Predicting Outcome of Defibrillation by Spectral Characterization and Nonparametric Classification of Ventricular Fibrillation in Patients With Out-of-Hospital Cardiac Arrest

Trygve Eftestøl, Sivling; Kjetil Sunde, MD, PhD; Sven Ole Aase, DrIng; John Håkon Husøy, DrIng; Petter Andreas Steen, MD, PhD

Background—In 156 patients with out-of-hospital cardiac arrest of cardiac cause, we analyzed the ability of 4 spectral features of ventricular fibrillation before a total of 868 shocks to discriminate or not between segments that correspond to return of spontaneous circulation (ROSC).

Methods and Results—Centroid frequency, peak power frequency, spectral flatness, and energy were studied. A second decorrelated feature set was generated with the coefficients of the principal component analysis transformation of the original feature set. Each feature set was split into training and testing sets for improved reliability in the evaluation of nonparametric classifiers for each possible feature combination. The combination of centroid frequency and peak power frequency achieved a mean±SD sensitivity of 92±2% and specificity of 27±2% in testing. The highest performing classifier corresponded to the combination of the 2 dominant decorrelated spectral features with sensitivity and specificity equal to 92±2% and 42±1% in testing or a positive predictive value of 0.15 and a negative predictive value of 0.98. Using the highest performing classifier, 328 of 781 shocks not leading to ROSC would have been avoided, whereas 7 of 87 shocks leading to ROSC would not have been administered.

Conclusions—The ECG contained information predictive of shock therapy. This could reduce the delivery of unsuccessful shocks and thereby the duration of unnecessary “hands-off” intervals during cardiopulmonary resuscitation. The low specificity and positive predictive value indicate that other features should be added to improve performance.

Key Words: cardiopulmonary resuscitation ■ fibrillation ■ defibrillation ■ Fourier analysis

Although early defibrillation of ventricular fibrillation (VF) increases the survival rate after cardiac arrest,1,2 a recent study from Cobb et al3 indicates that some patients with VF might have a better chance of return of spontaneous circulation (ROSC) after a period with chest compressions and ventilation before the defibrillation attempt. The cardiopulmonary resuscitation (CPR)-induced myocardial perfusion can cause changes in the power spectrum of the VF with an attendant increase in the ROSC rate.4–6 Futile defibrillation attempts are in themselves detrimental, because tissue damage and postresuscitation myocardial dysfunction may be caused by the shock itself7 and by the lack of tissue perfusion from chest compressions during the shock period (analysis, charging, defibrillation, and outcome evaluation). It would therefore be important to know whether a shock will cause ROSC.

From 20% to 80% of defibrillation attempts in clinical studies are reported to cause the discontinuation of VF8–11 (the great variability depends on different time definitions of VF reoccurrence or different shock waveforms), but we found a frequency of only 10% ROSC after 883 individual shocks in 156 patients.11 This is similar to the results of previous reports.8,9,12 Brown et al12 reported that the combination of centroid frequency (CF) and peak power frequency (PPF) of the VF could predict ROSC with 100% sensitivity and 47.1% specificity. We question the reliability of their results due to both the study design and the small data set, with only 9 incidents of ROSC after 128 shocks in 55 patients. Brown et al12 reported that the combination of centroid frequency (CF) and peak power frequency (PPF) of the VF could predict ROSC with 100% sensitivity and 47.1% specificity. We question the reliability of their results due to both the study design and the small data set, with only 9 incidents of ROSC after 128 shocks in 55 patients. The reliability of their results could have been confirmed if the prognostic criteria had been defined from 1 data set ("training set") and the sensitivity and specificity had been derived from a new...
data set ("testing set") instead of both having been determined from the same data set.

We have therefore attempted to predict defibrillation outcome in human cardiac arrest by combining features of spectral characterization, splitting the data into training and testing sets, and using classifier generalization techniques in an attempt to increase the degree of expected reliability. One of the combinations studied was that reported by Brown et al.12

Methods
In the observational prospective study from Oslo,11 data were collected from the medical control module of the defibrillator (Heartstart 3000; Laerdal Medical) and the regular Utstein registration.12 All patients with out-of-hospital cardiac arrest of cardiac origin13 that occurred between February 19, 1996, and February 18, 1998, were included if the advanced cardiac life support (ACLS) attempt was documented on the medical control module. Approval for the study was obtained through the Regional Committee for Research Ethics, Health Region III (Norway), and the Norwegian Data Inspectorate.

