be predicted accurately 93% of the time. The false positive rate was only 2%, which makes this a very specific procedure to use for identification of normal hearing at 4000 Hz. Very good hearing (0-15dB) was correctly predicted an normal (0-25dB) 96% of the time. This procedure therefore offers a very rapid and accurate objective measurement of normal hearing.

The false negative rates at 4000 Hz were 3% for C4, 2% for C5, 2% for C6, 5% for C7 and 1% for C8 (70 + 75dB). The total false negative rate for all categories spanning hearing impairment was 13%. This is the lowest incidence of false negative responses for the four frequencies in this study, and also stated elsewhere for 4000 Hz (Moulin, et al. 1994; Kimberley et al. 1994b). This study proved it possible to predict normal and impaired hearing at various frequencies with DPOAEs and ANNs but more research is needed to lower false negative rates to acceptable levels for screening and diagnostic purposes. (See 6.8 "Results of the present study in perspective: Readiness for broad clinical use".)

Prediction of specific classes of hearing impairment at 4000 Hz was once more not satisfactory due to insufficient data in certain categories for the neural network to train on. The ear count versus prediction accuracy correlation for 4000 Hz can be seen in Figure 6.2. C3 (20 + 25 dB) had only seven ears and was never predicted accurately opposed to C1 (0 + 5dB) that had 33 ears and was predicted accurately 85% of the time. It could be speculated that normal and impaired hearing could be predicted even more accurately if there were more ears in every category.

6.8 Results of the Current Study in Perspective: Readiness for Broad Clinical Use.

The need for an objective, rapid, non-invasive, inexpensive, and accurate measurement of hearing inspired the investigation of the distortion product otoacoustic emission (DPOAE) as a possible new test of hearing.

Otoacoustic emissions have been used for a number of clinical uses in audiology. Applications include (a) screening of neonates and infants; (b) differential diagnosis of cochlear versus retrocochlear hearing losses; (c) monitoring of the effects of noise exposure or ototoxic drugs on the outer hair cells; and (d) to monitor fluctuating hearing loss in persons with Meniere's disease (Lonsbury-Martin et al. 1992; Norton & Stover, 1994; Norton & Widen, 1990; Robinette, 1992; Koivunen, et al. 2000).

The distortion product has been proven as an acceptable screening procedure. It is present in all normal hearing ears and even though it is measurable in ears with a hearing loss of up to 65dB HL, the amplitude and threshold of the DPOAE indicate different qualities, revealing hearing impairment (Moulin et al. 1994; Smurzynski et al. 1990). It can be measured non-invasively, objectively and rapidly (Norton & Stover, 1994). It is not significantly affected by gender (Cacace et al. 1996; Lonsbury-Martin et al. 1990) and has good test-retest stability (Cacace et al. 1996). Furthermore, it can be measured over a wide range of frequencies (Bonfils et al. 1991). DPOAEs are not affected by state of consciousness and do not require sedation (Norton & Stover, 1994). Lastly, it is an economic test that yields ear specific information. Many researchers used these attributes to correctly identify normal hearing in populations with varying degrees of hearing ability (Gorga et al. 1993;

Kimberley et al. 1994; Moulin et al. 1994). The only limitation of DPOAEs as a screening procedure is the lack of sensitivity identified in some of the studies. Sensitivity of a screening procedure is affected negatively by a high incidence of false negative responses. False negative responses refer to hearing-impaired ears that are predicted as normal (Schwartz & Schwartz, 1991). Some studies, including the present one, revealed an incidence of false negative responses too high for clinical acceptability (Kimberley et al. 1994; Moulin et al. 1994). According to Brass and Kemp (1994), it is very important to have a very high sensitivity for a first pass screening procedure (low incidence of false negative responses), as close as 100% sensitive as possible. The specificity of a screening procedure (affected by the number of false positive responses), on the other hand, is less important and is quite acceptable even if the test is only moderately specific (such as a specificity of 75%). Brass and Kemp (1994) compared test effectiveness and efficiency quite effectively: "In terms of screening effectiveness, the final number of false negatives is very important as this is the group of those for whom we were screening but missed. In terms of screening efficiency, the final number of false positives is important as this increase the number passed on to and hence the cost of the next stage of screening." (Brass & Kemp, 1994: 386).

It would therefore seem that the number of false negatives in this study is too high for an acceptable screening procedure, even though it is better than reported elsewhere (Gorga et al. 1993; Moulin et al. 1994; Probst & Hauser, 1990; Stover et al. 1996a).

DPOAEs however, have never been used as a diagnostic test of hearing where specific thresholds for frequencies were determined (Kimberley et al. 1994b; Lee et al. 1993). Many researchers mentioned the increasing role that otoacoustic emissions have in diagnostic audiology (Kemp et al. 1990; Martin et al. 1990; Moulin, 2000a, Moulin 2000b) and described the possibility of DPOAEs as a diagnostic audiologic test (Durrant, 1992; Kimberley et al. 1994b; Lee et al. 1993). These authors also stated that additional research is necessary before otoacoustic emissions can be implemented as a diagnostic test of hearing.

This research project attempted to predict hearing ability at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz with DPOAEs in normal and hearing impaired ears with the use of artificial neural networks. It was attempted to predict specific hearing levels within 10dB for normal and hearing-impaired ears. Even though this study could correctly identify normal hearing quite accurately, the specific predictions of hearing levels at various frequencies were rather disappointing. One possible reason for the poor prediction of categories depicting hearing loss is the number of ears in every category that the neural network had to train on. Chapter 3 explained the learning and training of a neural network in 3.5.2.2, and that every category should have enough data for the neural network to train on, to form adequate representations and to make accurate predictions. To investigate this possibility, the accuracy of the prediction was correlated with the amount of data that the neural network had to train on. (See Figure 6.2.) Results indicated that there seems to be a threshold value for the number of ears in every category: when there were more than 32 ears in a category, prediction accuracy was always above 75%. These results suggests that categories depicting hearing impairment can be predicted accurately, if there is enough data in every category for the neural network to train on. If all the parameters of this research project were kept exactly the same with just one alteration of research design namely, the increase of subjects, hearing ability could be accurately predicted within 10dB from 500 Hz to 4000 Hz, for hearing levels up to 65dB HL.

This find is definitely a great contribution to the development of DPOAEs as a diagnostic test of hearing. Results proved that it is possible to predict hearing ability at frequencies as low as 500 Hz objectively, rapidly, non-invasively, and economically with DPOAEs and ANNs. The research in this study could be used as a foundation stone for further research to develop the distortion product as a new objective diagnostic test of hearing. Such a test would change the field of objective diagnostic audiology as we know it considerably and would be a tremendous aid in the evaluation of difficult-to-test populations.

