FEATURE EXTRACTION IN AUDITORY BRAINSTEM RESPONSES: USE OF THE DAUBECHIES WAVELET

G. Lightbody, P. J. McCullagh and H. G. McAllister

School of Computing and Mathematics, Ulster of University, Newtownabbey, Northern Ireland

g.lightbody@ulster.ac.uk

Abstract: The Auditory Brainstem Response (ABR) is a clinical trial whereby a stimulus click is applied to a subject's ear to determine hearing capability and health of the auditory pathways. If the stimulus is heard a response should be noted in their EEG. A significant body of work has been devoted to the enhancement and automated classification of these responses. This paper presents a methodology whereby wavelet decomposition is performed on the pre and post-stimulus sections of the ABR waveform. Wavelet coefficients best representing the ABR are selected and a power ratio using the post over pre-stimulus coefficients is used as a response indicator, classifying strong visible responses with an accuracy of 99.5%. Those left unclassified are passed to a further stage whereby repeat recordings are used to calculate cross-correlation features derived from the wavelet and time domains. These formed inputs to a C5.0 decision tree algorithm resulting in an accuracy for the lower level responses of 80%.

Introduction

Wavelets form a powerful tool for multi-resolution filtering and analysis. They decompose the signal into frequency bands that maintain a level of temporal information [1]. One such area that benefits from wavelet filter theory is in the pre-processing and feature extraction of Auditory Brainstem Responses (ABR) [2]. The conventional method of extracting the ABR waveform has been averaging which uses the deterministic nature of the signal to enhance the waveform while suppressing the background noise and EEG. Typically up to 2000 single trials may need to be averaged before the noise is sufficiently suppressed.

Figure 1 shows a very strong visible response for a healthy adult with a 70dB (normal hearing level, nHL) stimulus. This is referred to as the Jewett waveform and is characterised by seven peaks, of which the key five are labelled on Figure 1. As the level of stimulus is reduced, the different peaks of the waveform become less obvious, and their latency is increased. The shape of the waveform from the long slope at the top of peak V is the strongest part of the waveform to remain as the stimulus diminishes. This is depicted in Figure 2. It can be seen that the identification of wave V and its following negative slope are key to classifying the presence of a response in a less obvious case, when the stimulus level is set near the subject's hearing threshold.

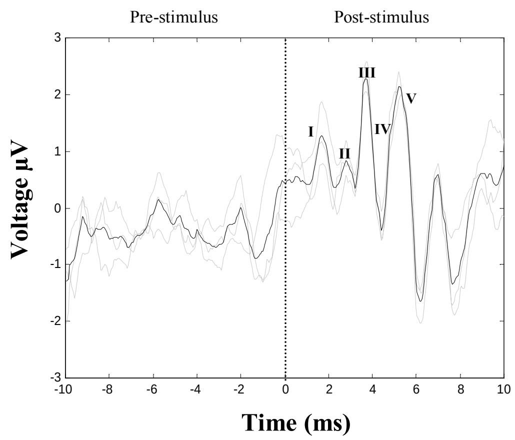

Figure 1: Jewett ABR waveform at a 70dB Stimulus

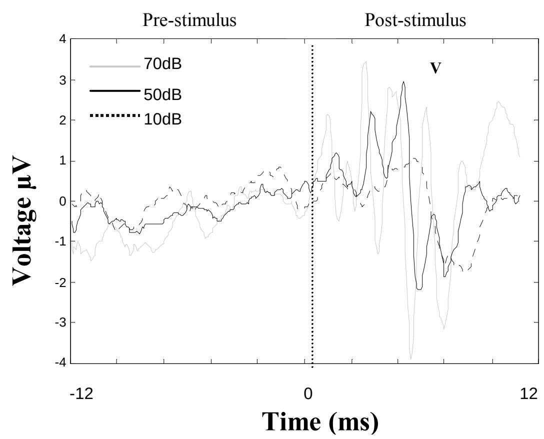

Figure 2: ABR waveform at a 70, 50 and 10 dB Stimuli

The ABR waveform will differ from subject to subject and may be affected by external conditions such as, electrode placement, filtering ranges, intensity levels, ear used, and even patient head shape. A range of factors need to be taken into consideration before a clinical expert can make an interpretation. They may need to:

