6 Discussion 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 and confirm these findings with a series of objective electrophysiologic 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 1, however, it is evident that there are some weaknesses in current objective diagnostic procedures. In the evaluation of special populations such as neonates from birth to 6 months, the crucially ill, and malingerers, audiologists often have to rely heavily on the objective electrophysiologic procedures to determine hearing ability. To determine hearing thresholds with electrophysiologic procedures is often costly, requires a large amount of time and 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 was investigated as a possible new test of hearing. It was attempted to predict pure tone thresholds at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz with the use of artificial neural networks. Very interesting and promising results were obtained. The following Chapter will attempt to discuss the implications of the present study's findings.

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. As stated in Chapter 3, 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 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 set of data that is not correlated with the other 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 distribution curve.

The prediction accuracy of this neural network is illustrated in the histograms in Figure 5.1, Figure 5.2 and Figure 5.3. At first glance the presence of a normal distribution curve can be seen in all these histograms. 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 Average Hearing Ability

Some studies attempting to predict hearing ability as normal or abnormal with DPOAEs have used the amplitude of the distortion product (Gorga et al., 1993). Other studies used DPOAE threshold information (the single response that could be elicited with the lowest primary stimuli, still measurable above the noise floor) (Gorga et al. 1996). Most studies use both amplitude and threshold information by combining these procedures or by performing both separately (Kimberley & Nelson, 1989; Moulin et al., 1994). This study did not use amplitude or threshold data as neural network input, but a complete new approach. For this study, the pattern of all present and absent distortion product otoacoustic emissions in the frequency range f1 = 500 Hz to f1 = 5031Hz over a 35dB span of L2 = 65dB SPL to L2 = 25dB SPL were used. The mere presence or absence of an emission and the pattern thereof across 8 DP Grams and 11 DPOAE frequencies were used as input information for the training of the neural network.

As a first level approach, the average hearing ability of every subject was determined by taking the average of 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz. The neural network was trained with the DPOAE responses of all frequencies and loudness levels and average hearing ability. The network had to predict average hearing into one of the seven 10dB categories of scenario four and also into one of the five categories of scenario five.

Predictions were overall better in scenario five for the prediction of average hearing ability. The overall prediction accuracy of the network (average of all the categories) improved from 40% accurate in scenario four to 52% accurate in scenario five. This improvement is due to the fact that there were more ears in the hearing loss categories of scenario five, than in the hearing loss categories of scenario four. The neural network had more data to train on and could make better representations of what a hearing impaired subject's DPOAE pattern look like. It could be speculated that an increase in hearing-impaired subjects included in this study would improve the accuracy of the prediction, for the network would have more data to train on.

It was clear that the network was able to make more accurate predictions in certain categories. Categories depicting normal hearing, especially category 1 (0-10dB) was always predicted most accurately. There are two possible reasons for this. First, category 1 always had the largest number of ears for the neural network to train on. The network therefore had a better representation of the qualities and pattern distributions of a subject with normal average hearing. Second, it is also likely that subjects with normal average hearing exhibited the best DPOAE responses. It is possible that once the average hearing becomes impaired, there is a decrease in DPOAE responses, resulting in more absent than present responses, making it difficult for the neural network to distinguish between different classes of hearing impairment.

Average hearing in category 1(0-10dB) could be accurately predicted 93% of the time in scenario five, and as normal (into the next 10dB category, 11-20dB) 5% of the time. Prediction of very good average hearing was therefore completely wrong (more than one category out) just 2% of the time. Category 2 (11-20dB) however, could be accurately predicted only 5% of the time, into and adjacent category 58% of the time and was completely wrong 37% of the time. This poor prediction for the upper half of normal hearing ability contributes to the fact that normal hearing (0-20dB) could only be predicted accurately, 85% of the time in scenario five and 87% of the time in scenario four. Ears demonstrating a hearing loss was falsely predicted as normal 16% of the time in scenario four and 11% of the time in scenario five. These high values of false negative responses raise questions regarding the sensitivity of this procedure. A procedure demonstrating such a high false negative rate can not make accurate and reliable predictions regarding average hearing ability.

6.4 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 (0-20dB HL) at 500 Hz was predicted as normal 92% of the time in scenario five. 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, pure tone thresholds were not predicted with DPOAE amplitude or threshold at a single DPOAE measurement, but with a pattern of all present and absent DPOAE responses 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 92% of the time in scenario five.

The purpose of this study however, was to predict pure tone thresholds in hearing impaired categories as well. Hearing ability was divided into categories, and the neural network had to predict the most probable category of hearing ability. 500 Hz was predicted in two scenarios, scenario four, with seven 10dB categories and scenario five, with five categories of different decibel intervals. The differences in results between these two scenarios will be discussed next.

The prediction of 500 Hz was usually better in scenario five than in scenario four, with the exception of a few categories. In scenario four, the accurate prediction of normal hearing at 500 Hz was 87%. The reason for the improved prediction accuracy in scenario five was due to the fact that the neural network had more ears in every category depicting hearing loss to train on. The number of ears in a category had a definite effect on the prediction accuracy of the neural network. This aspect is discussed as one of the variables influencing the outcome of this study, in **6.10.2.2 Amount of Data Available to Train on in Every Category**.

The prediction of hearing ability into specific categories for 500 Hz was rather disappointing. In scenario five, category 3 (21-35dB) could be accurately predicted only 13% of the time and into an adjacent category 47% of the time. The categories depicting hearing impairment of scenario five spanned 15dB, so even if hearing ability was predicted into an adjacent category, it could still be as far as 30dB wrong. Category 4 (36-50dB) was predicted accurately only 25% of the time and category 5 (51-65dB) was never predicted accurately. A possible reason for the poor prediction of categories depicting hearing loss, is the distribution of the number of ears in every category. Although subjects were initially selected in such a manner that their average hearing ability fell into one of three categories, normal, mild, and moderately severe hearing loss, category one and two (depicting normal hearing) had a total of 86 ears (out of 120). This is due to the typical pattern of sensorineural hearing losses, usually involving the higher frequencies (Yantis, 1994). Most of the subjects therefore had normal hearing at 500 Hz even though they might have had a hearing loss at higher frequencies. The neural network therefore did not have a good representation of the characteristics of a DPOAE pattern of a subject with a hearing loss at 500 Hz. Prediction accuracy in categories depicting hearing impairment improved from scenario four to scenario five. This is most likely due to the fact that the number of ears in the hearing loss categories increased and that the neural network had more data to train on. It is very possible, that results would improve even more, if more subjects were included in the study. Another possibility for the poor prediction in hearing impairment categories is that hearing impairment at 500 Hz might result in absent or diminished DPOAE results in the low frequencies. This aspect would make it difficult for the neural network to extract enough data from the pattern of absent and present responses to make a prediction.