To integrate results obtained from the prediction of 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz, a few case studies will be investigated where the neural network accurately predicted hearing ability in all four frequencies. A few case studies will also be investigated where the neural network made false predictions with possible reasons for inaccurate predictions. These results are discussed below.

6.9 Case Studies where the Audiogram was Predicted Accurately

The network predicted hearing ability into one of seven 10dB categories. Examples for correct predictions can be seen in Figure 6.3. Subjects that demonstrated very good hearing ability (0-15dB) at all four frequencies were usually predicted accurately. Figure 6.3. depicts the audiogram and predicted categories for ear 35, an ear with normal hearing that was predicted accurately, ear 85, an ear with a hearing loss predicted accurately except for 4000 Hz that was 10dB out and ear 107, an ear of an 80-year-old person, predicted accurately except for 500 Hz that was 10dB out. Subject information, such as age, gender, and complaints of tinnitus or vertigo is presented in Table 6.1.

Table 6.1: Subject information of cases predicted accurately

	Ear 35	**Ear 85**	**Ear 107**
Subject age	61	72	80
Subject gender	Female	Male	Female
Tinnitus?	No	No	High frequency
Vertigo?	No	No	No
Medication	None	None	None
Noise exposure?	None	None	None

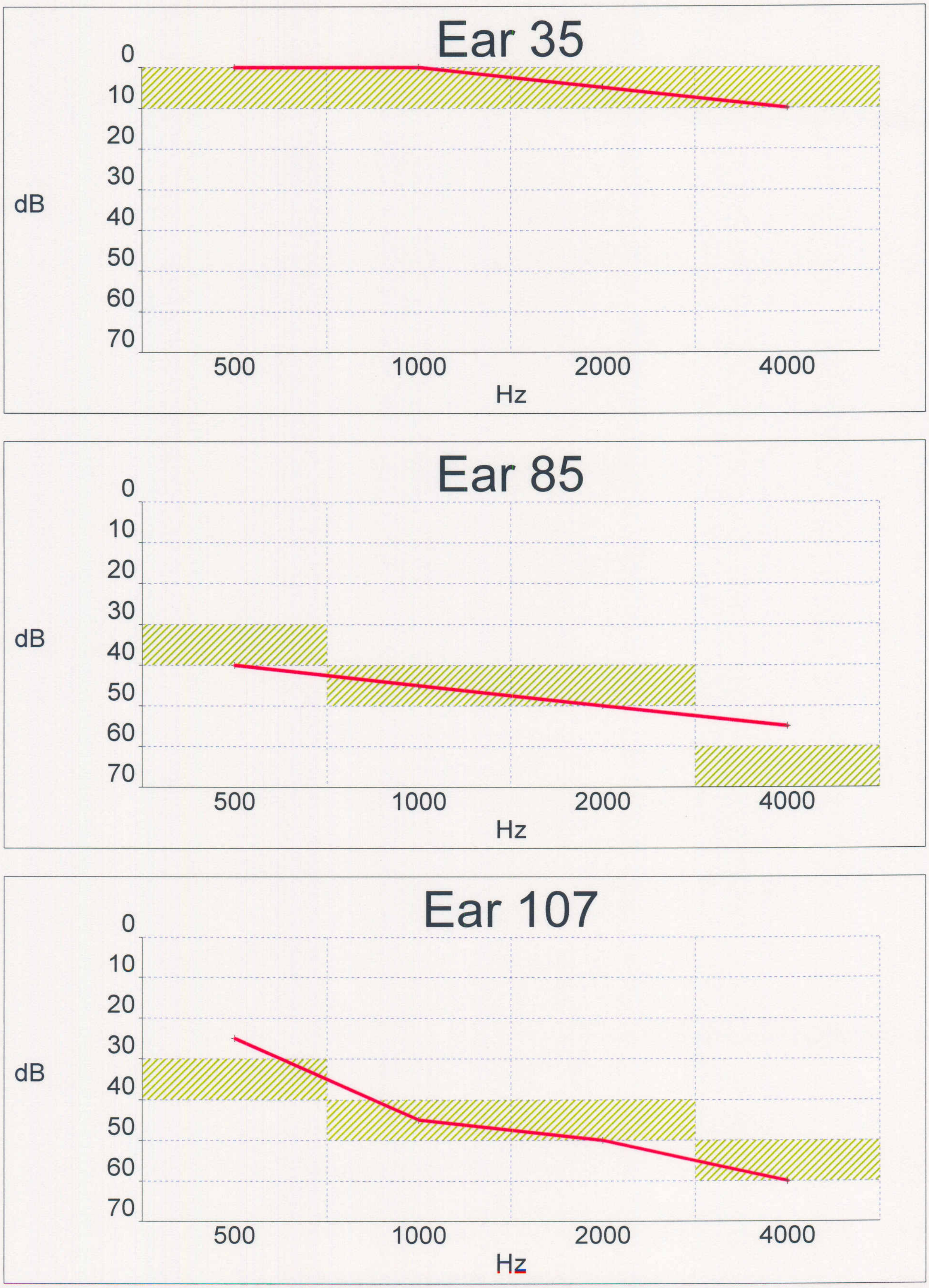

Figure 6.3: Case studies where the audiogram was predicted accurately (prediction in green, audiogram in red).

6.10 Case Studies where the Audiogram was Predicted Inaccurately

Subjects demonstrating hearing loss were sometimes predicted inaccurately. Six examples of inaccurate predictions will be given. Audiograms and predictions can be seen in Figure 6.4. The subject information of these six ears is presented in Table 6.2.

Table 6.2: Subject information of cases predicted inaccurately

	Ear 19	**Ear 33**	**Ear 71**	**Ear 73**	**Ear 74**	**Ear 91**
Subject age	31	35	16	39	39	43
Gender	Female	Female	Male	Male	Male	Male
Tinnitus	No	High Frequency	No	No	No	No
Vertigo	No	No	No	No	No	No
Medication	None	None	None	None	None	None
Noise exposure	None	None	None	20 years	20 years	15 years

6.10.1 Interesting Phenomena in Cases Predicted Inaccurately

With closer analysis of the cases that were predicted inaccurately, it became evident that there were a few circumstances in which the neural network could almost never predict hearing ability correctly. Some of these instances included noise exposure, very mild hearing losses, and possible retrocochlear hearing losses. These cases will be discussed below. It is however, also the case that in some instances, the neural network predicted hearing ability inaccurately for no apparent reason.