- Check latencies of waves I, III and V
- Examine overall morphology of waveform
- Evaluate consistency of subaverages

The interpretation of the waveforms can be subjective, thus clinical experts may not always draw the same conclusion [3], which is particularly true in threshold cases. Artificial intelligence techniques can provide an objective assistance in response

ISSN: 1727-1983 © 2005 IFMBE

interpretation [4,5]. Useful information may be extracted from the EEG records using analysis in the time and frequency domains. Davey [5] showed that power analysis of the post and pre-stimulus time domain waveforms could be used as an effective method to classify strong responses with an accuracy of 98.6%. A ratio was calculated from the powers in the post over pre-stimulus waveforms. Ratios above a set threshold of 5 indicated a response. Remaining ABRs were passed to a second stage of analysis, whereby repeat recordings were used to derive features based on cross-correlation parameters using both the time and frequency domains. These features provided inputs to train decision trees and neural networks to classify the responses, resulting in an accuracy of 82.5%.

In this paper an analysis using wavelet decomposition has been investigated using a subset of test cases from the same dataset as Davey [5]. Although, the exact dataset was not known, therefore, the time domain analysis was repeated to guarantee that a valid comparison could be made. A similar structure was followed using a two-stage process. Power ratios of the post over pre-stimulus wavelet coefficients were used to classify strong responses. Those remaining were then classified using correlation features of repeated tests derived from the wavelet domain.

Materials and Methods

The investigations were performed on a database of 89 test subjects, provided by the Audiology Department of the Royal Group of Hospitals in Northern Ireland. Each subject had a range of tests stimulus levels applied providing an ample mix of good, weak and non-response waveforms, all of which were classified by a clinical expert. The data was pre-processed by band-pass filtering (100Hz-3kHz) and then sampled at 20kHz. Each waveform consisted of 480 data samples, half before the stimulus and half after, which related to 12ms before the stimulus and 12ms after the stimulus.

Figure 3 depicts how the data was pre-processed before commencing the wavelet analysis. The first stage was to de-noise the data using a wavelet filter. The post and pre-stimulus components of the waveform were separated, and then decomposed individually using a Daubechies scaling wavelet [6]. Davey [5] showed that by selecting only a portion of the post-stimulus waveform the accuracy of classification could be enhanced. Three possibilities were considered [5]. Firstly, all the post stimulus data was considered. Then, only waveforms I to V were considered, relating to data from 1.5ms post-stimulus to 9.5ms post-stimulus. Lastly, wave V was isolated and used, which related to data from 5ms post-stimulus to 9.5ms post-stimulus. That is, the most signification oscillations of the ABR were maintained, while eliminating certain artifacts that were observed to occur at time of stimulus and at the end of the recording. The same procedure was followed for the wavelet analysis, as shown in Figure 3.

The wavelet decomposition required that the number of data samples be dyadic, therefore the samples needed to be extended to 256. Three methods [7] were considered for these experiments:

- **Zero Extension:** the data is extended by zeros
- **Odd Extension**: the data begins to repeat from the start of the waveform
- **Even Extension**: the data is extended by a reflection of the waveform, i.e. sample N+1 is sample N, sample N+2 is sample N-1, etc.

A comparison was made and it showed that using even extension provided moderately better results, as shown later on in Table 4.