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 in scenario five. Category one (0-10dB) had 12% false negative responses, in other words, subjects with hearing losses were predicted as having perfect hearing (0-10dB) 12% of the time. Category two had 8% false negative responses. If the two categories are combined to represent normal hearing (0-20dB), the false negative rate is 20%. Even though these values are lower than reported elsewhere (Gorga et al., 1993; Moulin et al., 1994; Probst & Hauser, 1990; Stover et al. 1996a), the high incidence false negative responses raises questions regarding the sensitivity of this procedure. A false negative value of 20% is too high for a reliable prediction.

As far as the test's specificity is concerned, false positive rates were lower than false negative rates. Normal ears were predicted as hearing impaired only 6% of the time in scenario five.

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 can accurately categorize hearing ability at 500 Hz as normal, 92% of the time.

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 found considerable influence of low frequency noise interfering with test measurements (Gorga et al., 1993; Kimberley et al. 1994; Moulin et al., 1994; Probst & Hauser, 1990). In a study by Kimberley et al. (1994), hearing ability at 1000 Hz could be accurately predicted as normal 71% of the time. Kimberley et al. (1994) did not state the false negative rate for 1000 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. From the previous results and can be seen that many researchers experienced difficulty with the prediction of normal hearing ability at 1000 Hz.

In this study, 1000 Hz could be accurately predicted as normal 87% of the time in scenario five and 84% of the time in scenario four. Again, predictions improved in scenario five when hearing loss categories were enlarged to present the neural network with more data to train on. Very good hearing ability (0-10dB) was predicted most accurately. Very good hearing ability (0-10dB) could be accurately predicted 93% of the time with a false negative rate of 9% in scenario five. Very good hearing ability was predicted into the adjacent 10dB category (11-20dB) 5% of the time, and predicted completely wrong only 2% of the time. Category 2 (11-20dB), however, had less accurate predictions in scenario five. Category 2 (11-20dB) could only be accurately predicted 22% of the time and had a false negative rate of 9%. This poor prediction of the upper half of the normal hearing range affected the prediction of normal hearing ability considerably, bringing it down to 87% in scenario five with the combined false negative rate of 18%. Even though the predictions in this study for 1000 Hz were more accurate than stated elsewhere (Gorga et al., 1993; Kimberley et al. 1994; Moulin et al., 1994; Probst & Hauser, 1990), the high incidence of false negative responses influences the sensitivity of this procedure. Such a high incidence of false negative responses lessen the clinical applicability of this neural network run as a possible screening procedure.

The prediction of specific categories of hearing ability at 1000 Hz was again somewhat disappointing. By enlarging the categories in scenario five and including more ears in every category depicting hearing loss, results improved somewhat and overall prediction accuracy across all categories improved from 54% to 58%. The improvement, however, was not enough to allow for prediction of pure tone threshold of hearing impaired categories at 1000 Hz. In category 3 (21-35dB) scenario five, for example, hearing ability could never be predicted accurately. Hearing ability was predicted into an adjacent category 67% of the time. In scenario five, to predict hearing ability into an adjacent category of category 3 (21-35dB), it means that it was either in category 2 (11-20dB) or in category 4 (36-50). If the neural network predicted hearing ability in category 2 (11-20dB), it means the results could be as far as 24dB out (somewhere between 11dB and 35dB). If the network predicted hearing ability in category 4 (36-50dB), it means the prediction can be as much as 29dB out (between 21dB and 50dB). Category two to four spans from 11dB to 50dB. This is a very broad region of hearing ability, not yielding specific results regarding the hearing ability of 1000 Hz at all. It is therefore very important to attempt to have as many accurate predictions (within the same class) as possible. The 10dB categories of scenario four therefore provide more frequency-specific information regarding predictions into an adjacent class, but in the case of scenario four, less accurate predictions were made because of the shortage of data in some of the categories. It could be speculated that the prediction of 1000 Hz might be better if more subjects with hearing impairment is included in the study, and the neural network is trained with 10dB categories only. Although the prediction of specific regions of hearing impairment was rather disappointing, the results obtained from the prediction of very good hearing ability (0-10dB) and normal hearing ability (0-20dB) were more promising.

6.6 Prediction of 2000 Hz

The prediction of normal hearing at 2000 Hz in other studies yielded far more promising results than the predictions at 500 Hz and 1000 Hz. Kimberley et al. (1994) predicted normal hearing at 2050 Hz with 92% 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 76% of the time with an average false negative response of 15%. Even though predictions of normal hearing were more accurately at 2000 Hz, the false negative rate was still unacceptably high.

The predictions of 2000 Hz for this study yielded the poorest predictions of all four of the frequencies. Normal hearing at 2000 Hz could be predicted accurately only 84% of the time. False negative rates were 12%, which was lower than in the other two studies (Moulin et al., 1994; Kimberley et al., 1994). In scenario four, prediction of normal hearing at 2000 Hz was only 82% and had a false negative rate of 19%. The prediction of very good hearing (0-10dB) was once again more satisfactory. Very good hearing could be accurately predicted at 2000 Hz 96% of the time in scenario five with a false negative rate of only 4%. If the objective is to identify very good hearing at 2000 Hz in the range of 0-10dB, this would be quite a sensitive and specific screening procedure.

It is interesting that normal hearing at 500 Hz and 1000 Hz could be predicted more accurately in this study than at 2000 Hz. Some studies that measured DPOAE amplitudes, reported a “trough” or a dip in DPOAE amplitude in the 2000 Hz region

(Gaskill & Brown, 1990; Gorga et al., 1993; Lonsbury-Martin et al., 1990; Spektor et al. 1991). Although the reason for this drop in DPOAE amplitude is still unknown, Lonsbury-Martin et al., (1990) suggested that it could be as a result of individual differences in middle and inner ear resonance. This study did not use DPOAE amplitude as an input stimulus, but the whole spectrum of present and absent responses. It could be argued however, that a decrease in DPOAE amplitude at 2000 Hz might influence the level where the distortion product vanishes into the noise floor. If this hypothesis is true, it could result in less present responses and more absent ones and that could account for the differences in prediction accuracy.

The comparison of the prediction abilities of scenario four and five for 2000 Hz indicates that scenario five predicted severe hearing loss much more accurately. For example, when the two categories that represents severe hearing loss in scenario four is combined, the result is a category that spans 51-70dB. This category overlaps with category five of scenario five (51-65dB) if one consider that only persons with pure tone thresholds > 65dB HL were selected for this study. The average accuracy of prediction of the two categories in scenario four was 10% and in scenario five it improved to 37%. The average prediction within one 10dB class in scenario four was 35% and it improved to 47% within one category in scenario five. Average prediction of the two classes of scenario four was more than one 10dB class wrong 56% of the time. Scenario five had only 16% completely wrong. The differences in category size between scenario four and five makes it difficult to make comparisons of prediction accuracy when predictions are one or more categories out, but the improvement of prediction accuracy within the same category can clearly be seen in scenario five. This improvement is probably also due to the fact that the neural network had more data to train on. Improvements in scenario five, however, was not enough to make accurate predictions regarding hearing status at 2000 Hz when a hearing loss is present.

