Fuzzy Sets for feature identification in biomedical signals with self-assessment of reliability

An adaptable algorithm modelling human procedure in BAEP analysis

Justus Piater [1, 2]
Franz Stuchlik [1]
Hellmut von Specht [2]
Roland Mühler [2]

1. University of Magdeburg, Computer Science Department, Institute for Information and Communication Systems
2. University of Magdeburg, Medical Department, Division for Experimental Audiology/Medical Physics

Correspondence to: J. Piater, IIK, Computer Science Department
University of Magdeburg
Postfach 4120
D-39016 Magdeburg, Germany

Abstract

The Analysis of biomedical signals has mostly been restricted to traditional signal processing methods. This article proposes a different approach, applied to the evaluation of Brainstem Auditory Evoked Potentials (BAEP). An automatic peak identification method is described which uses criteria applied by human evaluators in visual analysis. These criteria are defined as fuzzy sets and are combined using fuzzy operations, thus reflecting the weighting of different facts by humans. Membership values are interpreted as degrees of satisfaction which indicate the degree to which a sample satisfies a given criterion. The system judges its own performance in terms of degrees of reliability. Tests on a large set of clinical data showed a high performance on good and average quality curves. A substantial drawback was the assignment of too many peaks in poor potentials. The approaches presented here can easily be applied to similar one-dimensional (and higher) signal analysis tasks.

1 Introduction

Biomedical signals are usually analyzed either visually or by means of computers. In the latter case, signal processing methods are applied, such as Fourier transform or transformations for feature enhancement and suppression, clustering, and statistical methods. However, it can be very difficult to characterize the relationship between the computer-based and visual examination techniques. Consequently, such methods are hard to assess, which is the case in the evaluation of Brainstem Auditory Evoked Potentials (BAEP). In the following, a fuzzy set based system for BAEP peak identification is described, featuring self-assessment of the reliability of its output, and the possibility of adaptation to specialized objectives. Although the methods are described in this particular context, they can be extended and likewise applied to other areas of signal processing.

1.1 Brainstem Auditory Evoked Potentials

For some time now, Brainstem Auditory Evoked Potentials have been successfully used in audiometry and neuro-otology, specifically in the detection and localization of sources of hearing impairments. They are obtained by application of acoustic (e.g. rectangular) stimuli to one ear. The electric response is recorded between vertex and mastoid using scalp electrodes. In some cases, potentials are recorded from both sides (ipsi- and contralateral with respect to the stimulated ear). In order to obtain a reasonable signal-to-noise-ratio (SNR), up to several thousand so-called sweeps have to be averaged.

Evoked potentials show both intra- and interindividually reproducable features. In this article, five typical peaks (1), numbered I through V, are considered. Most of their diagnostic significance lies in the peak latencies (elapsed time after the stimulus) and their prominence. Figure 1 shows a distinct BAEP, recorded on both ipsi- and contralateral sides.

1.2 Automatic BAEP evaluation

Up to now, BAEP are visually evaluated by human experts. Their performance is based on subjective experience. In order to increase objectivity, numerous computer-based methods have been proposed. Most of these publications concentrate on one of the two central tasks in BAEP evaluation, namely (a) detection of an evoked potential, or (b) the determination of peaks.

Woodworth et al. (2), for instance, applied a *Matched Filter*, trying to detect the presence of a BAEP after few averaging steps and to identify peak V. However, as the signal to be recognized had to be known in advance, a

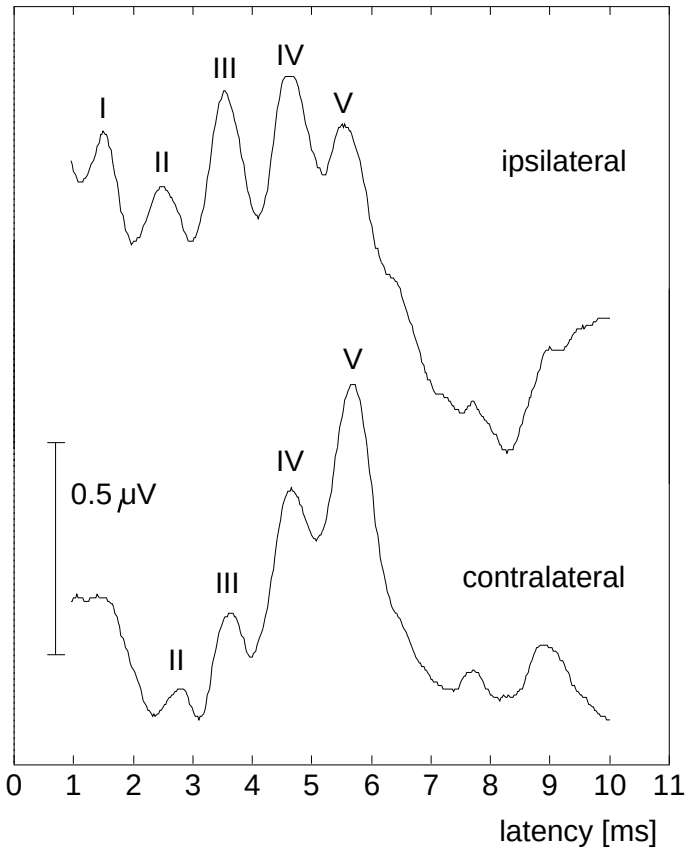

Figure 1: BAEP with very prominent ipsi- (I–V) and contralateral (II–V) peaks.

personal filter had to be designed for each subject. This disadvantage can be avoided by using statistical methods, like BACHEN (3) who suggested testing the phase angles of Fourier harmonics for uniformity of distribution.

Another indication of BAEP presence is the reproducability of the averaged signal which can be tested e.g. through cross-correlation coefficients (4). In a thorough investigation (5), several variance ratios and cross-correlation functions were compared. The latter were found to be superior; however, their results were highly dependent on "typical" features of the potentials.