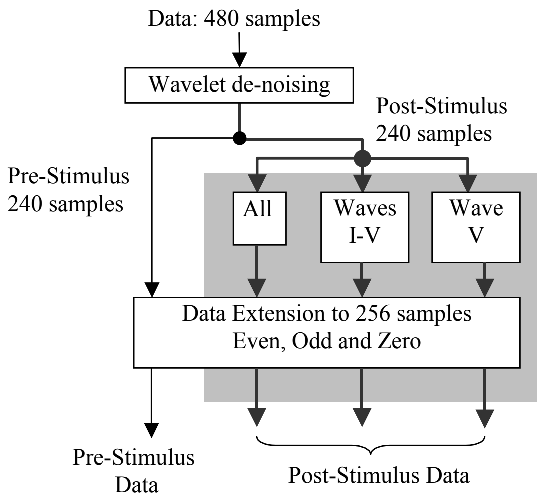

Figure 3: Pre-processing of ABR waveform

The decomposition was performed on the 256 input samples of data to seven levels, referred to as approximation level, A7, and detail levels, D7-D1. Each level produced coefficients that represented a frequency band of the test data. In total there were 256 coefficients produced as shown in Table 1. The higher the level, the lower the frequency band, and the lower the resolution (i.e. fewer coefficients). This enabled the separation of coefficients in levels D4, D5 and D6, which represent the frequency bands that dominate the ABR, that is, frequencies in the region of 200Hz, 500Hz and 900Hz [8]. Likewise, coefficients D1-D3 that represented the higher frequency components were excluded from the tests.

Table 1: Wavelet Decomposition Levels

Level	Coefficients	Number of Coefficients
D1	[129:256]	128
D2	[65:128]	64
D3	[33:64]	32
D4	**[17:32]**	**16**
D5	**[9:16]**	**8**
D6	**[5:8]**	**4**
D7	[3:4]	2
A7	[1:2]	2

This research concentrated on the coefficients of band D4, although D5 and D6 were also considered. It

ISSN: 1727-1983

was observed within D4 that there were obvious oscillations present in the post-stimulus coefficient set that were not present in the pre-stimulus case. Figure 4 shows the plotted D4 coefficients from the post-stimulus and pre-stimulus wavelet decomposition for an example showing a strong response to a 70dB stimulus. Figure 5 shows the same coefficients for an example where no stimulus was applied, using the same subject.

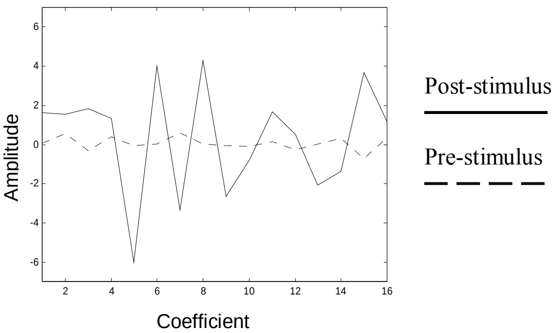

Figure 4: Key wavelet coefficients (70dB stimulus)

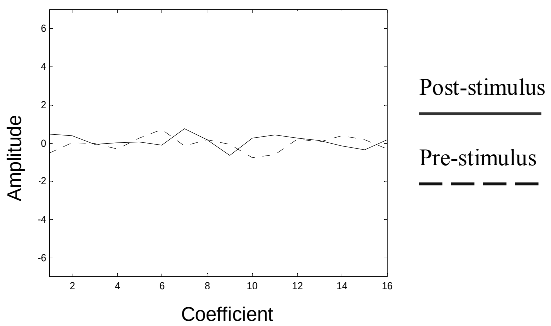

Figure 5: Key wavelet coefficients (No stimulus)

It was considered that the ratio of the mean absolute values of the post-stimulus coefficients over the mean absolute values of the pre-stimulus coefficients would provide a useful response indicator. In the example given in Figure 4 this value was 82.6 for a 70dB stimulus, compared with a result of 0.83 for data from the same subject but with no stimulus applied. In summary, the closer the parameter is to unity the less likelihood of a response. Conversely, the greater the value from unity indicates a stronger response.