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 (Gorga et al., 1993; Moulin et al., 1994). In the study by Gorga et al. (1993), normal hearing could be accurately predicted 90% of the time with a false alarm rate of only 5%. Moulin et al., (1994) successfully predicted the DPOAE frequency of 4000 Hz (primary frequencies are between 5000 Hz and 6000 Hz) as normal 79% 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%.

The prediction of normal hearing at 4000 Hz with the use of DPOAEs and neural networks revealed that normal hearing could be predicted accurately 91% of the time in scenario five. The false negative rate for this prediction was 10%. Very good hearing ability (0-10dB) at 4000 Hz was accurately predicted 92% of the time with a false negative rate of 7%.

Prediction of specific classes of hearing impairment at 4000 Hz was once more not satisfactory. Even category 2(11-20dB) in scenario four, depicting the upper half of normal hearing, was not predicted accurately once. Category 3(1-30dB) was predicted accurately only 13% of the time and category 4(31-40dB) was never accurately predicted.

The comparison of prediction accuracy for scenario four and five revealed significant improvement of the prediction of moderately severe hearing losses and overall prediction accuracy across all categories. In scenario five, the prediction of category 5(51-65dB), which represents the moderately severe hearing losses were accurately predicted 68% of the time, one category out 15% of the time and completely wrong 17% of the time. This is a great improvement for the prediction accuracy of moderately severe hearing losses in scenario four, where the average accurate prediction was 34%, the average one category out prediction was 38% and the average wrong prediction was 28%. The moderately severe hearing category had its best prediction at 4000 Hz, scenario five by far. One reason for this improvement in prediction accuracy was because the neural network had so many ears in that category to train on. Category 5 (51-65dB), scenario five had 41 ears and category 1 (0-10dB) had 47. This large number of ears in the moderately severe hearing loss category can be attributed to the general pattern of sensorineural hearing losses, often involving the higher frequencies to a greater extent (Yantis, 1994). The poor accurate prediction of the second category (11-20dB), correct only 14% of the time, can be attributed to the lack of data in that category. There were only seven ears that had a hearing loss of 11dB to 20dB at 4000 Hz and apparently it was not enough data for the neural network to train on. It could be speculated that prediction accuracy for all categories could be greatly improved if there were more subjects included in every one of the categories.

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 and possible reasons for inaccurate predictions. These results are discussed below.

6.8 Case Studies where the Audiogram was Predicted Accurately

The network predicted hearing ability into one of seven 10dB categories in scenario four or into one of five categories in scenario five. Examples of correct predictions are from scenario four and can be seen in Figure 6.1. Subjects that demonstrated very good hearing ability (0-10dB) at all four frequencies were usually predicted accurately. Figure 6.1. A is the audiogram and predicted categories for ear 35, an ear with normal hearing that was predicted accurately. Figure 6.1. B is the audiogram and prediction of ear 85, an ear with a hearing loss predicted accurately except for 4000 Hz that was 10dB out. The third example, Figure 6.1. C is from 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 XLIV.

Figure 6.1: Case studies where the audiogram was predicted accurately (prediction in green).

Table XLIV: 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

6.9 Case Studies where the Audiogram was Predicted Inaccurately

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

Table XLV: 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

6.9.1 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.

6.9.1.1 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 loss, 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 loss, such as in the cases of ears 73, 74 and 91.

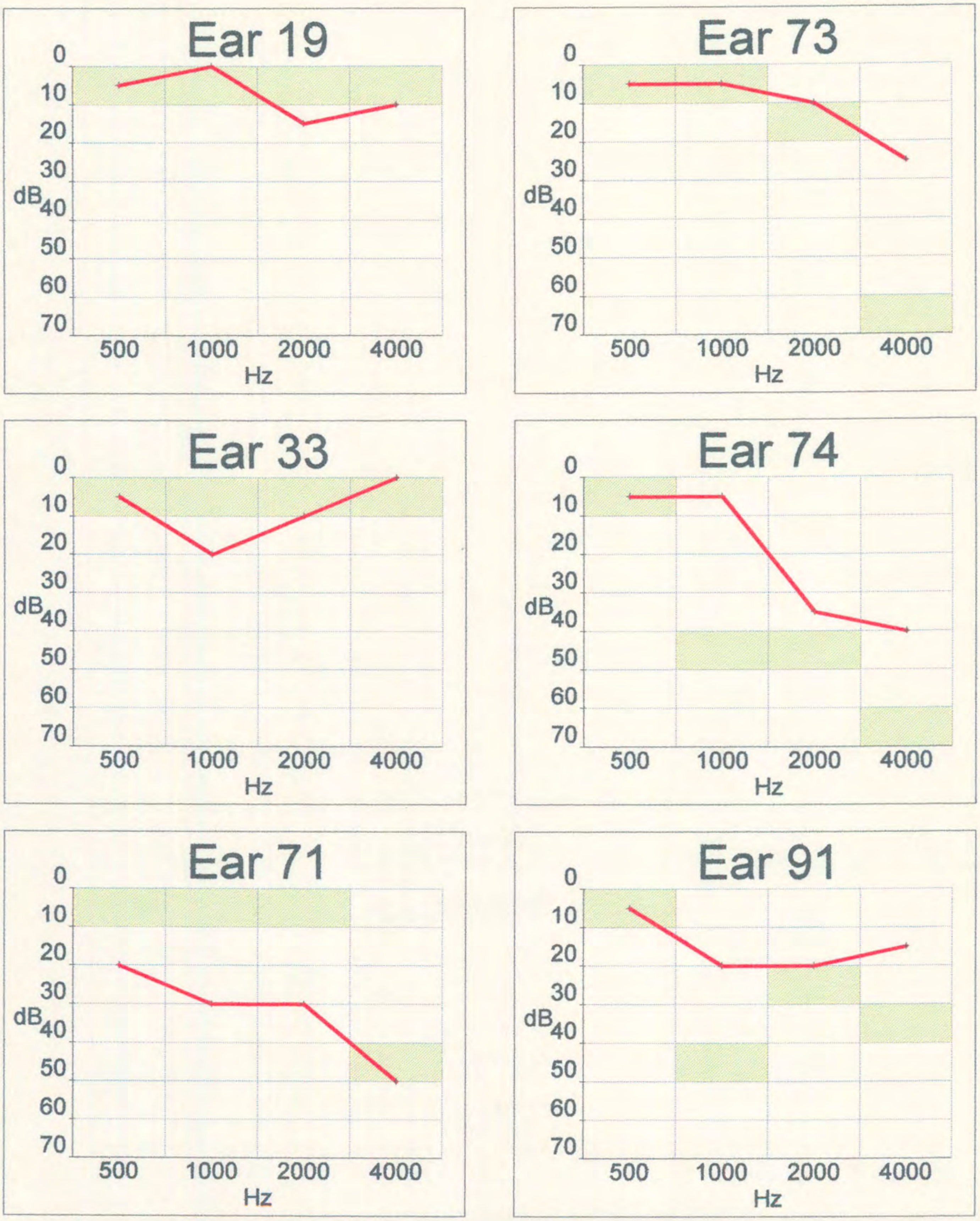


Figure 6.2: Case studies where the audiogram was predicted inaccurately (prediction in green).