It was mentioned earlier that the reliability of BAEP detection and peak identification strongly depends on the SNR. COPPOLA (6), WONG (7), and others investigated its influence and suggested methods for its estimation. However, until now peak identification has usually been considered separately. Various methods were proposed, like a backpropagation network (8) trained with visually evaluated data for identifying peak V. MADHAVAN et al. (9) tried to identify peaks I through V using a syntactic method. PRATT et al. (10) enhanced certain features of BAEP while suppressing others, thus improving identification rates.

The latter two suggestions were combined by GRÖNFORS (11) who analyzed the output of several bandpass filters by an attributed automaton. His algorithm, however, could be fooled by the occurence or absence of certain maxima.

None of these methods tried to assess the reliability of its output; neither did they attempt to apply criteria similar to those used by human scorers. In developing the peak determination system presented here, the following goals were set:

- incorporation of rules and heuristics applied by human experts,
- automatic assessment of the reliability of each peak identified,
- total assessment of the reliability of the set of peaks identified,
- improvement of the peak identification rate in medium-quality potentials and
- adaptability/trainability of the algorithm.

While most of the previously published methods rely on traditional signal processing, this algorithm is based on a similar procedure used by many experts in subjective visual evaluation. Typical rules are hard-wired into the system. They are evaluated on the basis of fuzzy set theory. (For an introduction, see e.g. (12).)

An overview of the compontents and their interaction is presented in Fig. 2.

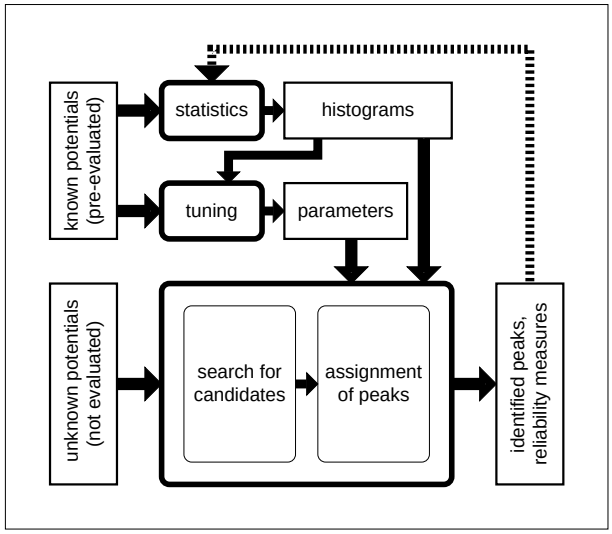

Figure 2: Components of the peak identification system. The upper half represents the parameter initialization, the lower the peak identification procedure.

1.3 Subjects and signal recording

This investigation is based on 895 curves obtained from 143 subjects during several months of clinical routine. Potentials from normal hearing as well as some hearing impaired subjects were included. All ages were represented. Strongly pathological potentials were excluded from the test set.

The potentials were recorded on ipsi- and contralateral sides using a commercial unit (Bera Double Scan by Hörniß & Zeisberg). A rectangular pulse (suction) of duration 100 μs served as the acoustic stimulus, applied to one ear at a rate of 18 Hz using Holmberg headphones. Electrodes were placed on vertex (positive), mastoids on either side (negative) and forehead (ground). The potentials were sampled at 40 kHz in a post-stimulus time window of 0.95–14.95 ms. 2000 so-called *sweeps* were averaged. Any sweep exeeding an amplitude threshold of 10 μV was regarded as containing an artifact and discarded.

2 Rules

In the following, the implemented rules will be explained. The criteria, on which they are based, are emphasized:

1. A peak is distinguished by its *shape*, e.g. curve maximum.

2. Peak V is often marked by its prominent *downslope*.
3. Similarly, peak V usually is also well established on the contralateral side, both latencies roughly *matching*.
4. Typical ranges of *inter-peak latencies* are adhered to.
5. The peaks appear in typical *latency regions* which depend on the intensity of the acoustic stimulus.

None of these rules alone is sufficient for reliable BAEP peak identification. They have to be combined using appropriate weights.

A sample l of an averaged BAEP satisfies a criterion C_i of a rule $i, 1 \leq i \leq 5$, to a *degree of satisfaction* $\eta_i(l) \in [0,1] \subset I\!R$. The value 1 is identified with "complete satisfaction", 0 with "no satisfaction". In determining the suitability or *reliability* $\theta(l)$ of a sample as a peak, the proposed algorithm, like a human expert, takes into account the matching degrees of all rules.

The calculation of the matching degrees η_i are based on the following

Presupposition: The more (less) frequently a certain property of visually identified peaks is encountered, the better (worse) it satisfies the respective criterion.

Hence, *typical* properties are identified with *good* properties with respect to a certain criterion. This assumption was found to be adequate, except for the shape criterion which required special treatment.

2.1 Criteria

Next, the criteria underlying the rules mentioned above will be detailed. Peaks I, III, and V usually are the most prominent ones and are also most important for clinical diagnostics. Hence, as they play a major role in the algorithm, we call them *major peaks* as opposed to the *minor peaks* (II and IV).

Shape: This criterion is the only two-dimensional one. It is represented by the first two derivatives $d^{(1)}(l), d^{(2)}(l)$ of the sample l in consideration. The ith derivative is computed recursively as follows:

$$d^{(i)}(l) = d^{(i-1)}(l+\delta) - d^{(i-1)}(l-\delta) \qquad [1]$$

The parameter δ was, like all parameters mentioned here unless otherwise noted, adjusted using the method described in Sec. 4. Since each of the five peaks is characterized by typical shapes, this criterion is computed separately for each peak class (I through V).

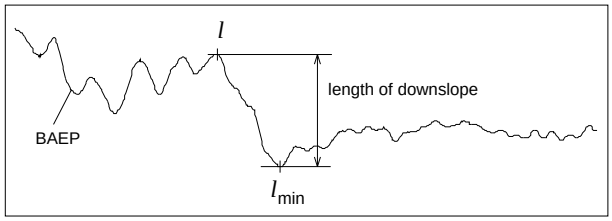

Figure 3: Definition of the downslope.