Figure 6.4: Case studies where the audiogram was predicted inaccurately (prediction in green, audiogram in red).

6.10.1.1 Subjects Demonstrating Hearing Loss Due to Noise Exposure

Subjects with a large amount of noise exposure revealed poor correspondence between pure tone audiograms and DPOAE measurements. The DPOAE measurements indicated a much larger hearing loss than the pure tone audiograms when DPOAE measurements were compared to the normal range. This confirms the research by Durrant (1992) that indicated that damage in outer hair cells could be measured before the actual hearing loss occurs. Even though these subjects already had noise-induced hearing losses, damage to these subject's outer hair cells indicated a far greater hearing loss than their pure tone audiograms. The neural network therefore predicted hearing ability inaccurately as much more hearing-impaired in most of the cases demonstrating noise-induced hearing losses, such as in the cases of ear 73, 74 and 91.

6.10.1.2 Subjects Demonstrating Very Mild Hearing Losses

Subjects demonstrating very mild hearing losses, especially if only some of the frequencies were hearing impaired, were often predicted as normal. Smurzynski et al. (1990) also referred to a subject with a very mild hearing loss that had normal DPOAEs. Probst and Hauser (1990) and Gorga et al. (1996) found that a mild hearing loss is sometimes not detected with DPOAE measurement. If the DPOAE measurements of a specific ear are within the normal range, the neural network will predict that ear as normal. For this study, all eleven DPOAE frequencies were used for input information. If most of the DPOAE responses were normal, it often happened that the neural network predicted all the frequencies as normal, as in the case of ear 19 and ear 33.

6.10.1.3 A Subjects Demonstrating A Possible Retrocochlear Hearing Loss

Another example of poor correspondence between DPOAE results and the pure tone audiogram was that in a few instances, DPOAE results appeared much better than the hearing ability depicted in the audiogram. This could possibly be one of the small percentages of subjects that could be classified as having a retrocochlear pathology based on otoacoustic emission results (Robinette, 1992). In the case of ear 71, the neural network predicted hearing ability as normal at 500 Hz, 1000 Hz and 2000 Hz, but ear 71 demonstrated a mild hearing loss. (The audiogram and predictions of ear 71 can be seen in Figure 6.4.) The DPOAE measurements, when compared with the normal range of DPOAE measurements, also indicated that DPOAEs were much better than would be expected from an ear with a mild hearing loss. Even though other site-of-lesion tests were not performed on this ear, the fact that the DPOAE measurements and neural network predictions were so much better than expected could be as a result of retrocochlear pathology. The outer hair cells on the basilar membrane in the cochlea could therefore be normal and capable of normal DPOAEs.

The next section discusses all the subject, DPOAE and ANN variables that were investigated to determine their effect on prediction accuracy.

6.11 Subject-, DPOAE- and ANN-Variables Experimented with to Determine Optimal PTT Prediction Accuracy.

The subject variable that was experimented with was AGE, presented to the network in 5-year categories or 10-year categories. DPOAE variables include the threshold of the distortion product defined as 1, 2 or 3dB above the noise floor, the frequency information of the DPOAE to use for ANN input, and the amplitude of the DPOAE presented to the neural network in a number of different ways. ANN variables include experimentation with the number of middle neurons and different error tolerances during training and prediction.

6.11.1 The Effect of the Subject Variable AGE Presented to the Network in 5 year or 10 year Categories on PTT Prediction Accuracy.

It seems that the investigation of the influence of age on the distortion product has been problematic for many authors (Avan & Bonfils, 1993; He & Schmiedt, 1996; Karzon et al. 1994; Nieschalk et al. 1989). The reason for this is, that it is very difficult to determine how much of the differences in the distortion product observed in elderly subjects are due to age, and how much is due to sensitivity changes associated with aging. It seems that these authors agree that the negative correlation between DPOAE levels and age is due to changes in hearing threshold associated with aging rather than age itself. However, some researchers found that when the age variable was included in pure tone prediction studies with DPOAEs, age had a positive effect on prediction accuracy (Lonsbury-Martin et al. 1991; Kimberley et al. 1994a; Kimberley et al. 1994b, De Waal, 1998). The age variable was therefore always included in this study, but in two different ways, either depicted with the dummy variable technique into one of 20 possible 5-year categories, or into one of 10 possible 10-year categories.

The results for differences in prediction accuracy for these two methods were given in Figure 5.5. Prediction accuracy for the two methods was always within 5%. The presentation of age as within 5 years, or within 10 years therefore did not have a great effect on prediction accuracy. There may be more than one reason for this. The first possible reason is the fact that even though age was presented more specifically to the network in the 5-year category method, the extra inputs that were needed to represent this information contributed to a more complex neural network topology. Many more inputs were needed which made the data set more complex and it took the network longer to converge. The advantage of a more specific age presentation might have been lost in the creation of a more complex data set. The second possible reason is that the differences in DPOAEs due to age related changes might not a phenomenon that is greatly influenced by only 5 years: the age related difference between a subject only five years older than another subject with identical PTTs might be too small to detect or to make a big difference in PTT prediction accuracy. It would make an interesting study to investigate the effect of age on prediction accuracy with DPOAEs and PTTs by including age categories larger than 10-year categories, such as 15-year and 20-year categories to find the cutoff point where age is not presented specifically enough anymore to have a positive effect on prediction accuracy.

6.11.2 The Effect of DPOAE Threshold Defined as 1, 2 or 3 dB Above the Noise Floor on PTT Prediction Accuracy.

As was described in 4.6.5.2 “Determination of optimal stimulus parameters”, the acceptance of a DPOAE test was either that the cumulative noise floor had to be –18dB SPL or if noise levels were higher, that the DPOAE amplitude had to be 10 dB above the noise floor. For about half of the tests run (47%), noise levels were low enough that test acceptance was achieved by the –18dB SPL cumulative noise level criteria. For all these cases, a DPOAE threshold had to be determined. Other researchers specified DPOAE threshold as 5 dB above the noise floor (Krishnamurti, 2000) or 3dB above the noise floor (Lonsbury-Martin, 1994). For this research project, it was investigated to see if a lower threshold for cases that passed the first criterion of cumulative noise levels of –18dB SPL would have more accurate predictions as a result. There were experimented with 1, 2 and 3 dB above the noise floor. Results were summarized in Figure 5.6.