Characterization
The ECG segments before shock were grouped according to shock outcome. Outcome was defined as ROSC if a palpable pulse was present in the postshock period (independent of duration). The remainder of the shocks corresponded to No ROSC, including conversions to electromechanical dissociation, asystole, VF (VF starting >5 seconds after the shock), or nonreset shocks (VF starting <5 seconds after the shock). If the initial postshock rhythm was present for >10% of the duration of the interval, it was defined as the postshock rhythm. Otherwise, the next rhythm was considered. This was done through an automated procedure, which handled all except 15 shocks. These failures were caused by illogics in the annotation structure. Each shock was analyzed as an independent event because we wanted to predict the result of each shock independent of clinical variables that may affect outcome.12 Thus, we made no distinction between shocks with or without prior ACLS in the class division scheme (Table 1). We evaluated whether ACLS affected the results from the ROSC and No ROSC groups, respectively.

The prediction analysis was performed in 2 stages that apply pattern recognition methods.14 First, the ECG was spectrally characterized (feature extraction), and second, decision regions for shock outcome prediction were determined and evaluated.

Feature Extraction
The characterizing features were computed from the estimated power spectral density (PSD) of each ECG segment.4–6,12,15–18

$$P(f) = \frac{1}{L} \sum_{n=0}^{L-1} x(n)e^{-j2\pi fn/L}$$

where f is the frequency and x(n) denotes sample n in an ECG segment of length L.

We attempted to discriminate between preshock ECG segments that correspond to ROSC and No ROSC outcome by computing the following features from the ECG segment PSD estimates.

The CF, or median frequency,4,5,13,14 is given by

$$CF = \frac{\int_{f_l}^{f_u} P(f)df}{\int_{f_l}^{f_u} P(f)df}$$

where f_l and f_u are the lower- and higher-frequency band limits, respectively. By varying these limits, we could study the effect of extracting features from different frequency bands.

PPF is given by

$$PPF = \arg \max_i \{\hat{P}(f)\}$$

The spectral flatness measure (SFM)19 of the VF is given by:

$$SFM = 2 - \frac{\int_{f_l}^{f_u} P(f)df}{\int_{f_l}^{f_u} \hat{P}(f)df}$$

SFM attains a value between 0 (peaky) and 1 (flat).

Various time domain measurements of signal amplitude characteristics of VF have been investigated.6,12,16,20–22 In the present study, we investigated an alternative frequency band–limited energy measurement (ENERG):

$$\text{ENERG} = \int_{f_l}^{f_u} \hat{P}(f)df$$

This enables direct analysis in a specific frequency band.

An alternative decorrelated feature set was generated by principal component analysis (PCA) transformation.14 The features were projected onto the eigenvectors that best represented the entire data set. Thus, the new decorrelated feature set is represented by the magnitudes of the projections along the eigenvectors. Before classification, a combination from either the original or the decorrelated feature set was placed into a feature vector, v.

Classification
In the classifier, each feature vector, v, was considered to belong to 1 of the K classes, o_i, i=1, . . . , K, which corresponded to the shock outcome rhythms (K=5). As shown in Table 1, o_1 and o_5 corresponded to the ROSC and No ROSC groups, respectively. Decision regions for these defined classes from annotated data can be retrospectively calculated with classification theory. The class membership of new data are decided through prospective comparison with these decision regions.

K decision regions, R_k, i=1, . . . , K, are computed by assigning costs for the possible wrong decisions. Each cost, C(o_i, \hat{o}_j), expresses the risk associated with classification of a pattern of the true class o_i as belonging to the decided class \hat{o}_j. A reject class, o_k, is added to handle ambiguous or out-of-range patterns. Each R_k is calculated by selecting the minimum component of the risk vector r=[r_1, r_2, . . . , r_k]^T, where

$$r_k = \sum_{j=1}^{K} C(o_j, \hat{o}_k) P(o_j|v).$$

TABLE 1. Class Division Scheme

<table>
<thead>
<tr>
<th>Class</th>
<th>Subclass</th>
<th>Conversion to</th>
<th>n</th>
<th>Prior ACLS, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROSC</td>
<td>Pulse rhythm</td>
<td>87</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>No-ROSC</td>
<td>Pulseless rhythm</td>
<td>337</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>No-ROSC</td>
<td>Isoelectric rhythm</td>
<td>98</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>No-ROSC</td>
<td>VF</td>
<td>35</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>No-ROSC</td>
<td>Non-reset shocks</td>
<td>311</td>
<td>87</td>
<td></td>
</tr>
</tbody>
</table>