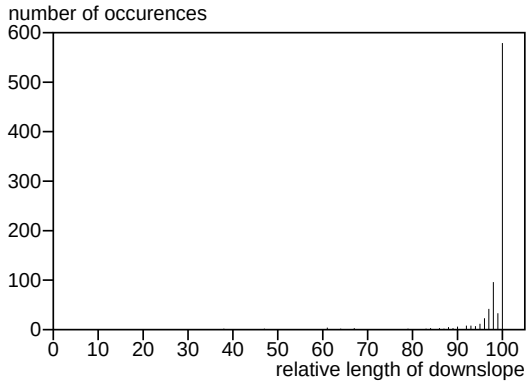

Figure 4: Histogram of the lengths of peak V downslopes as percentage of the longest downslope in the given BAEP.

Downslope: Typically, peak V is followed by the longest downslope of all peaks. To human experts, this fact is of great importance for the identification of peak V. Here, the length d of the downslope associated with a point l is defined in terms of a point $l_{\min}$, its first derivative not exceeding a given threshold t:

$$l_{\min} = \max\{l^* \geq l \mid d^{(1)}(l^*) \leq t\} \qquad [2]$$

As illustrated in Fig. 3, the length s of the downslope is then defined as the difference in amplitude between these two points:

$$s(l) = d^{(0)}(l) - d^{(0)}(l_{\min}) \qquad [3]$$

In order to obtain the amplitude-independent matching degrees $\eta_{\text{downslope}}(l)$, all lengths found within one BAEP are then normalized. A histogram of the relative downslope lengths is shown in Fig. 4.

Ipsi/contralateral match: Another important feature characterizing peak V is its occurence in the contralateral potential. Its latency is roughly

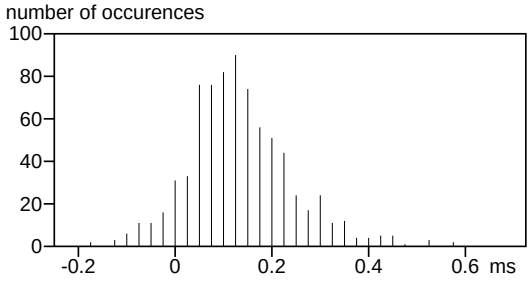

Figure 5: Histogram showing the differences of the contralateral from the ipsilateral peak V latencies.

the same, typically slightly larger than the ipsilateral one. This difference in latency is evaluated by this criterion.

The histogram in Fig. 5 shows typical values.

Distances: The distances between the peaks are important for their identification. These typical inter-peak latencies do not depend on the intensity of the stimulus (13).

This algorithm takes into account the distances of each peak to two *reference peaks*. Only major peaks serve as reference peaks, a major peak's reference peak being the other two (if present) major peaks, a minor peak's the two neighboring ones.

In order to keep this criterion as weak as possible, the two inter-peak latencies were not regarded as correlated (unlike the shape criterion). They were calculated separately, and then combined using a paramterized averaging operator. The cases where only one or none of the reference peaks were available had to be considered separately.

Latencies: In contrast to the inter-peak latencies, the absolute latencies depend on the stimulus intensity. Therefore, a new parameter *ilat* (intensity-related latency) was introduced as a function of absolute latency l and stimulus intensity I:

$$\mathrm{ilat}(I, l) = l - k_1 e^{k_2 I} + k_3 \qquad [4]$$

The parameters k_1 and k_2 were determined by a regression over all peak V latencies l_V available in the test data. Parameter k_3 was arbitrarily chosen such that all *ilat*-values were positive.

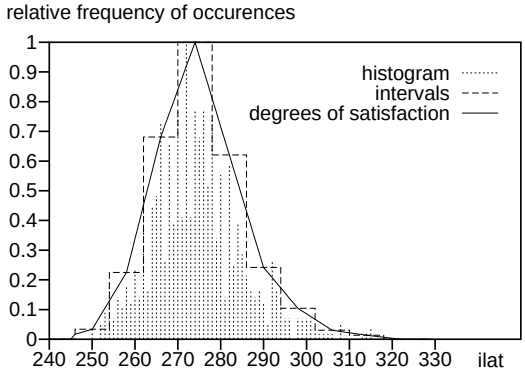

Figure 6: From a histogram to the degree of satisfaction of a criterion (example: Peak V latency after *ilat*-conversion). Presented are the normalized histogram, its subdivision into intervals (rectangular graph) and the interpolated histogram representing the degrees of satisfaction.

2.2 Histograms as a basis for degrees of satisfaction

Acceptable values of the criteria were expressed as fuzzy sets. In order to generate appropriate membership functions, these were derived from histograms of properties of visually identified peaks in the training data.

Several methods of generating fuzzy membership functions from statistical data have been proposed (14), most of which regard possibility measures. If, on the other hand, the requirements imposed on the fuzzy sets and operations are less restrictive, methods can be chosen arbitrarily. (The relationship between probability and fuzziness cannot be discussed here; DE FÉRIET (15), for instance, deals with this matter in depth.) Here, a most obvious method was chosen, as explained in the following, using the *ilat*-criterion as an example.

The peak identification program runs in two phases. During the *initialization phase*, all peak V latencies contained in the training data are converted into their corresponding *ilat*-values. They are inserted into a histogram which is divided into a small number of intervals (Fig. 6). The width of the intervals was empirically determined such that the resulting histogram was unimodal without losing too much accuracy. This histogram, its height normalized to $[0, 1]$, then forms the membership function of a fuzzy set containing acceptable *ilat*-values for peak V.