There are latency changes in ABR as stimulus response diminishes (Figure 2), so it was supposed that allowing the wavelet to take different morphologies could provide useful information. That is, as the length of the wavelet scaling filter increases the latency of the oscillations change, so the conjecture was that different filter lengths would suit responses of different strengths and characteristics. Hence, a range of scaling wavelets was used from filter length 2 to 24 (even lengths only). Figure 6 shows the process followed to extract the coefficients. For each scaling filter the ratio of the mean absolute D4 post over pre-stimulus coefficients was calculated resulting in 12 values. This was repeated using three possibilities for post-stimulus data:

- All post stimulus data
- Waves I-V: 1.5ms to 9.5ms post-stimulus
- Waves V: 5ms to 9.5ms post-stimulus

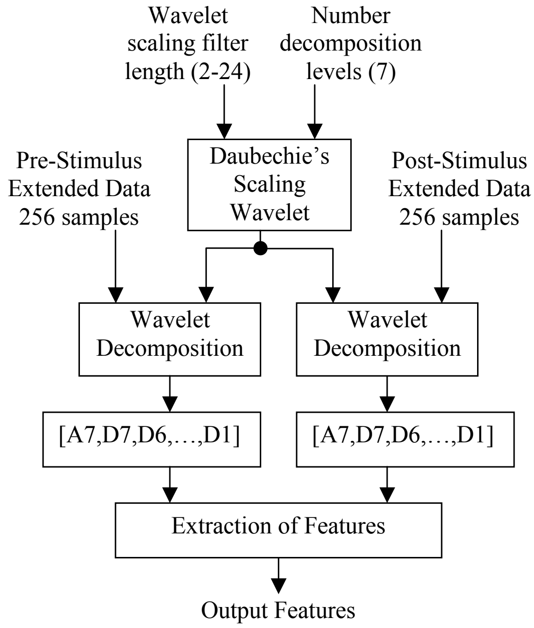

Figure 6: Feature Extraction

When an expert is unsure if a response is present, repeat recordings, using the same conditions, are performed so that a visual comparison can be made between the waveforms. In a similar manner, Davey [5] used correlation of the repeat recordings in the time and frequency domain to help classify the waveforms. Within the wavelet domain it was noted that there was also a correlation between the coefficients of repeated tests. It was believed that this could provide useful information and three variations were considered. Firstly, the cross-correlation value of the D4 coefficients from repeat recordings was calculated. Then, it was considered that only the portion of the coefficients that best represented the ABR waveform would be used by an inverse wavelet transform to convert the data back into the time domain, whereby cross-correlation values would then be calculated on the reconstructed waveforms. This was done using the D4 coefficients and also for coefficients D4-A7.

Tables 2 and 3 give a summary of the features extracted from the wavelet and time domain analysis. They were used as inputs to train decision trees to automate the classification. The Clementine data mining software system (http://www.spss.com) was used to model underlying relationships and features in the data using the C5.0 decision tree algorithm [9]. The rule set derived in the training of the decision tree gave insight into the features that dominated the classification and those that had no relevance.

Even extension was used with only a scaling filter of length 6 for the wavelet decomposition as experiments showed this to give some of the most consistent results, although further analysis could be performed to support this.

Table 2: Extracted Features
(R relates to post-stimulus data, N to pre-stimulus data)

Feature	Description
Wavelet coefficients, scaling filter length 6, waves I to V	
Rd6, Rd5, Rd4	Post-Stimulus D6, D5, D4
Nd6, Nd5, Nd4	Pre-Stimulus D6, D5, D4
Wavelet D4 coefficient ratios, waves I to V	
WD_RNd4_wlt2 WD_RNd4_wlt4 … WD_RNd4_wlt24	Ratio of post/pre-stimulus mean absolute D4 coefficients, for scaling filters of length: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 and 24
WD_RNd4_wlt2_14	Averaging ratios for scaling filters 2 to 14
WD_RNd4_wlt2_24	Averaging ratios for scaling filters 2 to 24
Additional Features	
TD_RN_w1to5	Power ratio of post-stimulus (waves I to V) waveform over pre-stimulus waveform
Stimulus level	Stimulus applied