6.9.1.2 Subjects Demonstrating Very Mild Hearing Loss

Subjects demonstrating very mild hearing loss, 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.

6.9.1.3 A Subject 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 percentage 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 6.2.). 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 variables that possible influenced the outcome of this study.

6.10 Variables that Influenced the Outcome of this Study

Numerous variables influenced the outcome of this study. The variables will be divided into variables of the distortion product otoacoustic emission, variables of the neural network, and variables of the subject.

6.10.1 Variables of the Distortion Product Otoacoustic Emission

Variables of the distortion product can be divided into technical parameters of DPOAE measurements and variables concerning the data analysis. Different aspects of the distortion product can be used to predict hearing threshold and these aspects are referred to as DPOAE analysis variables. These variables will be discussed in the following section.

6.10.1.1 Technical Parameters of DPOAE Measurements

In every research project that measures DPOAEs, there are a large number of stimulus variables that should be specified before results can be interpreted. Different frequencies, loudness levels, frequency ratios and loudness level ratios can influence the outcome of the results. The fact that every research project has a different technical setup makes it very difficult to compare the results. The stimulus parameters for this study were chosen with great care, based on the recommendations of many other researchers (Avan & Bonfils, 1990; Gaskill & Brown, 1990; Gorga et al., 1993; Harris et al., 1989; Kimberley et al., 1994; Mills, 1997; Moulin et al., 1994; Nielsen et al., 1993; Stover et al., 1996a).

The stimuli that were used for this study were specified in detail in Chapter 2. To summarize the stimulus parameters briefly, 11 pure tone frequencies spanning f1 = 500 Hz to f1 = 5031 Hz were used. The primary frequency ratio of f1/f2 was 1.2. Loudness levels for the primaries ranged from L2 = 65dB SPL to L2 = 25dB SPL. The loudness level ratio for L1/L2 was L1>L2 by 10dB.

6.10.1.2 DPOAE Analysis Variables

There are different aspects of the distortion product that can be used in the analysis of data. Information regarding the threshold or the amplitude of a distortion product can be used to attempt to predict hearing thresholds, or a combination of both.

There is not yet clear consensus on the best analysis procedure to identify normal and impaired ears (Stover et al., 1996a). Some authors used only I/O functions where DPOAE thresholds were determined (Gorga et al.,1996) where others used only DP Grams or DPOAE amplitudes to investigate DPOAE responses in normal hearing and hearing impaired subjects (Gorga et al., 1993).

Most researchers, however, use a combination of the two procedures or perform both procedures separately. Martin et al. (1990a), Spektor et al. (1991), and Smurzynski, et al. (1990), performed both procedures in their studies separately, while Moulin et al. (1994), and Kimberley and Nelson (1989), combined the two procedures in an interesting way. Moulin et al., (1994) conducted several DP Grams but at different loudness levels in 10 dB steps, therefore gaining almost the same information as performing both procedures. Kimberley and Nelson (1989) on the other hand, measured several I/O functions but at eight different frequencies.

Not knowing which procedure is currently the most applicable in the areas of diagnostic effectivity, it seems plausible to gain as much information as possible by combining the two procedures or performing both separately (Kimberley & Nelson, 1989; Martin, et al., 1990a; Smurzynski et al., 1990). Thus far, researchers used amplitude or threshold measurements to predict hearing ability. Even though these authors combined these procedures, they used only the lowest measurable responses and not every measured response, as in the case of this study. This study did not analyze the amplitude or the threshold of the distortion product, but the pattern of all present and absent responses from eight DP Grams. The mere presence of a DPOAE and its position in the pattern of emissions served as input information for the neural network.

The second aspect of DPOAEs that has been used for data analysis in the past is the frequency of the emission that should be correlated with the frequency of the distortion product. In other words, does a DPOAE indicate the state of the OHC on the basilar membrane in the region of the f1 frequency, the f2 frequency, the geometric mean of the two primary frequencies (GM) or at the distortion product itself (the 2f1-f2 frequency). Many authors also disagree on which frequency variable of the DPOAE should be compared to the pure tone threshold (PTT). Some authors suggest that it is the f2 value of the DPOAE that best correlate with pure tone thresholds (PTTs) (Harris et al., 1989; Kimberley et al., 1994). Other studies support the notion that the generation of the distortion product correlates best with the cochlear place near the geometric mean (GM) of the primaries (Bonfils et al., 1991; Harris et al., 1989; Kimberley et al., 1994; Lonsbury-Martin & Martin, 1990; Martin et al., 1990b; Moulin et al., 1994; Smurzynski et al., 1990).

For this study, the pattern of present and absent responses of all 11 DPOAE frequencies was used, and not only DPOAE information at one frequency, such as the f2 place or the geometric mean or the 2f1-f2 place. This aspect is possibly the reason why 500 Hz could be predicted so accurately. Other researchers attempting to predict normal hearing at 500 Hz, used DPOAE information around 500 Hz and experienced a lot of background noise and absent responses. In this study, the neural network could gain enough information from surrounding frequencies to predict the status of 500 Hz accurately 92% of the time.

The next set of variables that influenced this study, are variables of the neural network.

6.10.2 Variables of the Neural Network

Variables of the neural network can be divided in neural network topology and the amount of data available to train on. These variables are discussed below.

6.10.2.1 Neural Network Topology

Neural network topology can be divided into the size of the hidden layer and the accuracy of the prediction during training.

6.10.2.1.1 The Size of the Hidden Layer

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 extract significant features of the input data (Nelson & Illingworth, 1991). With too many middle neurons the network has difficulty to make generalizations (Nelson & Illingworth, 1991; Rao & Rao, 1995). 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 100, 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, a number of 140 hidden layer neurons were found to yield optimal results.

6.10.2.1.2 Accuracy of the Prediction During Training

Another neural network variable, is the acceptable error during training. As stated in Chapter 3, 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 5dB, 10db, 1dB or within any decibel amount during the training stage. The normal assumption would be that 1dB accuracy during training would yield the most accurate predictions. It was actually found in this study that if the neural network were trained with 1dB accuracy during training, the network had limited generalization abilities and made poorer predictions when presented with unfamiliar data. Blum (1992) also stated that it sometimes help to train on slightly “noisy” data to enhance generalization ability during the prediction of unfamiliar data.