In the *action phase*, degrees of satisfaction are calculated for a large number of samples of a new curve. The latency of a current sample l is first converted into its *ilat*-value. Then, its degree of satisfaction $\eta_{\text{ilat}}(l) \in [0, 1]$ of this criterion is taken from the normalized histogram mentioned above, applying linear interpolation between the values of adjacent intervals (Fig. 6).

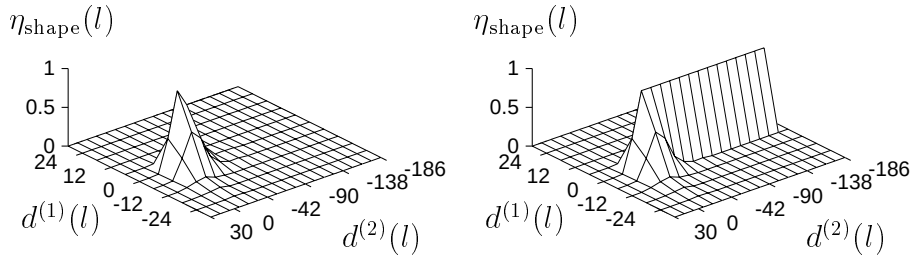

Figure 7: Assessment correction of the shape criterion. The grid represents the segmentation of the two-dimensional shape histogram into intervals. The third dimension displays the η values obtained by linear interpolation, before (left) and after (right) the assessment correction as described in Sec. 2.3.

2.3 Special case: Shape criterion

In the case of the shape criterion, the procedure mentioned above is trivially extended to the two-dimensional case. However, there is one peculiarity about the shape of a peak: The validity of the *Presupposition* is restricted. Very prominent peaks (characterized by highly negative second derivatives) are rare, leading to low degrees of satisfaction $\eta_{\text{shape}}(l)$.

This was compensated by filling the histogram in the area of negative second derivative with its maximum value (Fig. 7). The possiblility of explicitly introducing *a priori* knowledge is one useful feature of fuzzy systems, especially if used in combination with automatic generation of membership functions.

2.4 Weighting of rules

Appropriate weighting of rules is an important issue in the design of knowledge-based systems. The pioneer medical expert system MYCIN (16), for instance, introduced *certainty factors* which, as a heuristic concept, worked in this case but appeared inconsistent in other applications.

Fuzzy set membership functions can be appropriately shaped in order to implicitly reflect the weighting of the underlying rules. Here, this information is contained in the histograms: Narrow, steep histograms result in very *specific* rules, corresponding to a high certainty factor when the rule, "firing" to a high degree, is combined with a different rule. Accordingly, wide histograms with gentle slopes represent *unspecific* rules. The downslope criterion of peak V is an example of a very specific rule (cf. Fig. 4), in contrast to the weaker

criterion of latency match (Fig. 5).

3 Peak identification procedure

Identification of peaks is performed in two steps: First, the qualification $\theta_c(l)$ of each sample l is calculated. It indicates the degree to which l seems qualified as a peak. The two latency criteria are not included in this step. An l with $\theta_c(l) > 0$ we call peak-*candidate*.

During the second step, candidates are combined in various ways, taking into account all criteria. The resulting assignments A are associated with peak qualifications $\theta_p(l)$ and an assignment qualification $\Theta(A)$.

3.1 Collection of candidates

A separate list of candidates is established for each of the peaks I through V. In collecting candidates, a BAEP is scanned sample by sample, calculating each one's qualification:

$$\theta_c(l) = \eta_{\text{shape}}(l) \qquad [5]$$

This results in a distribution of suitabilities over latencies. In the following, only local maxima of this distribution, exeeding a given threshold, are considered.

This concludes the candidate collection procedure for peaks I–IV.

The qualifications of the peak V candidates are subject to further processing: The prominence of the downslope and the ipsi-/contralateral match are considered. The latter is accomplished by compiling contralateral peak V candidates. For each ipsilateral candidate, the best contralateral counterpart is found and the degree of match η calculated. Due to the use of a disjunctive operator, the better one of the two matching peaks dominates the final assessment.

In a nutshell, peak V candidate qualifications are computed according to the following rule:

$$\begin{aligned} \theta_c(l) = \quad & (\eta_{\text{shape}}(l) \ \wedge \ \eta_{\text{downslope}}(l) \vee \quad && \textit{(ipsilateral)} \\ & \eta_{\text{shape}}(l_c) \ \wedge \ \eta_{\text{downslope}}(l_c)) \quad && \textit{(contralateral)} \\ & \wedge \ \eta_{\text{ipsi-/contral.match}}(l) \end{aligned} \qquad [6]$$

The conjunctive ($\wedge$) and disjunctive ($\vee$) operators represent parametrized *t-norms* and *t-conorms* as proposed by Schweizer and Sklar (17). Their definition covers the entire range from *drastic product/sum* to *min/max* operators, respectively (12). These operators are adjusted separately (cf. Sec. 4).

3.2 Determination of the best combination

In the second step, the final assignment of peaks I–V is selected from a ranked list of combinations, each of which are built from one (or no) peak from each of the five candidate lists. This procedure can be summarized as follows:

1. Sequentially, all possible combinations of major peak candidates are formed. Suitability values of the inter-peak latencies $\eta_{\text{distances}}(l)$ are calculated into the respective qualifications:

$$\theta_p(l) = \theta_c(l) \wedge \eta_{\text{distances}}(l) \qquad [7]$$

   For each combination, the following steps are performed:

   (a) The best minor peak candidates are selected. Again, the $\eta_{\text{distances}}(l)$ are included into their $\theta_p(l)$.

   (b) Of the resulting combination A, the quality $\Theta(A)$ as a peak assignment is calculated. The absolute latency suitabilities of the proposed peaks are included using a fuzzy conjunctive operator.