Of the 156 patients, 110 received shock treatment. The 87 shocks in the ROSC class were administered to 46 of the patients. Of these shocks, 35 converted to sustained pulse rhythm, whereas 52 converted to pulse rhythms of limited duration (>20 seconds). Nineteen ROSCs were achieved after the first shock (4 with prior ACLS).
TABLE 2. Tested Combinations of Candidate Features

<table>
<thead>
<tr>
<th>Combination</th>
<th>Spectral</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[v₁]</td>
<td>[v₁]</td>
</tr>
<tr>
<td>2</td>
<td>[v₁ v₂]</td>
<td>[v₁ v₁]</td>
</tr>
<tr>
<td>3</td>
<td>[v₁ v₂]</td>
<td>[v₁ v₁]</td>
</tr>
<tr>
<td>4</td>
<td>[v₁ v₁]</td>
<td>[v₁ v₁]</td>
</tr>
<tr>
<td>5</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>6</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>7</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>8</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>9</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>10</td>
<td>[v₁ v₁ v₃]</td>
<td>[v₁ v₁ v₁]</td>
</tr>
<tr>
<td>11</td>
<td>[v₁ v₁ v₃ v₄]</td>
<td>[v₁ v₁ v₁ v₁]</td>
</tr>
<tr>
<td>12</td>
<td>[v₁ v₁ v₃ v₄]</td>
<td>[v₁ v₁ v₁ v₁]</td>
</tr>
<tr>
<td>13</td>
<td>[v₁ v₁ v₃ v₄]</td>
<td>[v₁ v₁ v₁ v₁]</td>
</tr>
<tr>
<td>14</td>
<td>[v₁ v₁ v₃ v₄]</td>
<td>[v₁ v₁ v₁ v₁]</td>
</tr>
<tr>
<td>15</td>
<td>[v₁ v₁ v₃ v₄]</td>
<td>[v₁ v₁ v₁ v₁]</td>
</tr>
</tbody>
</table>

Shown are vectors for SFM, ENRG, CF, and PPF and their respective decorrelated features by PCA transformation.

This corresponds to minimization of the expectation of the classifier risk,

\[ P_{\text{err}} = \frac{1}{K} \sum_{j=1}^{K} \left( 1 - P(R_j | \omega) \right) \]  

where \( P(R_j | \omega) \) is the probability of false alarm for class \( \omega \). 

The classifier performance characteristics are expressed by the sensitivity (probability of positive prediction of ROSC outcome) and specificity (probability of negative prediction of No ROSC outcome) given by

\[ P_{\text{sns}}(\omega_i) = P(R_i | \omega_i) \]

and

\[ P_{\text{spc}}(\omega_i) = \frac{1}{1 - P(\omega_i) P(R_i | \omega_i)} \]

respectively. \( P(R_i | \omega_j) \) expresses the proportion of true class \( \omega_i \) with the corresponding decision being \( \omega_j \).

The decision regions were calculated iteratively with minimization of the object function

\[ J = \sum_{i} \left( P_{\text{sns}}(\omega_i) - P_{\text{sns}}(\omega_i) \right)^2 \]

so that the classifier would meet the desired performance criterion given by \( P_{\text{sns}}(\omega_i) \). This is done by multiplying the costs, \( C(\omega_i, \omega_j) \), \( j \neq i \), by a factor \( \alpha \). By setting \( i = 1 \) (\( \omega_i \) corresponding to ROSC), this allowed specification of a sensitivity for the recognition of ROSC outcome.

The underlying statistics can be estimated with classification theory.

Multidimensional histograms were applied in which the feature space is divided into bins of equal volume, in which the PDF estimates are computed. Each feature set is normalized by dividing by the respective feature axis into \( n_b \) equal-sized intervals in the range from the minimum to the maximum feature value. The PDF estimates in each histogram bin are then distributed by applying an elliptic gaussian kernel function, resulting in a smoother continuous estimate.

Histogram bin resolution and kernel width are the 2 key parameters of the classifier. A small number of large bins provides low histogram resolution, whereas a large number of small bins provide high resolution. Each feature axis of the PDF is divided into \( n_b \) intervals. Thus, if the feature dimension is \( D \) for a specific feature combination, the feature space is divided into \( n_b^D \) bins of equal volume. Smoothness is governed by the width of the kernel function. A narrow kernel function provides a high-resolution estimate with high variance, whereas a wide kernel function provides a smoother low-resolution estimate with low variance.