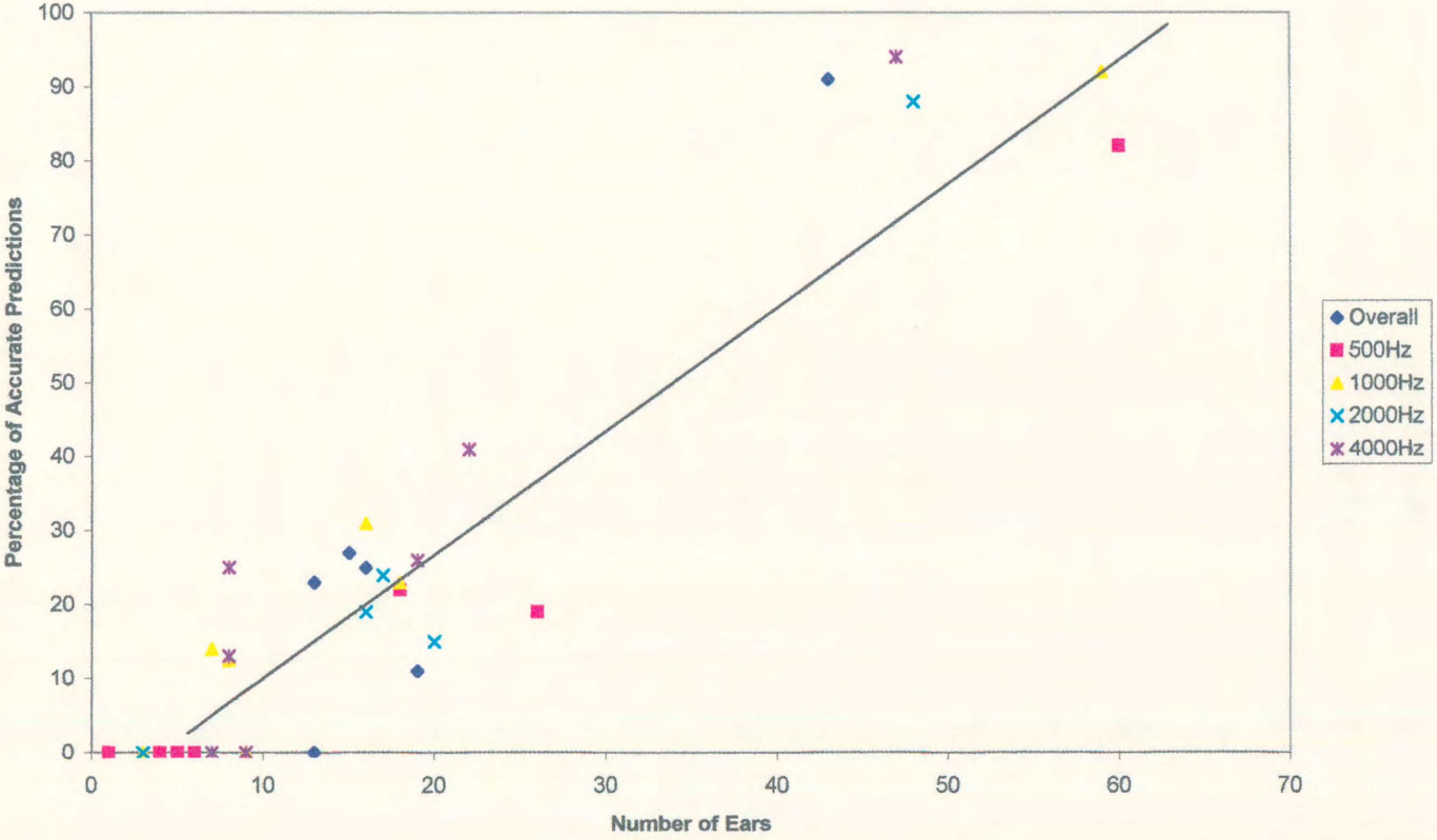
By setting the acceptable error during the training period to 5dB, most accurate predictions of unfamiliar data was obtained in this study.

6.10.2.2 Amount of Data Available to Train on in Every Category

This variable influenced the prediction accuracy of the neural network considerably. It was very evident that categories with large numbers of ears were predicted more accurately than categories with only a few ears. The prediction of 4000 Hz, scenario five for example, revealed 68% prediction accuracy for moderately severe hearing losses, due to the large number of ears in that category to train on (41 ears) but only 14% prediction accuracy for category 2 (11-20dB) with only seven ears. To investigate the effect that this variable had on prediction accuracy, the number of ears in every category were correlated with the percentage of accurate predictions. The prediction accuracy versus number of ears in every category for scenario four is illustrated in Figure 6.3. Scenario five is illustrated in Figure 6.4. The scattergrams in Figure 6.3 and Figure 6.4 indicate a very strong correlation between the number of ears in every category and the accuracy of the prediction. The linear fit shown with the data points in Figure 6.3 has a slope of 1.703 and Figure 6.4 has a slope of 1.714. The correlation coefficient for Figure 6.3 (scenario four) is 0.94 and for Figure 6.4 (scenario five) is 0.92.

Two conclusions can be derived from these scattergrams. First, there is a very strong correlation between the number of ears in every category that the neural network had to train on, and the accuracy of the prediction that was made. The more data the network had to train on, the more accurate the predictions. Second, derived