2. Finally, all assignments A_i are sorted in descending order of quality $\Theta(A_i)$.

The calculation of the qualifications of peaks and assignments can be summarized as follows:

$$\theta_p(l) = \theta_c(l) \wedge \eta_{\text{distances}}(l) \qquad [8]$$

$$\Theta(A) = h(\theta_p(l_{\text{I}}), \ldots, \theta_p(l_{\text{V}})) \wedge h(\eta_{\text{ilat}}(l_{\text{I}}), \ldots, \eta_{\text{ilat}}(l_{\text{V}})) \qquad [9]$$

h is a parametrized averaging operator. Some additional mechanisms are not described here. They deal with indistinct or ambiguous peaks which are hard to identify and thus cannot be categorized.

Peaks, which are found by this procedure, can be fed back into the histograms (cf. Sec. 2.2), thus updating the fuzzy membership functions and allowing the system to "learn from experience." Through proper use of this feature, the system may further improve its performance.

4 Parameter tuning

4.1 Specification of the goal

For the evaluation of an automatic peak identification system, its peak assignments must be compared with those of human experts. In each BAEP of the

training set, up to five ipsilateral peaks and, where applicable, a contralateral peak V, had been marked by visual evaluation. These peaks we call *pre-set*, those that intentionally had not been marked *not pre-set*.

Peaks assigned by the automatic system are named *identified*, those not assigned *not identified*.

When the automatically and visually obtained results are compared, each of the five possible peaks falls into exactly one of the following five *categories of identification*:

1. *correct*: pre-set and correctly identified
2. *correctly not identified*: not pre-set and not identified
3. *overdetermined*: not pre-set but identified
4. *underdetermined*: pre-set but not identified
5. *bad*: pre-set and incorrectly identified

A peak was defined to be correctly identified if the difference between the automatically and visually determined latencies did not exceed 0.1 ms.

4.2 Algorithm

The parameters of the fuzzy operators (and other system parameters) had to be fine-tuned according to the following objectives:

- high number of *correct* peaks,
- low number of *bad* peaks,
- low number of *underdetermined* and
- low number of *overdetermined* peaks.

A performance index Γ was defined as follows:

$$\Gamma = \frac{n_{\text{correct}}}{2n_{\text{bad}} + 1/2n_{\text{under}} + \gamma n_{\text{over}} + 1} \qquad [10]$$

The values n indicate the numbers of peaks obtained in the given categories. The term +1 keeps the fraction finite even in the ideal case. Equation [10] was empirically determined. Several values of γ were investigated as discussed in Sec. 5.4. In the following, its value is assumed to be zero.

The problem of optimizing the system parameters is thus transformed into maximizing Γ, which means the practically unsolvable problem of finding the global maximum in a high-dimensional hyper-surface of unknown shape. Such problems are typically dealt with using non-deterministic methods like Simulated Annealing or Genetic Algorithms. However, the computational cost to obtain one Γ value appeared too high.

Therefore, a simplified solution had to be found. In this case, a greedy hill-climbing algorithm was applied: All parameters were iteratively varied, the variance shrinking in every iteration. The algorithm terminated when the variance fell below a given threshold. This diminished the chance of getting trapped on small local maxima, but in the end lead to a fine-tuned system.

This step is intended to be performed only once (after the first initialization phase described in Sec. 2.2), while action and initialization phases may be iterated and typically lead to different fuzzy membership functions, according to the underlying BAEP data.

5 Results

As summarized in Fig. 8, most peaks were *correctly* identified. The portions of the *bad* and *underdetermined* peaks were by far the lowest (around 3 %). Few peaks were *correctly not* identified (because few had not been pre-set either), but nearly 14 % were *overdetermined*. The algorithm obviously had problems judging the ambivalence or distinctness of a peak.

Significantly more major than minor peaks were *correctly* identified, whereas more minor peaks were not identified or *overdetermined*. This was due to the fact that the major peaks could be more reliably detected and thus had been more frequently pre-set than minor peaks.

Table 1 shows the high performance of the system if restricted to verifiable peaks. Almost all peaks which were both pre-set and identified were *correctly* identified. The differences in latency, which occurred in these cases, are displayed in Fig. 9.

5.1 Subjective judgment of the performance

Generally, prominent peaks received high reliability values $\theta_p \approx 1$, while indistinctly shaped ones got values near zero. Accordingly, the reliability values Θ of the total assignments showed a high intuitive significance: Potentials commonly considered unequivocal mostly received clearly higher values than poor or ambiguous ones.

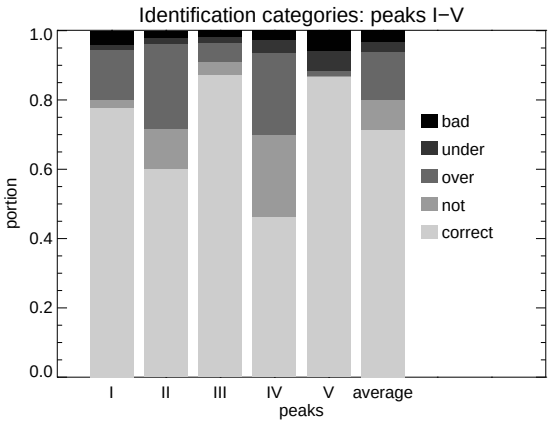

Figure 8: Performance of the peak identification system. Each bar shows the portions of peaks belonging to each of the five categories of identification (Sec. 4.1). The bar on the right displays the average across all peaks.

Table 1: Performance of the peak identification system. Only peaks which were both pre-set and identified are considered. The rows contain the quota distinguished by peaks; the bottom row displays the arithmetic means. The values in each row add up to 1.0.