It was very interesting to note that there was virtually no effect on prediction accuracy for threshold defined as 1, 2 or 3 dB above the noise floor. A threshold of 1dB could predict hearing ability in normal and impaired ears as well as a threshold of 3dB. All predictions were within 1%. This finding leads the researcher to believe that the definition of DPOAE threshold may be lowered to 1dB. A lower DPOAE threshold might improve prediction accuracy for other studies if valuable DPOAE data close to the noise floor is not discarded but used as valid present responses.

6.11.3 The Effect of the Emission or Inclusion of Low Frequency DPOAE Information for ANN Training on PTT Prediction Accuracy

For experiments where low frequency DPOAE data was omitted, f1= 500 Hz, 625 Hz and 781 Hz were omitted in the input data to train the neural network on to investigate if certain frequencies could be predicted more accurately. The summary for PTT prediction accuracy with DPOAEs and ANNs where noisy low frequency emissions were omitted or present was given in Figure 5.7.

Results with low frequencies present or absent revealed only a slight difference. Differences in prediction accuracy were always within 2%. All frequencies, except

2000 Hz, were even predicted slightly more accurately when low frequency DPOAEs were present. It seems like the "noisy" low frequency DPOAEs had very little effect on the training or prediction capabilities of the neural network, which confirms the viewpoint of Blum (1992) that neural networks excel in dealing with noisy incomplete data. Nelson and Illingworth (1991) described the nature and strengths of artificial neural networks very effectively: "The characteristics of intuition, prediction and statistical pattern recognition allow neural networks to deal with situations in which input data may be fuzzy, incomplete, or ambiguous, or may even have some corrupted data. Because most of the data in this world is inexact, this characteristic becomes highly significant." (Nelson & Illingworth, 1991:69). Blum (1992) even suggested training a network with "noisy" data to enhance generalization capability of the network's predictions. It seems like the presence of noisy low frequency data in ANN training had a slightly positive effect on prediction accuracy.

6.11.4 The Effect of DPOAE Amplitude Presentation to the Neural Network on PTT Prediction Accuracy

For this research project, the amplitude of the DPOAE was presented to the network in four different ways: AMP 100, AMP 40, ALT AMP and No AMP (see 4.8.1.4 for descriptions of characteristics for each representation.) These results were summarized in Figure 5.8.

As was described in 4.8.1.4, each method of amplitude representation influenced the ANN in a unique way. AMP 100's neural network had trouble converging (reaching optimal error tolerance levels during training to begin prediction), probably because the amplitude was depicted as a fraction of 100, which made the number rather small (a DPOAE response of 30 dB which can be regarded as quite a strong DPOAE was depicted as 0.3). The largest DPOAE response measured in this study was 39dB, so no amplitude input in the AMP 100 method were ever depicted as more than 0.4. The significance of the size of DPOAE amplitude might have been lost in this method and that might be one reason why it took the network so long to converge. As a result of this, AMP 100 never had the most accurate prediction at any frequency.

AMP 40's ANN had the exact same topology than the AMP 100 method but convergence was much faster because the input values were larger (a fraction of 40 instead of a fraction of 100) and the network therefore found it easier to make midway representations to reach error tolerance levels faster. The same 30dB response would now be depicted as 0.75. This method created more "room" to distinguish between DPOAE amplitudes of various sizes and indicate the significance thereof. AMP 40 had the most accurate predictions of overall prediction accuracy at 500 Hz, 1000Hz, and 2000 Hz. At 4000 Hz, AMP 40 was second in accuracy, only 2% less accurate than the No AMP method. At 2000 Hz, AMP 40 was most accurate for identification of very good hearing (0-15d) and normal hearing (0-25dB).

The ALT AMP technique had the one advantage that information was presented to the network in the same fashion than the output predictions, which was by depicting information in categories with the dummy variable technique. Input mode and output mode for that neural network was therefore the same. A Disadvantage of the ALT AMP method however, was that the complexity of the topology of the network increased drastically due to the fact that 352 input neurons were needed to present amplitude in this fashion instead of the usual 88. This method was able to predict very good hearing (0-15dB) and normal hearing (0-25dB) most accurately at 500 Hz. ALT AMP came a close second to No AMP in the prediction of very good and normal hearing at 1000 Hz (within 0.8%).

With the No AMP method where the amplitude of the distortion product was left out entirely, most accurate predictions were made at prediction of normal and very good hearing at 1000 Hz and 4000Hz. At 4000 Hz, overall prediction accuracy was also best for the No AMP method.

When it comes to the overall prediction accuracy across all categories spanning normal hearing and hearing loss, the AMP 40 method seems to be the best way to present amplitude to the neural network. When it comes to the prediction of normal hearing (0-25dB) or very good hearing (0-15dB), then the No Amp and ALT AMP methods, work the best.

The finding that the No AMP method showed good prediction for normal hearing is consistent with the findings of Moulin et al. (1994) who concluded that DPOAE threshold is a more sensitive predictor of hearing than DPOAE amplitude. It could be argued that in the case of normal hearing, DPOAE thresholds are low enough to predict hearing ability without any amplitude information, such as in the No AMP method.

When it comes to the prediction of hearing loss, however, it seems that the inclusion of amplitude data had a positive effect. The fact that hearing loss categories were predicted best with the AMP 40 method in this study is consistent with the findings of

Probst and Hauser (1990) who found that DPOAE amplitude is strongly correlated with PTTs in hearing impaired ears.

Future studies that attempt to predict PTTs in hearing-impaired ears should therefore include both the threshold and amplitude information of the distortion product to ensure optimal prediction accuracy.

6.11.5 The Effect of the ANN Variable MIDDLE NEURON COUNT on PTT Prediction Accuracy

As was stated in chapter 4, the size of the hidden layer is function of the diversity of the data (Blum, 1992). The number of middle layer neurons determines the accuracy of prediction during the training period. With an insufficient number of middle neurons, the network is unable to form adequate midway representations or to subtract significant features of the input data (Nelson & Illingworth, 1991). With too many middle neurons the network has difficulty to make generalizations (Rao & Rao, 1995; Nelson & Illingworth, 1991). The number of middle layer neurons was determined by trial and error, based on the accuracy of the prediction during the training period. If the neural network was unable to converge during training, the number of middle level neurons was increased and the prediction attempted again.

Various network runs were conducted with varying numbers of hidden layer neurons, ranging from 20 to 180. With a number of hidden layer neurons below 60, the network was unable to extract significant features from the input data and sometimes would not converge during training. A number of hidden layer neurons more than 160 resulted in poor generalization ability. For this study, three possible middle neuron counts were identified, namely 80, 100 and 120 and experiments were run with all three possibilities to determine its effect on prediction accuracy.