Figure 6.3: Prediction Accuracy and Ear Count Correlation
- Scenario 4 -



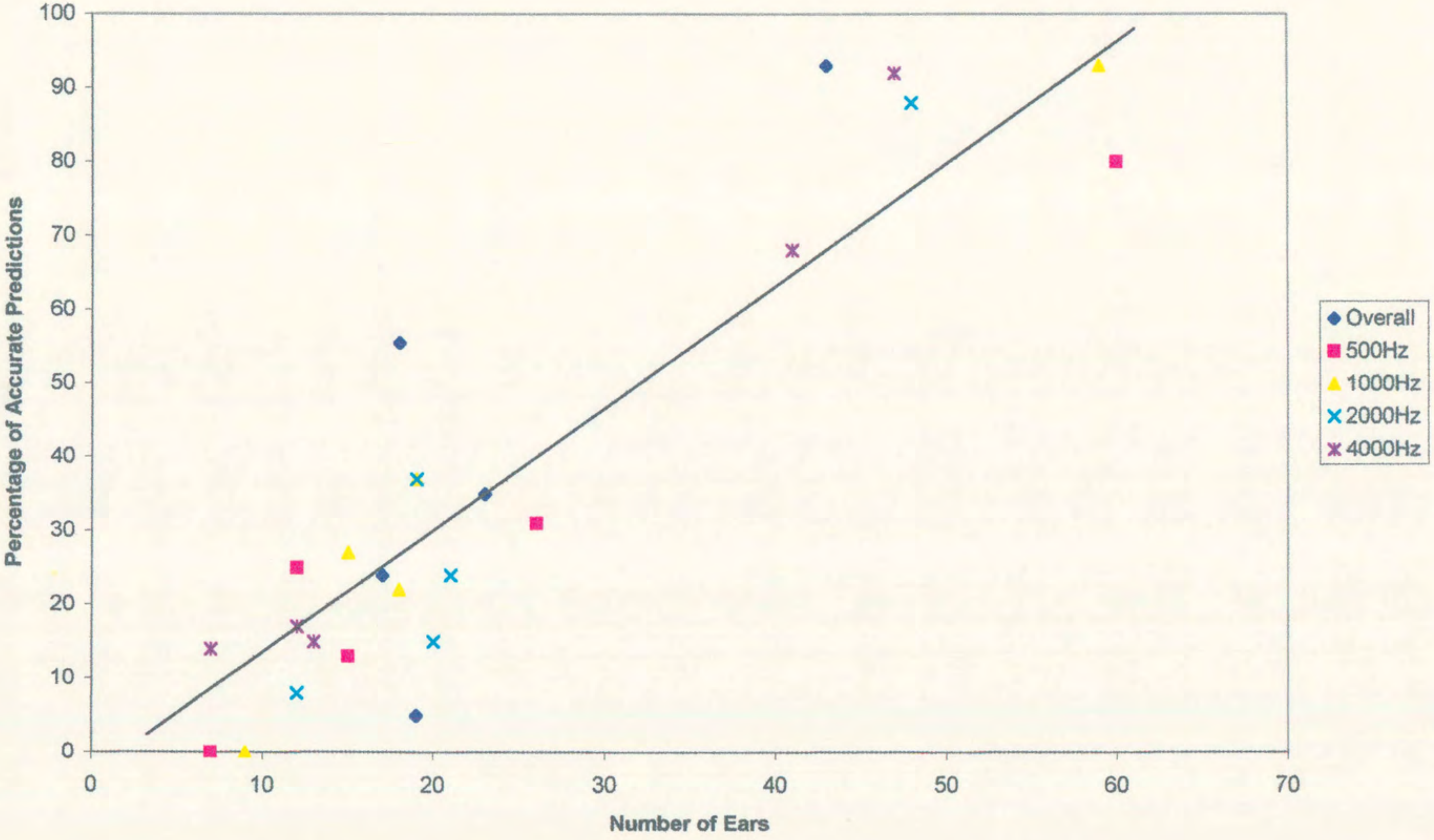


Figure 6.4: Prediction Accuracy and Ear Count Correlation
- Scenario 5 -

from the linear fit of the data points, one can expect to find a saturation level in prediction accuracy when the number of ears reach more or less 65. For example, a study with 1000 ears in every 10dB category would not yield significantly better predictions than a study with 70 ears in every 10dB category.

This strong correlation between number of ears to train on and prediction accuracy indicates a very good possibility that pure tone thresholds can be predicted in normal hearing and hearing-impaired ears with DPOAEs and neural networks. If more ears are included in the training of a neural network, accurate predictions of all categories of hearing impairment seems possible. To predict hearing ability within 10dB in normal hearing and hearing-impaired ears objectively, non-invasively and accurately in the frequency range of 500Hz to 4000 Hz (and even in higher frequencies), would improve diagnostic hearing assessment of difficult-to-test populations greatly. Such a test would change and improve the field of pediatric audiology, as we know it, dramatically.

Other variables that influenced the outcome of this study will be discussed next.

6.10.3 Subject Variables

Subject variables that influenced the outcome of this study include age, gender and the combination of age and gender. Another subject variable is the presence of a spontaneous otoacoustic emission close to a DPOAE frequency. These variables are discussed in the following section.

6.10.3.1 The Age Variable

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, when the age variable was included as a neural network input, prediction of average hearing ability improved from 40% overall accuracy to 50% overall accuracy. It is quite remarkable that one more input variable could provide enough information to the neural network to enhance prediction abilities so drastically. One can not help but wonder if the negative correlation between age and DPOAEs is really just a reflection of sensitivity changes related to aging, or if there is perhaps more to the influence of the age variable on the distortion product. Based on the improvement in prediction ability when the age variable is included in the neural network, it is suggested that future neural network runs should include this variable in the prediction of hearing ability at different frequencies.

6.10.3.2 The Gender Variable

The gender variable did not have a remarkable effect on the prediction accuracy of the neural network. Prediction accuracy improved from 40% overall prediction accuracy to 44% overall prediction accuracy. Some of the categories showed no improvement and prediction accuracy in some of the categories was even affected negatively. These results confirm the studies of Lonsbury-Martin et al. (1990), Gaskill and Brown (1990) and Cacace et al. (1996) that gender effects on DPOAEs are apparently limited to minor differences in DPOAE amplitudes and thresholds. It is not recommended that the gender variable should be included in the neural network's input alone, but it could be included as a variable in conjunction with the age variable.

6.10.3.3 The Combination of the Age and Gender Variables

The combination of the age variable and gender had an interesting effect on the prediction accuracy of the neural network for average hearing ability. Even though overall prediction accuracy for the neural network was slightly poorer than with the age alone, (48% with age and gender, 40% without any extra variables, 50% with age alone, 44% with gender alone, see Table XLII), the prediction accuracy of very good hearing (0-10dB) improved to 98% (Table XLIII). Very good average hearing (0-10dB) was predicted as normal (0-20dB) 100% of the time. The age variable alone predicted all categories, except the first one, better than the combination of age and gender did. It is therefore recommended that for the development of DPOAEs and neural networks for screening purposes, the age and gender variables should both be included, to ensure optimal prediction of the normal hearing ability categories. For prediction of hearing ability for diagnostic purposes however, (the prediction of specific categories of hearing ability) it seems that the age variable should be included alone as neural network input.

The last variable that could possibly have affected the outcome of this study is the presence of a spontaneous emission close to the distortion product.

6.10.3.4 The Presence of a Spontaneous Otoacoustic Emission Close to the Distortion Product

It was mentioned in Chapter 2 that a spontaneous otoacoustic emission within 50Hz of the primaries of a distortion product otoacoustic emission could enhance the amplitude of a DPOAE significantly under certain experimental conditions (Kulawiec & Orlando, 1995; Probst & Hauser, 1990). It could be argued that SOAEs were never measured in the data collection procedure and that it could have influenced the outcome of this study.

The presence of SOAEs close to the primaries did not influence this study significantly for the following reasons. First, SOAEs are recordable only in 50% of subjects with normal hearing and is completely absent in subjects demonstrating a hearing loss of more than 30dB HL (Lonsbury-Martin, 1994). Even if some of the subjects with normal hearing demonstrated spontaneous emissions close to the primary frequencies of the distortion product, the neural network would still have predicted their enhanced DPOAE amplitudes as normal. The only possible difference in the prediction accuracy of the neural network if a normal hearing subject demonstrated a SOAE, is that a category two (11-20dB) could possibly be predicted as a category one (0-10dB). It would however not have any effect on the prediction of normal hearing (0-20dB) as normal. Second, SOAEs are only present in normal hearing subjects and could therefore not have affected the neural network's prediction of hearing impaired categories. Studies investigating the normal characteristics of the distortion product should however be aware of this phenomenon and measure SOAEs in their data collection procedures.