Peak	correct	bad
I	0.95	0.05
II	0.97	0.03
III	0.98	0.02
IV	0.94	0.06
V	0.94	0.06
all	0.96	0.04

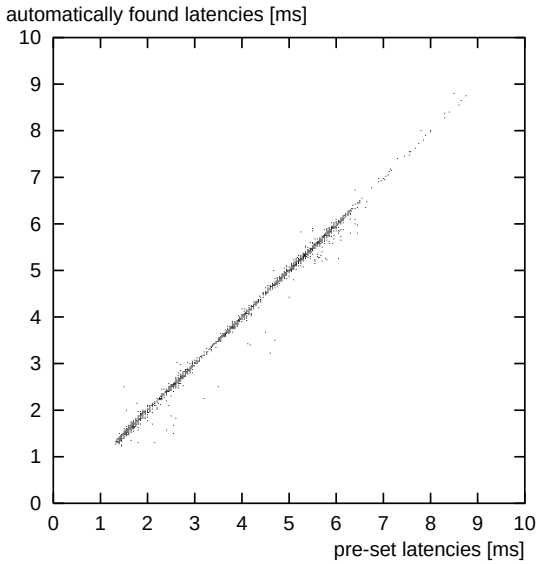

Figure 9: Accuracy of the automatic peak assignment.

Figures 10–13 present some examples of automatic peak assignments. Pre-set peaks are marked on the top, identified peaks on the bottom graphs. The numbers on the right indicate the respective latencies in milliseconds. To their left, the reliability values θ_p are indicated where appropriate. The bottom number is the reliability Θ of the total assignment.

Figure 10 shows a BAEP where peaks IV and V are not clearly separated (IV-V complex). Still, both were *correctly* identified. An appropriately low reliability value was assigned to the ambiguous and poorly shaped peak II. $\Theta = 0.34$ indicates a fair reliability of the total peak assignment.

The BAEP shown in Fig. 11 lacks peak I (the obvious maximum appears too early). Peak IV is wide and round, hence hard to determine precisely, and therefore received a low reliability. The poor Θ value resulted from high absolute latencies.

The main weakness of the proposed algorithm lies in its failure to consequently disregard poor or ambiguous peaks. In the BAEP shown in Fig. 12, only peak V had been pre-set. Its location could be verified by its downslope and its contralateral counterpart (not shown in the figure). Still, the system suggested all five peaks, even assigning high reliabilities to some of them. However, the uncertain total assignment was reflected in the low total reliability.

Well-established potentials, as illustrated in Fig. 13, received high reliabilities throughout.

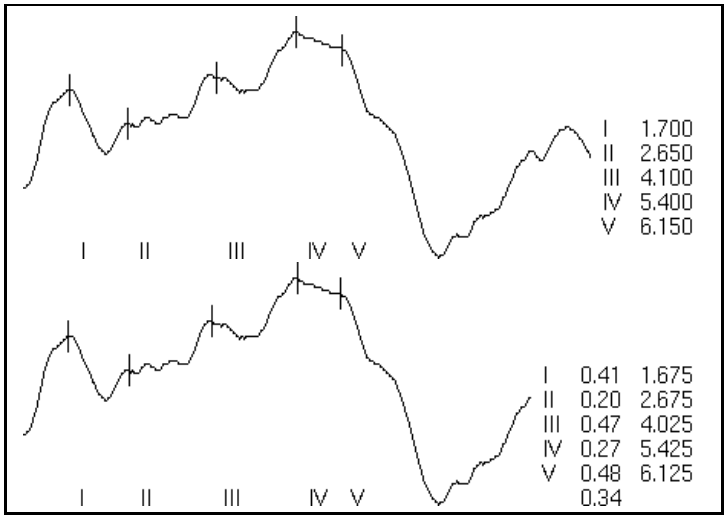

Figure 10: Acceptable assignment and assessment of a IV-V-complex.

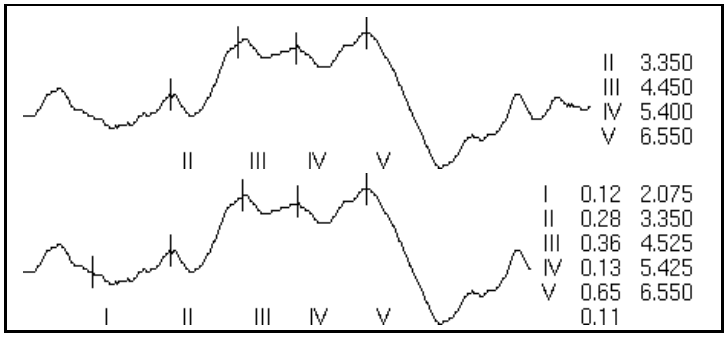

Figure 11: Missing peak I.

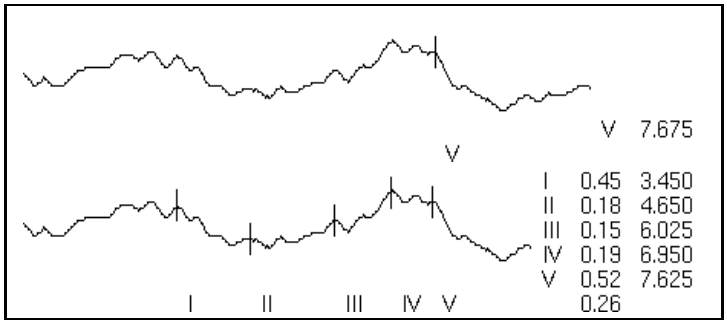

Figure 12: Problem: low stimulus intensity. Too many peaks were assigned.

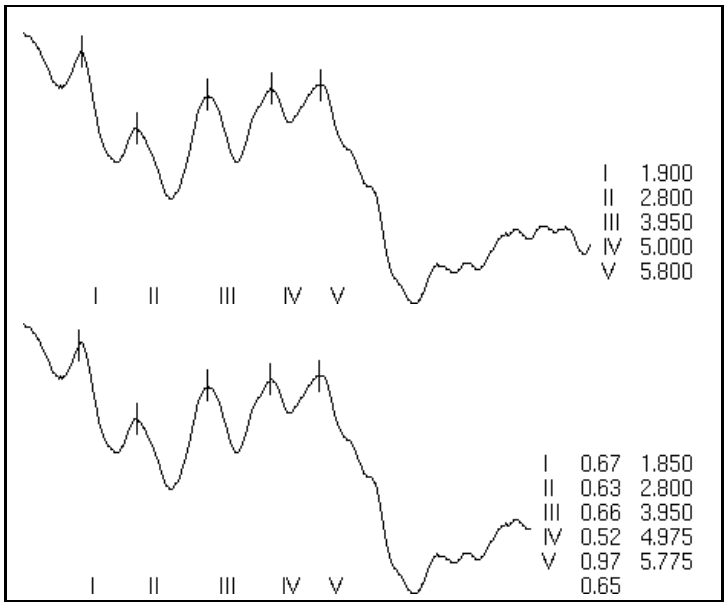

Figure 13: Clearly shaped potential.