Results for prediction accuracy for the three middle neuron counts were summarized in Figure 5.9. Prediction accuracy for all three middle neuron counts was always within 1%. There was therefore no significant enhancement for prediction accuracy when middle neuron counts were slightly increased.

Many authors agreed that there is not a clear-cut formula for the determination of the optimal number of middle level neurons for an artificial neural network but that it is determined by the complexity of the data and should be determined with trial-and-error experimentation (Rao & Rao, 1995; Blum, 1992; Nelson & Illingworth, 1991; Fujita, 1998).

It seems that all three middle level neuron counts for this study worked equally well and any one of the three middle level neuron counts can be applied to this research project for the prediction of PTTs with DPOAEs.

6.11.6 The Effect of the ANN Variable TRAINING ERROR SENSITIVITY on PTT Prediction Accuracy

Another neural network variable is the acceptable error during training. A neural network learns from its mistakes. The first step in the learning process is to compute the outputs, the second step to compare the outputs with the desired answers and the last step to adjust the set of weights to enable a better prediction the next time. The second step, namely the comparing of outputs with desired answers, can be made in various levels of accuracy. The neural network can be required to make predictions within any percentage of accuracy during the training stage. The normal assumption would be that the more accurate the prediction during training, the more accurate the predictions would be. It is however often found that if training error is set too close to zero, that generalization abilities of the network decreases and that it experience difficulty predicting cases slightly out of the ordinary (Rao & Rao, 1995).

Error tolerance levels for this study were set at 0.001, 0.002 and 0.003. Results were summarized in Figure 5.10. Prediction accuracy for the three tolerance levels was always within 1%. For this network configuration, the three error tolerances performed equally well.

6.12 Summary

Other studies	**This study**
In the investigation of DPOAEs as a possible new hearing screening or diagnostic procedure, many authors found a correlation between DPOAEs and PTTs (Gaskill & Brown, 1990; Gorga et al. 1993; Kummer et al. 1998; Lee et al. 1993; Vinck et al. 1996).	Results of this study confirmed the correlation between DPOAEs and PTTs and found that artificial neural networks found a significant correlation between DPOAEs and PTTs to enable ANNs to predict PTTs accurately given only DPOAEs.
Some researchers predicted hearing ability as normal or hearing-impaired at various frequencies, with more success in the high frequencies (Kimberley et al. 1994; Moulin et al. 1994). There were no reports of prediction at 500 Hz due to difficulties experienced with the rising noise floor below 1000Hz (Gorga et al. 1993; Moulin et al. 1994; Stover et al. 1996a).	Normal hearing ability was accurately predicted at frequencies as low as 500 Hz in the present study. Prediction accuracy for normal hearing was 94% for 500 Hz, 88% for 1000 Hz, 88% for 2000 Hz and 93% for 4000 Hz. These predictions are better than reported elsewhere (Kimberley et al. 1994 a+b; Moulin et al. 1994).

Even though this study could correctly identify normal hearing quite accurately, even at 500 Hz, the specific predictions of impaired hearing levels at various frequencies were rather disappointing.

One possible reason for the poor prediction of categories depicting hearing loss is the limited number of ears in every category that the neural network had to train on. With closer analysis of the possible correlation between prediction accuracy and data quantities, it became clear that the neural network would perform much better with more data to train on.

This finding serves as a strong indication that DPOAEs can be used to predict hearing status in populations with varying degrees of hearing loss objectively and accurately. **The results in this study forms a solid foundation stone for further research to continue the quest for optimal pure tone prediction with distortion product otoacoustic emissions.**

Chapter 7: Summary, Evaluation of the Study and Conclusion

7.1 Summary

An overview of current objective diagnostic procedures revealed that many technological advanced tests exist for the successful evaluation of hearing ability and site of lesion testing in adults. It is in the evaluation of difficult-to-test populations however, that limitations in current objective diagnostic procedures were identified. It seemed that, despite all the strengths and positive attributes of ABR, tympanometry, MLR, and LLR, weaknesses in these techniques made it difficult to measure exact hearing ability and site-of-lesion in populations such as neonates, infants, malingerers, the crucially ill and foreign speakers. Some of these weaknesses include a limited frequency area in which hearing ability can be determined, lengthy test times, the unfortunate possibility of sedation and the level of expertise and expense required for successful implementation of the diagnostic test battery and interpretation of results (Ferraro & Durrant, 1994; Musiek et al., 1994; Robinette, 1994; Weber, 1994). It is therefore with much hope that many researchers turned their investigations to otoacoustic emissions.

Kemp (1978) identified different classes of otoacoustic emissions, depending on the stimuli used to evoke them. Spontaneous otoacoustic emissions (SOAEs) are only prevalent in half of normal hearing persons and can therefore not be implemented as a screening test or diagnostically (Lonsbury-Martin, 1994; Norton & Stover, 1994). Stimulus frequency otoacoustic emissions (SFEs) are not currently clinically used due to difficulties in separating in-going stimuli and out-going emitted responses

(Lonsbury-Martin & Martin, 1990). Transient evoked otoacoustic emissions (TEOAEs) have been proven as a clinical acceptable hearing screening procedure, but the fact that they are only recordable in normal ears limited their diagnostic hearing applications (Kemp & Ryan, 1993; Lonsbury-Martin et al., 1992; Stevens et al., 1990). Distortion product otoacoustic emissions (DPOAEs) on the other hand, revealed many possibilities as a potential test of auditory functioning. First, it has been proven useful in both clinical and research settings, for it is the only emission type that can easily be recorded in many laboratory animals, allowing for experimental control of certain factors (Mills, 1997). Second, it can be measured in ears with a hearing loss of up to 65dB HL, therefore revealing information regarding outer hair cell functioning of hearing-impaired populations as well (Moulin et al., 1994). Third, it is the emission type that can most easily be compared to the pure tone audiogram, due to the pure tone nature of the stimuli that can be chosen to represent pure tone audiogram frequencies (Durrant, 1992; Lonsbury-Martin & Martin, 1990). Fourth, DPOAEs correlate well with pure tone thresholds and the configuration of the hearing loss (Durrant, 1992; Kimberly & Nelson, 1989; Stover et al., 1996a). Fifth, DPOAEs are not significantly influenced by gender (Cacace et al., 1996; Karzon et al., 1994).