Table 3: Features extracted from repeated recordings

Feature	Description
Time domain cross-correlation (CCR) features	
TD_Pre_CCR	Pre-stimulus part
TD_w1to5_CCR	Post-stimulus part: waves I to V
TD_w5_CCR	Post-stimulus part: wave V
Wavelet domain cross-correlation of D4 coefficients	
WD_Pre_CCR	Pre-stimulus coefficients
WD_Post_CCR	Post-stimulus coefficients
WD_w1to5_CCR	Post-stimulus coefficients: waves I to V
WD_w5_CCR	Post-stimulus coefficients: wave V
Wavelet domain cross-correlation of reconstructed waveforms using only the D4 coefficients	
XWD_Pre_CCR	Pre-stimulus part
XWD_Post_CCR	Post-stimulus part
XWD_w1to5_CCR	Post-stimulus part: waves I to V
XWD_w5_CCR	Post-stimulus part: wave V
Wavelet domain cross-correlation of reconstructed waveforms using coefficients D4-A7	
X1WD_Pre_CCR	Pre-stimulus part
X1WD_Post_CCR	Post-stimulus part
X1WD_w1to5_CCR	Post-stimulus part: waves I to V
X1WD_w5_CCR	Post-stimulus part: wave V

Results

A study was performed on the full dataset using the WD_RNd4_wlt2_14 as a feature for classification. A comparison was made between the three types of data extension. For each method the different formats for post stimulus data were analysed; all post-stimulus data, waves I to V, and finally, just wave 5. The results are presented in Table 4. All WD_RNd4_wlt2_14 values above a set threshold of √6 indicated a response. Values beneath the threshold were left unclassified. From Table 4 it can be seen that the best accuracy was produced when even extension was applied using only waves I to V in the post-stimulus data. Hence, this was the set up used in stage one of the classification.

Table 4: Using WD_RNd4_wlt2_14 feature on full data set with a threshold of √6

Average values	Data Extension Type		
	Zero	Even	Odd
All Post-Stimulus Data			
Total Classified	205.1	181.9	174.0
Classified %	35.3	31.3	30.0
Accuracy %	**96.4**	**91.1**	**90.2**
Post-Stimulus Data 1.5–9.5ms			
Total Classified	214.7	191.3	191.6
Classified %	37.0	32.9	33.0
Accuracy %	**95.1**	**97.6**	**97.1**
Post-Stimulus Data 5–9.5ms			
Total Classified	166.1	216.6	215.6
Classified %	28.6	37.3	37.1
Accuracy %	**94.5**	**96.8**	**97.3**

The analysis in the time domain of the power ratios was repeated to ensure that the comparisons were made on the exact same dataset. Table 5 and Table 6 give results for the first stage of classification using the full dataset and a range of thresholds. It can be seen that using WD_RNd4_wlt2_14 provided marginally better results. At a threshold of 5, a quarter of the test cases were classified using the time domain feature TD_RN_w1to5 with two wrong interpretations. Using the wavelet domain feature WD_RNd4_wlt2_14 a third of test cases were classified, using a threshold of √5, with only 1 incorrect interpretation.

Table 5: Time Domain Power Ratio Results: Waves I-V

Threshold	Total	Yes	No	% Correct	% Of tests
4	190	186	4	97.9	32.7
5	**146**	**144**	**2**	**98.6**	**25.1**
6	121	119	2	98.3	20.8
7	104	103	1	99.0	17.9
8	85	85	0	100	14.6

Table 6: Wavelet Domain (WD_RNd4_wlt2_14) D4 Ratio Results: Even extension, waves I-V

Threshold: √ of value	Total	Yes	No	% Correct	% Of tests
4	237	233	4	98.3	40.8
5	**195**	**194**	**1**	**99.5**	**33.6**
6	167	166	1	99.4	28.7
7	145	144	1	99.31	24.0
8	132	131	1	99.2	22.7

The 195 test cases, classified using WD_RNd4_wlt2_14 as the feature, were removed from the dataset. The remaining records were passed to a second stage of the classification process whereby additional features based on the cross-correlation (CCR) of repeat recordings were extracted and used to train a decision tree to aid the classification. In Table 7 a comparison was made between the time domain (TD) cross-correlation features and three variations of wavelet correlation features:

WD: D4 Coefficients

XWD: Reconstructed time domain waveform using D4 Coefficients

X1WD: Reconstructed time domain waveform using D4-A7 Coefficients

Table 7 shows example variations of decision tree inputs. Averages were calculated for all valid variations, that is, conditions where no features were selected and just the pre-stimulus feature was selected were not considered. The results given in Table 7 show that there is an increase in accuracy when using the time domain waveform reconstructed from coefficients D4-A7. This shows that the wavelet features provide comparable value in automated response detection.