5.2 Total assessment Θ

Figure 14 shows the quality of the total assessment with respect to the categories of identification. The portion of *correct* identifications grew monotonously with the total assessment Θ, while all other categories decreased. Hence, high values of Θ reliably indicated correct assignments.

The portions of bad and underdetermined peaks were generally small. Apparently, hardly recognizable peaks with small Θ fell mainly into the categories of *overdetermined* or *correctly not* identified peaks. This overvation confirms that uncertain potentials were generally associated with low values of Θ, even if more peaks were automatically identified than an expert would assign (cf. Fig. 12).

5.3 Peak assessment θ_p

The quality of the peak assessment with respect to the five categories of identification is shown in Fig. 15. The runs of the five portions resemble those in Fig. 14. The growing portion of *overdetermined* peaks from the first to the second interval results from the fact that the first interval includes the peaks not identified, as can be seen in the portions of *correctly not* identified and *underdetermined* peaks.

Again, the portion of badly identified peaks was low and still shrank with

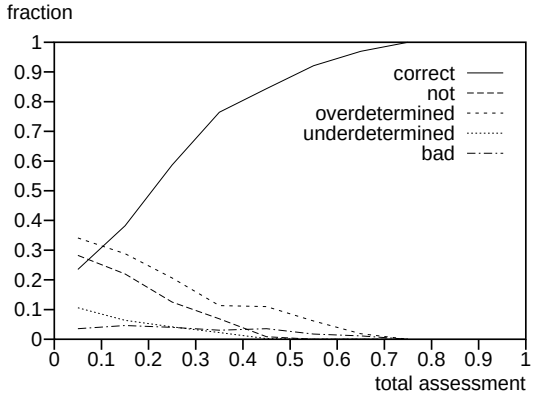

Figure 14: Significance of the total assessment Θ with respect to the determination categories.

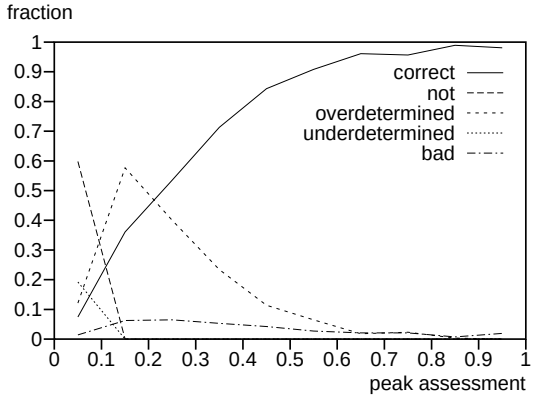

Figure 15: Significance of the Peak Assessment θ_p with respect to the identification categories (mean value of peaks I–V).

growing θ_p. Likewise, the number of *overdetermined* peaks decreased, as is to be expected in clear curves where almost always all five peaks had been pre-set. The curves show that peaks associated with high θ_p can be regarded as highly reliable. At a value of $\theta_p \geq 0.8$, the *correct* identification rate approaches 100 %.

The graphs presented in Figs. 14 and 15 provide a basis for the interpretation of total and peak assessments Θ and θ_p, respectively. For instance, it would make sense to verbalize the assessment values as shown in Tab. 2.

5.4 Changing the performance assessment

As mentioned above, one major problem of the peak identification system was its failure to discard ambiguous or poorly shaped peaks. To compensate for

Table 2: Verbalization of assessment values.

$0.0 \leq \Theta < 0.1$	questionable
$0.1 \leq \Theta < 0.2$	uncertain
$0.2 \leq \Theta < 0.5$	fairly reliable
$0.5 \leq \Theta < 0.8$	reliable
$0.8 \leq \Theta \leq 1.0$	certain

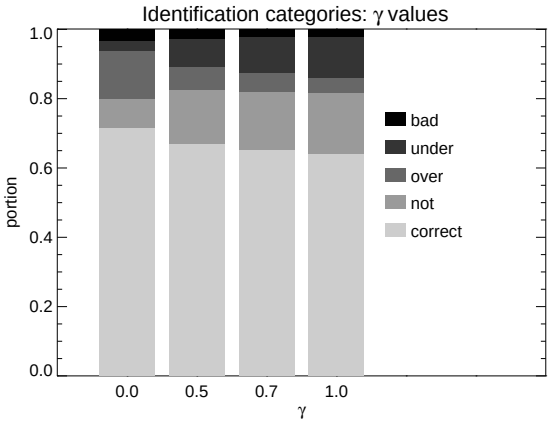

Figure 16: Influence of the parameter γ in the definition of Γ [10]. Only the averages of all peaks are shown. The leftmost bar, which equals the rightmost one from Fig. 8, is shown here for convenience.

this, the parameter adjustment algorithm was forced to keep the number of *overdetermined* peaks low. This was achieved by assigning a value greater than zero to γ [10]. Three values were tried as summarized in Fig. 16.

It can easily be seen that, in fact, the number of *overdetermined* peaks could significantly be reduced. Even the number of *bad* peaks decreased a little, because the system acted more cautious in assigning peaks. On the other hand, the number of *underdetermined* peaks dramatically increased, thus reversing the problem. It was not possible to keep both of them low. This proves that further information has to be taken into account in order to reliably detect and discard uncertain peaks.