6.11 DPOAE Measurements as a Diagnostic or Hearing Screening Procedure

In Chapter 1, current objective physiologic hearing assessment procedures were discussed as well as their shortcomings in the assessment of difficult-to-test populations. ABR is currently the most popular and widely used objective test of hearing for special populations (Weber, 1994), but ABR has a series of shortcomings limiting its effectiveness as a diagnostic procedure. Some of these limitations include that ABR measurement is often costly, it provides only limited frequency information, requires a large amount of time and highly trained and specialized personnel, and may require sedation.

The need for an objective, rapid, non-invasive, inexpensive, and accurate measurement of hearing was identified in Chapter 1. In the following chapter, one type of emission, the distortion product otoacoustic emission (DPOAE) was investigated as a possible new test of hearing and Chapter 2 concluded that DPOAEs have the necessary characteristics to be developed as a new objective test of hearing.

Otoacoustic emissions have been applied for a number of clinical applications 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).

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, amplitude and threshold information 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 yield ear specific information. Many researchers used these attributes to correctly identify normal hearing in populations with varying degrees of hearing ability successfully (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 sometimes prevalent 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).

Otoacoustic emissions however, have never been used as a diagnostic test of hearing where specific thresholds for frequencies were determined (Kimberley et al., 1994; 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) and described the possibility of DPOAEs as a diagnostic audiologic test (Durrant, 1992; Kimberley et al., 1994; 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.

The aim of this research project was, as stated in Chapter 4, 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 10-15dB 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, 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. Results indicated that these two aspects are strongly correlated. The correlation coefficient was 0.94 for scenario four and 0.92 for scenario five. Every prediction was presented by a point on a scattergram, indicating the number of ears in a category and the accuracy of that prediction. These results confirm the speculation 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. The research in this study should be viewed as a stepping stone for further research to develop the distortion product as a new objective diagnostic test of hearing.

6.12 The Effectiveness of the Application of Neural Networks to the Field of Audiology

Artificial neural networks (ANNs) have been applied to the field of Speech Pathology quite effectively in the past. One such an example is the study by Metz et al., 1992, who used artificial neural networks to estimate speech intelligibility of hearing-impaired subjects from acoustic variables. These authors were also confronted by complex data sets with potential nonlinear relationships between acoustic speech parameters and speech intelligibility. Metz et al. (1992) also used a back propagation neural network, but with two hidden layers that had between 4- 20 hidden layer neurons. The neural network was capable of dealing with the systematic nonlinearities in their complex data sets. The network very successfully classified hearing impaired persons into the first and last group (most and least intelligible) but the neural network experienced difficulty classifying intermediate categories probably due to the criterion chosen to separate the different classes. This experiment is currently being expanded to improve network performance.

The application ANNs to the field of audiology in this study revealed promising results. For the first time, normal hearing (PTTs < 20dB HL) could be predicted at 500 Hz with DPOAEs with 92% accuracy. This has never been possible with conventional statistical methods. The prediction of normal hearing at 1000 Hz was also improved considerably, from 73% accuracy (Moulin et al. 1994) to 87% accuracy in this study with ANNs. Results obtained at higher frequencies in this study were similar or slightly poorer than the predictions of normal hearing at the higher frequencies in other studies.

Another aspect that the results in this study revealed, it that is there is a very good possibility that pure tone thresholds can be accurately predicted (within 10dB) with artificial neural networks, given that there is enough subject data for the neural net to train on.

The application of neural networks to the field of audiology was therefore very successful. The neural network extracted significant information from the input data and identified a correlation between pure tone thresholds and DPOAEs. The network then used this learned correlation to make predictions of pure tone thresholds effectively. If all categories of hearing ability had sufficient data for the neural network to train on, hearing ability would have been predicted accurately in all categories. Even with a shortage of data in some of the categories, the neural network was still able to distinguish normal from impaired hearing ability effectively and accurately.

6.13 Summary

In the investigation of DPOAEs as a possible new hearing screening or diagnostic procedure, many authors used conventional statistical methods to find a correlation 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). Some researchers used statistics to predict 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).

The aim of this research project was 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.

Even though this study could correctly identify normal hearing quite accurately, even at 500 Hz, 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 limited number of ears in every category that the neural network had to train on. With closer analysis of the 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 recommendation for future research in hearing prediction with artificial neural networks and DPOAEs.

7 Summary, Evaluation of the Study and Conclusion

7.1 Summary

Ever since Kemp (1978) first described otoacoustic emissions, there has been an interest in these measurements to develop another diagnostic tool to evaluate hearing ability objectively, non-invasively and accurately. An overview of current objective diagnostic procedures revealed that many technological advanced procedures 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, a few weaknesses in these procedures 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 possibility of sedation and the level of expertise and expense required (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 testing 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 most frequency-specific emission type, due to the frequency specificity of the stimuli that can be chosen to stimulate any specific region on the basilar membrane (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 only slightly influenced by aspects such as age and 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., 1994; 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., (1994), 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.

For this study, a mathematical model, called artificial neural networks, 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 possible non-linear correlation. The neural network was used to predict pure tone thresholds given only the distortion product responses.

Artificial neural networks were initially developed to gain a better understanding of how the human brain works (Nelson & Illingworth, 1991). It is an algorithm for a cognitive task, such as learning or pattern recognition (Muller & Reinhardt, 1990). 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). It was with great expectations that neural networks were applied to the field of Audiology in this study.

The rationale for this study was to investigate DPOAEs as a possible new objective test of hearing. The research design of this study was a multivariable correlational study. The correlation between selected variables of DPOAEs and PTTs were studied with artificial neural networks. The correlation found by the neural network was then used to predict hearing ability at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz with DPOAEs (the main aim of the study).

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 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 traditional audiogram.

Data collection procedures consisted of the specification of stimulus parameters and DPOAE testing procedure. 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 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 threshold was that the distortion product had to be at least 3 dB above the noise floor and accepted by the GSI 60 DPOAE system during measurement.

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 (see Table II for the eleven frequency pairs). The eleven frequency pairs were presented to the ear in a sweep, one at a time starting with the low frequencies, ending with the high frequencies.

A data file for each ear was created, consisting of 19 columns and 88 rows of numbers (an example of a data file for one DP Gram can be seen in Table III).

A back propagation network was chosen for this study for two reasons: **1)** A possible nonlinear correlation is suspected between DPOAE thresholds and traditional pure tone thresholds. Metz, et al. (1992) reported the back propagation neural network to be very successful in dealing with nonlinearities that potentially occur in complex data sets. According to Blum (1992), the back propagation 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 back propagation 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 served as input stimuli. Hearing ability was divided into categories and the neural network had to predict hearing ability into one of the categories. Scenario four had seven 10dB categories and scenario five had five categories, spanning 10-15dB each. The network had 140 middle neurons, 88 input nodes and seven output neurons in scenario four, five output neurons in scenario five. The network's acceptable prediction error during training was set at 5%.