Table 7: Comparison of time domain versus wavelet domain correlation features (6-fold validation tests)

Features used by decision tree: portion of ABR used				**Accuracy in Classification (%)**			
Pre	**Post**	**I-V**	**V**	**TD**	**WD**	**XWD**	**X1WD**
			Y	71.8	68.7	61.8	**73.8**
		Y		71.0	65.3	62.5	**67.9**
	Y			64.2	61.4	61.3	**69.3**
		Y	Y	73.2	69.3	61.9	**74.7**
	Y	Y	Y	73.8	70.9	62.1	**73.9**
Y		Y	Y	74.4	71.1	62	**74.9**
Y	Y	Y	Y	75.0	71.2	63.1	**74.2**
Overall average over all valid variations				**71.9**	**68.3**	**62.1**	**72.7**

The results in Table 8 show a varied range of input features to the decision tree. Tests 1 to 4 investigate the effect of WD_RNd4_wlt2_14 on the results in Table 7. The additional feature increased the accuracy when coupled with all the cross-correlation features, reaching an accuracy of 80% when coupled with the X1WD cross-correlation features. All the correlation features derived for both the time and wavelet domain were considered in Test 5, obtaining a 75.8% accuracy, which was enhanced further in Test 6 to 79.4% by including WD_RNd4_wlt2_14.

Test 7 looked at the value of the individual wavelet D4 ratios (defined in Table 2) for each wavelet scaling filter length, resulting in an accuracy of 78.7%. These resulted in 12 input features, however, by looking at the rule set derived by the C5.0 algorithm it could be seen that scaling wavelet lengths 14, 22 and 24 were not included in the decision tree, and could thus be removed. The accuracy did not improve in Test 8 to by adding all the cross-correlation features derived from the time domain waveforms reconstructed by wavelet coefficients D4 to A7 (X1WD).

The raw wavelet coefficients were also studied as potential features (Tests 9-12). From Table 8 it can be seen that by inputting the raw D4 coefficients for both the pre and post-stimulus waveforms that there were no significant findings.

Table 8: Results using C5.0 Decision Tree Algorithm (6-fold cross validation tests)

Test	**Features**		**% Correct**
1	All TD CCR features	WD_RNd4_wlt2_14	78.1
2	All WD CCR features	WD_RNd4_wlt2_14	73.8
3	All XWD CCR features	WD_RNd4_wlt2_14	74.1
4	**All X1WD CCR features**	**WD_RNd4_wlt2_14**	**80.0**
5	All TD, WD, XWD & X1WD CCR features		75.8
6	All TD, WD, XWD & X1WD CCR features	WD_RNd4_wlt2_14	79.4
7	WD_RNd4_wlt2 WD_RNd4_wlt4 … WD_RNd4_wlt24		78.7
8	WD_RNd4_wlt2 WD_RNd4_wlt4 … WD_RNd4_wlt24	All X1WD CCR features	78.4
9	Rd6 (4 coefficients)	Nd6 (4 coefficients)	61.1
10	Rd5 (8 coefficients)	Nd5 (8 coefficients)	60.7
11	Rd4 (16 coefficients)	Nd4 (16 coefficients)	52.0
When applied to full data including high level responses classified in stage 1:			
12	Rd4 (16 coefficients)	Nd4 (16 coefficients)	69.5

Discussion

Previous studies indicate that ABRs may be classified based on time and frequency domain features. In this paper it is shown that the wavelet domain also provides useful features for automated ABR detection. Using an extensive set of data comprising of high and low level ABRs from a wide age range, a study was performed on the use of wavelet decomposition to extract features to assist classification. The data set was the same as that used for a previous study that focussed on the time and frequency domain [5]. This enabled an accurate comparison to be made of the wavelet features detailed in this paper to existing methods.