In practice, the definition of Γ cannot be changed arbitrarily as the optimization process affects all parameters. Such changes naturally affect the resulting ranges and distributions of the peak and total assessments and, hence, their interpretation.

6 Conclusion

The results illustrate the high performance of fuzzy operations in modelling human recognition and decision processes. Among medium and well distinguished potentials, a high correspondence of automatically identified and visually determined peak latencies was achieved. The reliability values, allotted automatically by the system, had a high intuitive significance. They correlated well with the desired recognition rates and were easily verbalized for intuitive interpretation.

However, weak and ambiguous potentials were not properly dealt with. Often, too many peaks were identified, some of them with overestimated reliability values θ_p. On the other hand, the total assessment Θ usually indicated a low reliability in these cases.

Attempts to keep the number of *overdetermined* peaks low resulted in an overly cautious behavior, discarding too many peaks. For use in clinical routine, this problem has to be eliminated by introducing additional criteria such as signal theoretic SNR estimates.

In the initialization phase, the tuning of parameters allows the adaptation the system to specialized tasks, e.g. typical pathologic potentials. On the other hand, if improperly used, this feature can also lead to a badly tuned system, resulting in failure to deal with the variety of data occuring in routine clinical use.

The methods presented in this article are suitable for use in other biomedical signal analysis tasks as well. They can also be extended to two-dimensional signals and applied to e.g. tumor localization in magnetic resonance images, as well as technological problems like fault detection in surfaces.

References

1. Jewett, D., and Williston, J. Auditory-evoked far fields averaged from the scalp of humans. *Brain* **94**, pp. 681–696 (1971).

2. Woodworth, W., Reisman, S., and Fontaine, A. B. The detection of auditory evoked responses using a matched filter. *IEEE Transactions on Biomedical Engineering* **30** (7), pp. 369–376 (1983).

3. Bachen, N. I. Detection of stimulus-related (evoked response) activity in the electroencephalogram (EEG). *IEEE Transactions on Biomedical Engineering* **33** (6), pp. 566–571 (1986).

4. Sininger, Y. S. Auditory brain stem response for objective measures of hearing. *Ear and Hearing* **14** (1), pp. 23–30 (1993).

5. Özdamar, O., Delgado, R. E., Eilers, R. E., and Widen, J. E. Computer methods for on-line hearing-testing with auditory brain stem responses. *Ear and Hearing* **11** (6), pp. 417–429 (1990).

6. Coppola, R., Tabor, R., and Buchsbaum, M. S. Signal to noise ratio and response variability measurements in single trial evoked potentials. *Electroencephalography and Clinical Neurophysiology* **44**, pp. 214–222 (1978).

7. Wong, P. K., and Bickford, R. G. Brain stem auditory evoked potentials: The use of noise estimate. *Electroencephalography and Clinical Neurophysiology* **50**, pp. 25–34 (1980).

8. Freeman, D. T. Computer recognition of brain stem auditory evoked potential wave V by a neural network. *Ann Otol Rhinol Laryngol* **101**, pp. 782–790 (1992).

9. Madhavan, G., de Bruin, H., Upton, A., and Jernigan, M. Classification of brain-stem auditory evoked potentials by syntactic methods. *Electroencephalography and Clinical Neurophysiology* **65**, pp. 289–296 (1986).

10. Pratt, H., Geva, A., and Mittelman, N. Computational waveform analysis and classification of auditory brainstem evoked potentials. *Acta Otolaryngol (Stockh)* **113**, pp. 279–284 (1993).

11. Grönfors, T. Peak identification of auditory brainstem responses with multifilters and attributed automaton. *Computer Methods and Programs in Biomedicine* **40**, pp. 83–87 (1993).

12. Klir, G. J., and Folger, T. A. "Fuzzy Sets, Uncertainty, and Information." Prentice-Hall, Englewood Cliffs, 1992.

13. von Specht, H., and Kraak, W. Akustisch evozierte Potentiale. *In* "Angewandte Akustik" (Kraak, W., Ed.), pp. 76–101. Verlag Technik, Berlin, 1990.

14. Dubois, D., and Prade, H. Unfair coins and necessity measures: Towards a possibilistic interpretation of histograms. *Fuzzy Sets and Systems* **10**, pp. 15–20 (1983).

15. de Fériet, J. K. Interpretation of membership functions of fuzzy sets in terms of plausibility and belief. *In* "Fuzzy Information and Decision

Processes" (Gupta, M., and Sanchez, E., Eds.), pp. 93–98. North-Holland Publishing Company, Amsterdam, 1982.

16. Shortliffe, E. H., and Buchanan, B. A model of inexact reasoning in medicine. *Mathematical Biosciences* **23**, pp. 351–379 (1975).

17. Schweizer, B., and Sklar, A. Associative functions and statistical triangle inequalities. *Publ Math Debrecen* **8**, pp. 169–186 (1961).

List of Symbols

$=$	math symbol *equals*
$\approx$	math symbol *approximates*
$<$	math symbol *less than*
$\leq$	math symbol *less than or equal to*
$>$	math symbol *greater than*
$\geq$	math symbol *greater than or equal to*
$\in$	math symbol *element of*
$\subset$	math symbol *subset of*
$\wedge$	math symbol *and*
$\vee$	math symbol *or*
$\mid$	math symbol *vertical slash*
$*$	math symbol *asterisk*
1	arabic number *one*
δ	greek lowercase letter *delta*
e	latin lowercase letter e (Euler's Constant)
η	greek lowercase letter *eta*
γ	greek lowercase letter *gamma*
Γ	greek uppercase letter *Gamma*
l	latin lowercase letter l
I	latin uppercase letter I
μ	greek lowercase letter *mu* (micro)
n	latin lowercase letter n
$\mathbb{R}$	alternate latin uppercase letter R (set of real numbers)
θ	greek lowercase letter *theta*
Θ	greek uppercase letter *Theta*