The concept of generality is important in the design of classifiers. The decision regions are calculated with a training set of feature

---

Figure 1. Effects of generalization. Training performances are shown in (a) specificity increase with number of bins and decrease with increased kernel width and (b) all classifiers are trained to 95% sensitivity. Corresponding testing performances are shown in (c) the specificities correspond roughly with that of training and (d) the specificities do not correspond at high bin numbers and low kernel widths. Generality is achieved at low bin numbers and high kernel widths.
vectors that represent the experience on which the classifier will base future decisions. Testing is done on an independent set. In a well-designed classifier, the testing performance should approach the training performance. Both the histogram bin resolution and the kernel width applied in the estimation affect generality.

Training and testing were conducted with a cross-validation technique. In each of S consecutive experiments, an \((S^2 - 1)/S\) portion of the entire data set is used to train classifier number \(i\). The remaining \(1/S\) portion is kept out for testing. \(i\) is varied from 1 to \(S\), thus producing \(S\) classifier performance results.

Experimental Setup

The ECG was sampled at 100 Hz with 8-bit resolution, and PSD was estimated from segment lengths \(L = 400\) zero padded to 512 samples.

Three feature sets were extracted with frequency ranges \((f_l - f_u)\) of 0 to 50 Hz, 0 to 25 Hz, and 0 to 12.5 Hz. The spectral features produced in each of these experiments were \(v_{SFM}, v_{ENRG}, v_{CF},\) and \(v_{PPF}\). The PCA transformation of these features gave the corresponding decorrelated feature set of \(v_{PCA1}, v_{PCA2}, v_{PCA3},\) and \(v_{PCA4}\). The ECG immediately before defibrillation was analyzed, and the measurements were grouped according to the postshock rhythm for classifier design (Table 1).

Classifiers were designed and tested with the use of all possible combinations of spectral features and decorrelated features (Table 2). The statistical functions were estimated with multidimensional histograms. Resolutions were adjusted according to setting \(n_b\) equal to 4, 8, 16, 32, 64, and 128 bins. For each of these resolutions, the smoothness was varied by setting the variance of the gaussian kernel function, \(k_w\), equal to 0, 1, 5, 10, 15, and 20.

![Figure 2. Test performances for generalized classifiers for all feature combinations and frequency ranges. Filled columns show sensitivities, and open white columns show specificities. Feature combination numbers correspond to those given in Table 2. Performance of different frequency ranges is \(a = 0\) to 50 Hz, \(c = 0\) to 25 Hz, and \(e = 0\) to 12.5 Hz for combinations of original spectral feature sets and \(b = 0\) to 50 Hz, \(d = 0\) to 25 Hz, and \(f = 0\) to 12.5 Hz for combinations of decorrelated spectral feature sets.](http://circ.ahajournals.org/)

<table>
<thead>
<tr>
<th>TABLE 3. Distribution of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Shown are SFM, ENRG, CF, and PPF and their respective decorrelated features by PCA transformation \((y_{ex} = y_{10x})\).
TABLE 4. Performance Characteristics

<table>
<thead>
<tr>
<th>Feature Combination</th>
<th>Frequency Range, Hz</th>
<th>$\hat{P}_{ss}$ (o), %</th>
<th>$\hat{P}_{sp}$ (o), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\nu_{CF}, \nu_{PPF}]$</td>
<td>0–12.5</td>
<td>92±2</td>
<td>27±2</td>
</tr>
<tr>
<td>$[\nu_{PCA}, \nu_{PCA}]$</td>
<td>0–25</td>
<td>92±2</td>
<td>42±1</td>
</tr>
</tbody>
</table>

The sensitivity and specificity for predicting ROSC for the reference classifier$^{12}$: CF and PPF at 0 to 12.5 Hz and the highest performing classifier based on PCA decorrelation at 0 to 25 Hz.

This combination of changing bin size resolution and smoothness enabled a search for the classifier that met the generality criterion, which we defined to be that the test sensitivities and specificities should approach the training sensitivities and specificities to within a 5% tolerance range. In Figure 1, the training and testing specificities and sensitivities are shown as functions of resolution and kernel width. Training with high bin resolution and narrow kernel width generates a classifier with 100% performance in both sensitivity and specificity as the result of overtraining, as verified by the large deviation in sensitivity in test performance. Generality in sensitivity is achieved either by increasing the kernel width or by using lower bin resolution, with both resulting in lower specificity.

A full-scale evaluation with respect to generality of the classifiers corresponding to all possible feature combinations (Table 2) was performed. Finally, for a given feature combination, the classifier with the best general performance was defined as that corresponding to the highest average test performance, lowest bin resolution, and narrowest kernel width, requiring that the training and test performances satisfied the generality criterion.