Many studies described the relationship between DPOAEs and pure tone thresholds (Avan & Bonfils, 1993; Bonfils et a; 1991; Gaskill & Brown, 1990; Gorga et al., 1993; Kimberley et al., 1994b; Probst & Hauser, 1990; Stover et al., 1996a). Statistical methods used to date, such as multivariate (discriminant) analysis in the case of the study of Kimberley et al., (1994b), but also in all the other studies previously named, indicated a correlation between DPOAE measurements and behavioral pure tones. These studies however, could not predict the actual pure tone thresholds given only the distortion product responses (Lee et al., 1993). The complexity of the data, the numerous variables involved and the possibility of a nonlinear correlation have been some of the reasons why conventional statistical methods could not predict pure tone thresholds given only DPOAEs, but only distinguish between normal hearing and hearing-impaired ears (Kimberley et al. 1994a; Kimberley et al. 1994b).

For this study, a mathematical model, called **artificial neural networks** (ANNs), was used to investigate the relationship between DPOAE measurements and pure tone thresholds. This technique has excellent correlation finding capabilities, even in the case of a non-linear correlation. The neural network was used to predict pure tone thresholds given only the distortion product responses.

ANNs were initially developed to gain a better understanding of how the human brain functions (Nelson & Illingworth, 1991). It is an algorithm for a cognitive task, such as learning or pattern recognition (Muller & Reinhardt, 1990) and can be used to make predictions based on learned correlations from previous similar data (Blum, 1992). Various disciplines became interested in the use of ANNs to address complex problems in the last two decades, ranging from cognitive psychology, physiology, medicine, computer science, electrical engineering, economy and even philosophy. ANNs have been successfully applied to the field of Speech Pathology for speech recognition purposes (Metz, et al., 1992) and Audiology to predict normal hearing ability (Kimberley et al. 1994a). It was therefore with great expectations that neural networks were applied to the field of diagnostic Audiology (Kimberley et al. 1994a; De Waal, 1998).

The rationale for the current study was to further investigate DPOAEs as a possible new objective test of hearing by improving prediction accuracy of PTTs with DPOAEs and ANNs. Factors that influenced DPOAEs levels were identified (such as the frequency ratio and levels of the primaries) and were controlled to ensure optimal DPOAE responses. Factors that influenced ANN efficiency were identified (such as middle level neuron count, input format and error tolerance levels) and were manipulated to reach optimal prediction accuracy levels. The correlation between the measured variables of DPOAEs and PTTs were studied with artificial neural networks and then used to predict hearing ability at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz with DPOAEs.

Data was obtained from 70 **subjects** (120 ears, in some cases only one ear fell in the subject selection criteria), 28 males and 42 females, ranging from 8 to 82 years old. Selection criteria included normal hearing, sensorineural hearing losses of varying degrees and normal middle ear functioning. The subject selection procedures included a short case history, otoscopic examination, tympanometry and a pure tone audiogram.

The distortion product has numerous stimulus variables that should be specified to ensure optimal testing conditions. The choice of **stimulus parameters** in this study was based on an extensive literature study and confirmed in the preliminary study. For this research project, eight DP Grams at 5dB intervals ranging from L1=70dB SPL to

L1=35dB SPL were measured. A frequency ratio of 1.2 was selected for the two primaries and the loudness level ratio of the two primaries was L1>L2 by 10dB. The frequency range of F1 = 500 to F1 = 5031 was tested. The **criterion for DPOAE presence** was that test conditions had to be "accepted" which implies either a DPOAE response of 10dB above the noise floor or a cumulative noise level of – 18 dB SPL. For the cases where the noise level was low enough to pass test acceptance conditions for a – 18 dB SPL noise floor, **DPOAE threshold** was defined as 1, 2 or 3 dB above the noise floor to investigate the effect of a lower threshold criterion as stated in the literature on prediction accuracy (the lowest criterion in literature was set by Lonsbury-Martin, 1994, as 3dB above the noise floor).

Eight tests or DP Grams were performed in each ear. Every DP Gram consisted of eleven frequency pairs. Every frequency pair consisted of two pure tones, f1 and f2 presented to the ear simultaneously.

A **backpropagation network** was chosen for this study for two reasons: **1)** A nonlinear correlation is suspected between DPOAE thresholds and traditional pure tone thresholds. Metz, et al. (1992) reported the backpropagation neural network to be very successful in dealing with nonlinearities that potentially occur in complex data sets. According to Blum (1992), the backpropagation neural network is capable of nonlinear mappings and able to generalize well. **2)** The purpose of this study is to predict pure tone thresholds with distortion product thresholds with the use of neural networks. According to Blum, (1992) and Tam and Kiang, (1993), the backpropagation neural network is highly applicable in the areas of forecasting and prediction.

Several different **trial runs** were conducted to determine neural network topology and the way the input data should be presented to the neural network. The input data was presented to the neural network in a binary mode, and the data pattern of all absent and present DPOAE responses as well as DPOAE amplitude, subject age and gender served as input stimuli. Hearing ability was divided into categories and the neural network had to predict hearing ability into one of the 10dB categories.

Data analysis consisted of analyzing the actual and predicted values of all 120 ears and to determine how many were predicted accurately, how many within one class and how many were predicted incorrectly.

Results indicated that normal hearing ability could be distinguished from hearing-impaired hearing quite accurately, on all selected frequencies and as low as 500 Hz. Many researchers failed to predict normal hearing ability at 500 Hz due to the rising of the noise floor at the lower frequencies (Gorga et al., 1993; Moulin et al., 1994; Probst & Hauser, 1990; Stover et al. 1996a). **In this study, normal hearing ability at 500 Hz was predicted accurately 94% of the time. Normal hearing at 1000 Hz was correctly identified 88% of the time, at 2000 Hz 88% of the time and at 4000 Hz 93% of the time.**

In the **previous study** (De Waal, 1998), normal hearing at 500 Hz was predicted with 92% accuracy, 1000 Hz with 87%, 2000 Hz with 84% and 4000Hz with 91% accuracy. False positive and negative rates were also higher than in the present study. It seems that differences in neural network topology and input data manipulation from the previous study to the present, had a positive effect on prediction outcome. It should also be noted that the accuracy of the previous study's results could have been influenced by the fact that a subject's related ear was included in the training set, the neural network had an "unfair" advantage to be trained with a subject's one ear that might look very similar to the ear to be predicted, such as in the case of noise exposure. For the present study, both ears of a subject were removed from the prediction set. Despite this disadvantage, more accurate predictions were made at all four frequencies with less false positive and false negative responses.

The **false negative values** in this study are lower than reported elsewhere (Gorga et al., 1993; Moulin et al., 1994; Probst & Hauser, 1990; Stover et al. 1996a), but are still considered too high for optimal sensitivity and specificity for screening and diagnostic purposes.