A high accuracy of 99.5% in classifying strong responses was obtained by using a power ratio derived from the mean absolute values for the D4 wavelet coefficients of the post-stimulus waveform over pre-stimulus waveform. A threshold was set ($\sqrt{5}$) above which a response was indicated and the classified waveform was removed from the test set. The remaining waveforms consisted of lower level, threshold, and no responses conditions and hence would be more difficult to classify. Weaker responses may rely on repeated tests to allow cross-correlation values to be calculated and used as response indicators. This adopts the approach employed by experts when interpreting lower stimulus level recordings.

Features were derived from the cross-correlation of wavelet coefficients from repeated tests. In addition, D4 coefficients and also D4-A7 coefficients were used to reconstruct the ABR signals back into the time domain, from which cross-correlation parameters were calculated for the repeated recordings. It was believed that the smoothing of the ABR would enhance the features. A moderate improvement on accuracy was obtained when coefficients D4-A7 were used to reconstruct the ABR back into the time domain.

A selection of features based on power values, raw wavelet coefficients and cross-correlation parameters were studied as inputs to train decision trees using the standard C5.0 algorithm. Results showed only slight improvements when using the wavelet domain cross-correlation features over the equivalent time domain features. However, a highest accuracy of 80% was obtained when wavelet cross-correlation features were coupled with the time domain cross-correlation features.

Using the raw D4 coefficients for the post and pre-stimulus waveforms (scaling filter length of 6) did not provide useful feature information.

Conclusions

The features extracted from the wavelet domain provide some improvement over equivalent time domain features. In particular a third of the test cases displaying high-level responses were classified to a 99.5% accuracy using the ratio of the post over pre-stimulus mean absolute D4 coefficients. Using the power ratio in the time domain a quarter of the same test set was classified to an accuracy of 98.5%.

Cross-correlation features were derived from the wavelet domain, and provided a minor increase in accuracy over the equivalent features in the time domain. Better results were obtained by combining the time and wavelet correlation features which resulted in an accuracy of 80%.

Acknowledgements

Appreciation goes to H. G. Houston, from the Audiology Department of the Royal Group of Hospitals in Northern Ireland, for providing the ABR test data.

References

[1] MALLAT S. G., (1989): 'A Theory for Multiresolution Signal Decomposition: the Wavelet Representation', *IEEE Trans. on Pattern Anal. Machine Intelligence,* **11,** pp. 674-693

[2] BRADLEY A. and P. WILSON W. J., (2004): 'On Wavelet Analysis of Auditory Evoked Potentials', *Journal of Clinical Neurophysiology,* **115**, pp. 1114-1128

[3] PRATT, T. L., W. O. OLSEN, ET AL., (1995): 'Four-channel ABR recordings: Consistency in interpretation,' *American Journal of Audiology*, vol. **4,** no. **2,** pp. 47-54

[4] DELGADO, R. E. AND O. OZDAMAR (1994): 'Automated Auditory Brainstem Response Interpretation,' *IEEE Engineering in Medicine and Biology*, pp. 227-237

[5] DAVEY R. T., MCCULLAGH P. J., MCALLISTER H. G., HOUSTON H. G., (2006): 'The use of artificial neural networks for the objective determination of hearing threshold, using the auditory brainstem response', *Neural Networks in Healthcare: Potentials and Challenges*, IDEA Group Inc (to be published January 2006)

[6] DAUBECHIES, I. (1992): 'Ten Lectures on Wavelets', in 'CBMS-NSF Regional Conference Series in Applied Mathematics', vol. 61, (SIAM, Philadelphia)

[7] GUTIERREZ, A.; SOMOLINOS, A. (2000): 'Influence of wavelet boundary conditions on the classification of biological signals,' Bioengineering Conference, 2000. Proceedings of the IEEE 26th Annual Northeast, pp. 25-26

[8] OZDAMAR, O., R. E. DELGADO, ET AL., (1994): 'Automated electrophysiologic hearing testing using a threshold-seeking algorithm,' *Journal of Am Acad Audiol*, **5**, pp. 77-88

[9] QUINLAN, J. R. (1986): 'Induction of decision trees,' *Machine Learning*, **1**, pp. 81-106.

ISSN: 1727-1983