As a first level approach, average hearing ability (average of 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz) was predicted first, and hearing ability at the four frequencies thereafter. The variables age and gender were included in neural network runs where average hearing ability was predicted to determine the effect of these variables on the distortion product.

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, 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 92% of the time. Normal hearing at 1000 Hz was correctly identified 87% of the time, at 2000 Hz 84% of the time and at 4000 Hz 91% of the time. Another aspect that should be kept in mind, however, is the false negative rate of every test that evaluated auditory functioning. Even though 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), the high incidence false negative responses raises questions regarding the sensitivity of this procedure.

The good predictions of normal hearing at the four frequencies can be attributed to two reasons. First, the different data analysis technique, or DPOAE variable that was used as input data, namely the use of all present and absent responses and not only the amplitude or threshold of one DPOAE measurement. Second, the different data processing technique, artificial neural networks, which excelled in the finding of a correlation between these two complex data sets.

The age variable had a positive effect on the prediction accuracy of average hearing ability. When the age variable was included as a neural network input, prediction of average hearing ability improved from 40% overall accuracy to 50% overall accuracy. Based on this improvement, it is suggested that future neural network runs should include this variable in the prediction of hearing ability at different frequencies. The gender variable did not have a remarkable effect on the prediction accuracy of the neural network. Prediction accuracy improved from 40% overall prediction accuracy to 44% overall prediction accuracy. The combination of the age variable and gender improved the prediction of normal hearing ability considerably.

The purpose of this study was to predict pure tone thresholds for 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz. Hearing ability was divided into categories, and the neural network had to predict the most probable category of hearing ability. The frequencies were predicted in two scenarios, scenario four, with seven 10dB categories and scenario five, with five categories of different decibel intervals. 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. A correlation between prediction accuracy and the number of ears in every category revealed that these two aspects are correlated strongly (correlation coefficient for scenario four was 0.94 and scenario five 0.92). All the categories and their predictions were plotted on a scattergram and the linear fit of the data plots suggests that the accuracy of the prediction increases as the number of ears in every category increases (linear fit of 1.703 for scenario four and 1.714 for scenario five). It can therefore be speculated that hearing ability can be predicted accurately (within 10dB) at various frequencies (500 Hz to 4000 Hz and possible even higher) over a range of 0 – 65dB HL with artificial neural networks. This can be achieved when every category is well represented and the network has enough data to train on, and also if the stimulus parameters for DPOAEs are chosen carefully.

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

The application of neural networks to the field of audiology was therefore very successful. The neural network extracted significant information from the input data and identified a correlation between pure tone thresholds and DPOAEs. The network then used this learned correlation to make predictions of pure tone thresholds effectively. If all categories of hearing ability had sufficient data for the neural network to train on, hearing ability would have been predicted accurately in all categories. Even with a shortage of data in some of the categories, the neural network was still able to distinguish normal from impaired hearing ability effectively and accurately.

The results in this study indicate strongly that DPOAEs are suitable as a diagnostic audiologic test of hearing. The research in this study should be viewed as a stepping stone for further research to develop the distortion product as a new objective diagnostic test of hearing.

7.2 Evaluation of Research Methodology

In the evaluation of research methodology, so many factors are at play that could influence the results in this study. Chapter 2 discussed the variables of the distortion product in detail and Chapter 3 the variables of artificial neural networks. Chapter 4 discussed reasons why certain variables and stimulus parameters were deemed as crucial for this type of study and chosen for this research project. The evaluation of research methodology will briefly look at 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. Certain chosen variables of the distortion product were used as input information in an artificial neural network. The network was trained with DPOAE variables and selected variables of pure tone thresholds. During training, the network determined a correlation between these two data sets. After completion of training, the network used the found correlation to predict pure tone thresholds of an unfamiliar subject with DPOAE results alone. 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 the data obtained. 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 avoid this factor from influencing 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, in April 1997. 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 in January 1998. 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.

Lastly, the accuracy of the prediction was determined by how accurately the neural network was trained. The training accuracy of the neural network was measured after every training set, and the neural network was trained until the accuracy of the prediction in the training set was at least 95%.

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 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 audiograms 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, 1995; Teleschi, Widick, Lonsbury-Martin & McCoy, 1995).

Neural networks are also very reliable. Neural networks are completely deterministic, in other words, 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 also based on finds during the pilot study. The third aspect that influences measurement and analysis according to Hall III et al. (1993), is the signal-to-noise ratio and the criteria for when a DPOAE is present. Based on recommendations in literature, the presence of a DPOAE (DPOAE threshold) was taken as 3dB above the noise floor for this study (Lonsbury-Martin, 1994; Lonsbury-Martin et al., 1990). The last factor identified by Hall III et al. (1993) that could influence data analysis was subject variables such as age and gender. Both these variables were included in network runs to determine their effect on the distortion product.

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 Microsoft 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 Microsoft Excel.

According to Leedy (1993) validity looks at 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 and to use that correlation to predict PTTs with DPOAEs and neural networks. 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 can not 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 that it has not yet been determined what the acceptable percentage error should be that was chosen during the training of the neural network. Even though prediction and generalization abilities of the neural network were optimal in this study with a percentage error of 5%, this percentage is not necessarily acceptable for the clinical application of DPOAEs as a diagnostic procedure. It is possible that 5% error during training is not accurate enough for a diagnostic procedure. This aspect requires further investigation.

The fourth limitation is the fact that this study did not use the amplitude of the DPOAE in conjunction with its presence (in this study, the pattern of present and absent responses were used, depicted as a one or a zero). In some of the initial neural network runs, it was attempted to include amplitude and threshold data, but due to inability of the neural network to converge with all these extra inputs, the inputs were simplified by presenting it to the neural network in a binary fashion. By not using amplitude information, it is possible that a whole dimension of the data was lost, that could have enabled the network to make much more accurate predictions.

The fifth limitation is the high incidence of false negative responses recorded in this study. This high incidence of false negatives influence 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.

The last limitation identified in this study is the duration of 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

The first recommendation for a study attempting to predict PTTs with DPOAEs and ANNs is to increase the number of subjects. According to Figure 6.3 and Figure 6.4, prediction accuracy of the neural network will be most accurate if there is at least 65 ears in every category.

The second recommendation is to include the age variable as one of the inputs of the neural network when hearing ability at specific frequencies is predicted. The age variable influenced the overall prediction accuracy of the neural network considerably.

The third recommendation is to attempt to find a network configuration that would allow the researcher to include amplitude information of the distortion product. By including this dimension of the distortion product as well, the network might be able to make much more accurate predictions of hearing ability.

The last 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.

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 distortion product otoacoustic emissions 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 the 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.