The improvement of predictions of normal hearing at the four frequencies from the previous study to the present study can be attributed to a few reasons: First, the previous study struggled to incorporate the amplitude of the DPOAE response into training data and eventually used only the presence or absence of a response as neural network input. The present study experimented extensively with ways to incorporate DPOAE amplitude and successfully presented this information to the network. Secondly, there were experimented more extensively with neural network topologies, and ways to present input data to the network and more optimal procedures have been discovered and used in the prediction of PTTs with DPOAEs. It could be speculated that even more extensive experimentation could lead to better results than obtained in the present study.

Predictions of categories depicting hearing impairment were not satisfactory. Subjects were initially selected in such a manner that there were 40 ears with normal hearing, 40 ears with mild hearing loss and 40 ears with moderately severe hearing loss. The distribution of hearing loss at the four frequencies, however, resulted in an unequal amount of data in the different categories. Some categories were poorly represented and the neural network did not have enough data to train on, resulting in inaccurate predictions. In retrospect it also seems that the **number of subjects** included in this study, was too small to provide enough ears in every prediction category. Figure 6.2 clearly demonstrates that there is a threshold for the optimum number of ears in a category to enable a neural network to make accurate predictions. When less than about 32 ears were present in a category, prediction accuracy was very poor (less than 50%), but all categories with more than 32 ears showed a significant increase in prediction accuracy (more than 75%). If this hypothesis is true, then very accurate prediction of hearing ability from 500 - 4000 Hz for hearing losses up to 65dB HL can be expected with DPOAEs and ANNs where categories are well represented. **Despite the limited data set and shortage of ears in most categories, normal hearing was still identified with high levels of accuracy with DPOAEs and ANNs and this study effectively proved that DPOAEs can be developed as a diagnostic test of hearing that meet all the requirements for such a procedure.**

Interesting cases were identified that had irregular neural network predictions. Cases that were exposed to long periods of noise were always predicted as more hearing impaired, cases with possible retrocochlear hearing losses were predicted as having normal hearing due to their normal emissions and cases with minimal hearing losses were sometimes predicted as normal. These irregularities once again stress the importance of the case history as part of the diagnostic battery.

7.2 Evaluation of Research Methodology

The evaluation of research methodology will discuss the research design, the validity and reliability of this study and some limitations that were identified.

7.2.1 The Research Design

The research design chosen for this study was a multivariable correlational study. The correlation between the measured variables of DPOAEs (DPOAE thresholds at eleven 2f1-f2 frequencies) and PTTs (thresholds at 500, 1000 2000 and 4000 Hz) was studied with artificial neural networks and then used to predict hearing ability in normal and sensorineural hearing impaired ears up to 65 dB HL. Factors that influenced DPOAEs levels were identified (such as the frequency ratio and levels of the primaries) and were controlled to ensure optimal DPOAE responses. Factors that influenced ANN efficiency were identified (such as middle level neuron count, input format and error tolerance levels) and were manipulated to reach optimal prediction accuracy levels. DPOAE responses, DPOAE amplitude, subject age and subject gender formed the input of a feedforward neural network with a backpropagation learning algorithm. The network was trained with data of 118 ears to "learn" the correlation between DPOAE and PTT responses, and then used the learned correlation to predict an unknown subject's PTTs given only the subject's DPOAE responses. It has been proved that it is possible to predict audiometric thresholds with DPOAEs and ANNs, based on the correlation that the network found between the two sets of data and then used to make the prediction. The multivariable correlational study method in this research project was therefore applied successfully.

7.2.2 Validity and Reliability

Ventry and Schiavetti (1980) identified several factors that can influence the validity and reliability of data. The validity of the data can be divided into internal validity and external validity.

Internal validity deals with factors such as history, where the amount of time elapsed between the first and last test could include certain factors such as medication or treatment which could affect the readings of the second test differently than the first test. To preclude the influence of this fact on test data, the pure tone audiogram, tympanogram and distortion product measurement were performed in one session, lasting about an hour.

Internal validity also deals with instrumentation. The accuracy of the data obtained for the pure tone audiogram is a result of how well the audiometer was calibrated, how recently the audiometer has been calibrated, and the cooperation of the subject (Leedy, 1993). The audiometers used in this research project (calibrated annually) were calibrated less than a year before this project. Pure tone thresholds were double checked with speech reception thresholds when poor cooperation of the subject was suspected, the instructions for pure tone audiometry was repeated and a threshold was determined as 3 responses out of 6 stimuli presented.

The GSI 60 DPOAE system was calibrated for a particular quiet room. Regarding the fit of the probe, closure was obtained or DPOAE testing even though closure is not considered necessary but helpful by some authors (Bright, 1994). The closure fit of the probe reduced any external noise.

Another factor that influences internal validity according to Ventry & Schiavetti, (1980) is the differential selection of subjects. The subjects selected for this study were divided into three groups, normal hearing, slight to mild hearing losses and moderately severe sensorineural hearing losses. The subjects were selected carefully to ensure that no other factors than a sensorineural hearing loss is present that could influence the test data such as middle ear pathology in this case. Tympanograms were interpreted carefully to ensure normal middle ear pressure. Subjects that had normal hearing but no tympanogram due to a perforation in the tympanic membrane were not included in the study. The selection of subjects was strictly according to the subject selection criteria as set out in Chapter 4.

Reliability deals with the accuracy of the data obtained (Leedy, 1993) or precision of measurement (Ventry & Schiavetti, 1980). Reliability can be assessed by examining the stability and consistency of the test or measure. Gaskill & Brown (1990) conducted a study to investigate stability and reproducibility of DPOAE Grams over time and with different ear probes. These authors found DPOAE measurements to be extremely stable over time and that different probe fits do not significantly influence DPOAE measurements. DPOAE measurements therefore seem to be reliable. The fact that DPOAEs are so reliable makes it an ideal procedure to monitor cochlear function in Meniere's disease, the administering of ototoxic medication or during surgery of structures close to the cochlea (Cane, Donoghue & Lutman, 1992; Subramaniam, Henderson & Spongr, 1994; Teleschi, Roth, Stagner, Lonsbury-Martin, 1995a; Teleschi, Widick, Lonsbury-Martin & McCoy, 1995b).

Neural networks are also very reliable. Two neural network runs with exactly the same inputs yield exactly the same results (Blum, 1992).

Hall III et al. (1993), identified more factors influencing measurement and analysis, and therefore the validity and reliability of the study. First, it is important to determine the status of the middle ear and external ear canal, for DPOAEs depend on both an inward and outward propagation of stimulus energy. These two factors were carefully assessed during subject selection procedures. Second, the measurement parameters for DPOAEs should be carefully chosen to ensure optimal measurement conditions. In this study, measurement parameters were chosen after an extensive literature study and confirmed with finds during the pilot study.