Statistical Analysis
The comparisons between ROSC and No ROSC and between ACLS and No ACLS were tested with the Wilcoxon rank sum test and presented as median values (25th and 75th percentiles). $P<0.05$ was regarded as statistically significant. Classifier performance results are presented as the mean±SD of the cross-validated sensitivities and specificities.

Results
The results for the parameters are summarized in Table 3. The test performance results of the classifiers that met the generality criterion are shown in Figure 2.

![Figure 3. Estimated class-specific probability density functions and corresponding decision regions. Subclasses are identified by color scheme of yellow=ROSC, blue=electromechanical dissociation, red=asystole, and green=VF reappearing >5 seconds after the shock; cyan=VF reappearing within 5 seconds after the shock (nonreset shocks); and white=reject. Highest performing classifier: (a) probability densities, (b) decision regions. Reference classifier: (c) probability densities, (d) decision regions.](image-url)
myocardial dysfunction. Moreover, because the spectral characteristics of the VF have been reported to reflect myocardial perfusion, the defibrillator also might guide the CPR attempt, because the myocardial perfusion depends on compression force, rate, and duration.

On the other hand, 7 shocks that resulted in a pulse-giving rhythm would not have been administered. These shocks presumably would have been administered later if CPR changed the characteristics of the VF. The effects of this could not be evaluated. The comparison of ACLS with No ACLS features illustrates this aspect of use of the features for online monitoring of the CPR efficiency. The use of the features as monitoring parameters for performance feedback during CPR is an interesting idea that is closely related to the prediction problem. Retrospectively, we studied the influence of ACLS on a single feature and demonstrated changes in values according to treatment. The present study demonstrates how a general classifier can be designed through cross-validation, which allows training and testing on independent data sets in combination with different resolutions and kernel widths in the estimation of the statistics that describe the features. This method gives an indication of how well the classifier will perform when challenged with new data in the future.

In a similar study of 128 shocks in 55 patients with only 9 successful shocks (defined as a conversion of VF to a supraventricular rhythm with a palpable pulse or blood pressure of any duration within 2 minutes of the shock without ongoing CPR), Brown et al extracted 4 parameters from the recorded ECG. The combination of CF and PPF gave the best predictive potential (sensitivity 100%, specificity 47.1%). The same combination of features gave a poorer predictive potential in the present study (sensitivity 92±2%, specificity 27±2%). We believe that the results of our generalized classifier are more realistic due to the larger database and the use of independent testing and generalization that were not done by Brown et al. Those authors generated the sensitivity and specificity with the same data from which the threshold values were computed, with no independent evaluation.

Noc et al reported in pigs that maximum and mean VF amplitude and dominant VF frequency were all acceptable shock outcome predictors. They derived the threshold values from 1 group and tested these in a separate validation group but had different results in the 2 groups, indicating that the results might not be reliable. Our results indicate that Brown et al would have experienced the same if their threshold values had been tested on new data.

Our method includes independent testing and generalization to avoid these problems. To ensure reliability, the data were split in 2. Training performance for ROSC and No ROSC prediction was computed from half of the data, whereas the other half was used to compute the corresponding test performance.

There are some limitations in the present study. First, the number of ROSC observations is low. Second, in the cross-validation processing of the data, the test performances were considered in the design of the classifiers to choose the generalizing parameters. Ideally, a final evaluation should
have been performed on yet another data set that did not influence the design process. Third, we used only 1 type of classifier: the histogram method. To obtain even more reliable results, the experiments should be repeated with other types of classifiers.

Spectral characterization of VF can be of clinical importance if it can be incorporated into the software of defibrillators. We have demonstrated a method to develop an outcome predictor for defibrillation attempts in out-of-hospital cardiac arrest patients, although the sensitivity of 92±2% and specificity of 42±1% are not satisfactory for clinical use. Therefore, other features should also be investigated to add discriminative power to the feature set.

Acknowledgments
This work was supported in part by the Research Council of Norway, the Norwegian Air Ambulance, the Laerdal Foundation for Acute Medicine, and Anders Jahre’s Foundation.

References
Predicting Outcome of Defibrillation by Spectral Characterization and Nonparametric Classification of Ventricular Fibrillation in Patients With Out-of-Hospital Cardiac Arrest
Trygve Eftestøl, Kjetil Sunde, Sven Ole Aase, John Håkon Husøy and Petter Andreas Steen

_Circulation_. 2000;102:1523-1529
doi: 10.1161/01.CIR.102.13.1523

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2000 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/102/13/1523

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/