The last aspect that could have an effect on the validity of this research project is human error during data preparation, analysis, and processing. Human error was eliminated or reduced where possible by electronic preparation, processing and analysis of data. DPOAE results were read into excel directly from the GSI-60 DPOAE system database to eliminate human error during the creation of subject files in data preparation. The computer extracted data that was used for the training of the neural network. Data analysis, where the correct answers were compared to the predicted answers, was also conducted on the personal computer to eliminate human error. Even the Figures, depicting prediction accuracy and correlation between number of ears and prediction accuracy were done on the computer with data directly from excel.

According to Leedy (1993) validity investigates the end results of the measurement. "Are we really measuring what we think we are measuring?" (Leedy, 1993:41). This research project attempted to find a correlation between DPOAEs and PTTs with ANNs and to use that correlation to predict PTTs with DPOAEs. It can be stated with reasonable certainty that this research project did in fact do what it was intended to do. Reliability, according to Leedy (1993) deals with the accuracy of the measurement. All measurements in this study were measured as accurately as technology currently allows on calibrated equipment.

7.2.3 Limitations of the Study

A few limitations were identified in this study. These are all aspects that should be kept in mind in the interpretation of results.

First, as stated previously, some of the categories depicting hearing impairment, were not represented adequately by the amount of data that the neural network had to train on. Even though subjects were initially selected to include an equal number of ears in three different categories of hearing ability, the pattern distribution of many sensorineural hearing losses is such that hearing loss is more prevalent in the higher frequencies than in the lower frequencies (Yantis, 1994). This resulted in an unequal number of ears in the categories that the neural network had to predict. Some of the categories were represented so poorly, that the neural network did not have enough data to train on. The network could not form adequate midway representations of the hearing ability of a subject in a category where only a few examples were present. To address this problem, more subjects should be included in neural network studies to ensure more data in every category.

The second limitation is the fact that this study did not investigate every possible neural network type and configuration available to determine their effectiveness as predictors of hearing ability. There are so many combinations of neural network configurations available, and even though numerous combinations were tried and tested for this application, it cannot be stated with certainty that this network type and configuration is the optimal choice. It is quite possible that better results can be obtained with other neural network types, or different topologies.

The third limitation is the high incidence of false negative responses recorded in this study. This high incidence of false negatives influences the sensitivity, and therefore the clinical acceptability of DPOAEs as a potential screening or diagnostic procedure. Further research is necessary to investigate possible neural network runs with different topology, different inputs and better measurement of DPOAEs to attempt to lower this high rate of false negatives. Changes in network topology had a significant improvement on false negative values from the previous study to the present, and further research might possibly reveal that DPOAEs can be used with ANNs to predict PTTs within acceptable levels of false negative responses for diagnostic purposes.

The last limitation identified in this study is the length of time needed for DPOAE measurement to obtain adequate information for one ear. The way in which this research project was constructed, 8 DP Grams were conducted in each ear. The pattern of all present and absent DPOAE responses from all eight DP Grams was used as input information. The duration of one DP Gram was about 2 minutes. It therefore took about 15 minutes per ear to obtain the necessary information. Even though it is still only half the time that is required to obtain a single threshold for one ear in ABR testing (Weber, 1994), it could be argued that 15 minutes is not such a rapid test of auditory functioning as was hoped for.

7.3 Recommendations for Future Research

There are a number of recommendations for studies attempting to improve PTT prediction accuracy of DPOAEs with ANNs using the method described in this study:

- The first recommendation is to increase the number of subjects. Figure 6.2 indicated that the “threshold” number of ears in every category to enable prediction accuracy of more than 75% is around 32 ears. With around 80 or more ears in a category, prediction accuracy could be expected to be very accurate, possibly more than 95%.
- The second recommendation is that the application of neural networks to this particular field of Audiology should be further investigated. Neural networks offer so many possibilities. It is possible that different types of networks or different types of configurations would yield more accurate predictions or lower false negative values. Slightly higher error tolerance levels such as 1, 2 or 3% may be able to generalize better and may improve prediction accuracy. It is also possible that other areas in Audiology could also benefit from artificial neural networks.

- The third recommendation is to experiment with lower "acceptance" criteria for DPOAE measurements, possibly a cumulative noise level slightly higher than –18 dB SPL (such as –10 dB SPL) or a DPOAE level slightly lower than 10dB above the noise floor (such as 5dB). If the larger number of responses obtained in this method is invalid and part of the irregular noise floor, then prediction accuracy of the neural network will be less accurate. In this case, the higher standards for test acceptance conditions can always be reintroduced, and neural network runs can continue with the higher standards, such as a DPOAE level of 10dB or a cumulative noise level of –18dB SPL and invalid responses can be discarded. If however, neural network prediction is more accurate due to the presence of more responses regarded as accepted, it could possibly influence the efficiency of DPOAE testing as it is currently conducted. When testing with higher "acceptance" criteria levels, responses regarded as "timed out" or "noisy" are lost forever. When testing with lower criteria, the validity of responses can always be determined later. It is possible that this technique might identify new levels for acceptable DPOAE measurement and improve prediction accuracy.

7.4 General Implications of the Study and Concluding Remarks

Audiologists are currently relying heavily on objective audiological tests to assess hearing ability in difficult-to-test populations. There are however, still many limitations in current objective procedures despite the enormous progress in the last few decades. Some of these limitations include the limited frequency area of objective hearing assessment, the expenses, time and expertise required, and the possibility of sedation. It is with much hope that many researchers turned to the investigation of

DPOAEs as a possible new rapid, objective, accurate and cost effective test of auditory functioning. The distortion product has been proven as an acceptable screening procedure. Otoacoustic emissions however, have never been used as a diagnostic test of hearing where specific thresholds for frequencies were determined due to shortcomings in conventional statistical methods (Kimberley et al., 1994; Lee et al., 1993).

The investigation of DPOAEs indicated strongly that DPOAEs are suitable as a diagnostic audiologic test of hearing. It is suggested that pure tone thresholds can be accurately predicted within 10dB as low as 500 Hz and for hearing levels of up to 65dB HL with ANNs. The successful application of ANNs in this field of Audiology opened the door to the development of an objective, rapid, accurate and economical test of hearing to aid in the assessment of difficult-to-test populations. It is strongly believed that this breakthrough will play a leading role in the efficiency with which the pediatric population will be assessed in the next decade.