33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Reconstructing ECG Precordial Leads from a Reduced Lead Set using Independent Component Analysis

Michael H. Ostertag and Gill R. Tsouri

Abstract—**In this paper, precordial lead reconstruction from a reduced set of leads is considered. We propose the use of independent component analysis to train patient-specific transforms from a reduced lead set to the six precordial leads of the standard 12-lead electrocardiogram. The proposed approach is applied to a publicly available database comprising 549 ECG recordings of patients with varying cardiovascular conditions. The fidelity of reconstruction is measured using percent correlation between the actual and reconstructed signals following a 30 seconds time lapse. The mean correlation is over 95% with a standard deviation under 12.7% for all reconstructed leads. The results demonstrate the potential of the suggested approach to provide a reliable solution to precordial leads reconstruction.**

I. Introduction

THE *ElectroCardioGram* (ECG) is the most common procedure for diagnosing cardiovascular problems and a critical tool for long-term monitoring of patients [1]. While most physicians prefer the diagnostic capabilities of traditional 12-lead systems, it is commonly not fully implemented for patient comfort and caregiver convenience.

Past and recent work focused on developing ECG systems that maintain the same diagnostic ability of 12-lead ECG with a reduced number of leads. See [2-4] for examples. These systems reconstruct the missing leads from a set of basis leads using linear transforms. A set of universal coefficients for this transform would have been an ideal solution but it was proven to be an inaccurate approach because the coefficients are dependent upon numerous biological and environmental factors [5-6]. Population-based and patient-specific coefficients produce much better results than universal coefficients. Patient-specific coefficients perform better than population-based coefficients at the beginning of observation, and as the time of observation progresses, their performance becomes the same [7]. These results suggest that patient-specific coefficients could provide better performance. In this paper we investigate the use of *Independent Component Analysis* (ICA) to provide patient-specific coefficients for long term observations and lead reconstruction.

ICA is used to represent a signal by a set of statistically *Independent Sources* (ICs). These sources (s) can be linearly combined to recreate the original observations (x) through a mixing matrix (A) as seen below in (1):

$$x = As. \tag{1}$$

The mixing matrix is square in most popular ICA algorithms in order to exploit the inverse matrix to shorten the time of convergence. ICA is very useful for biomedical applications because it can separate desired signals from unwanted signals, even if they are temporally or spatially correlated. It can be applied to biomedical signals so long as we assume that linear mixing of the sources occurs, which implies that the signal travels through the mixing medium and reaches all sensors simultaneously [8]. ICA has been used in separating maternal and fetal ECGs, filtering interference out of EEGs, and removing noise and artifacts from ECG signals with varying degrees of success [8-10]. Other blind source separation methods, such as *Second Order Blind Source Separation* (SOBI), exist but have not been used in ECG processing with as much frequency as ICA.

Of the 10 electrodes used in a 12-lead system, the most problematic are the unipolar leads across the precordium, of which leads V3 and V4 can complicate diagnostic procedures, such as echocardiograms and chest x-rays, and the life-saving procedure of defibrillation [6]. Since the precordial leads have substantial redundant information due to proximity and since their sources can be approximated by a dipole, they are prime candidates for reduction by ICA.

In this contribution, we propose and investigate reconstructing precordial leads in a patient-specific approach from their underlying sources using ICA. We are unaware of previous attempts to use ICA for ECG lead reconstruction.

The rest of the paper consists of an overview of the Proposed Method in Section II, a summary of the Results in Section III when applying the approach to a well-known and publicly available ECG database, a Discussion in Section IV, and theConclusion in Section V.

II. Proposed Method

The proposed precordial lead reconstruction follows a multi-stage approach. First, preprocessing is performed on the ECG signal to condition it and to locate the QRS complexes. Next, a training sequence develops a set of patient-specific transforms from one pair of precordial leads to the others. Then, the excess electrodes may be removed and the algorithm will continue to reconstruct the missing leads. In what follows, each step is described in more detail.

Manuscript submitted April 15, 2011. This work was supported by Blue Highway (A Welch-Allyn Company).

Michael H. Ostertag is a M. Sc. candidate at Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: mho7745@ rit.edu).

Gill R. Tsouri is Director of the Communications Research Lab at Rochester Institute of Technology, Rochester, NY 14623 USA. (phone: 585-475-6452 e-mail: grteee@ rit.edu).

A. ECG Dataset and Preprocessing

The work presented in this paper used the *Physikalisch-Technische Bundesanstalt* (PTB) diagnostic ECG database as made available through PhysioNet [11]. The database contains 549 ECG recordings from 290 subjects with a variety of diagnostic classes: 52 as healthy controls, 148 as myocardial infarctions, and the remainder with other cardiac diagnoses. For each recording, the 12 traditional and three Frank leads were captured simultaneously at a sampling frequency of 1 kHz.

The preprocessing stage encompasses two steps: filtering and beat detection. A set of cascading digital filters were used. The first was a high-pass filter with a cutoff frequency (f_c) of 0.5 Hz to remove the baseline drift. The second was a low-pass filter with an $f_c$of 150 Hz to reduce the amount of noise in the signal. Both filters were developed to meet AHA standards (outlined in [1]) using the Parks-McClellan algorithm. The filtered signal was then put through a QRS detection algorithm similar to the Pan-Tompkins QRS detection algorithm [12]. A moving average of 100 samples (100 ms) was taken of the square of the approximate derivative as obtained with the function seen below:

$$y[n] = -x[n-10] - 2x[n-5] + 2x[n+5] + x[n+10]. \quad (2)$$

The resulting function was normalized and the peaks were found that had a value above the threshold of 0.125 and that occurred at least 200 ms from the previous peak. This length of time was selected because it is the minimum time between beats due to physiological limitations. Once the peaks were found, the beat domain was defined as spanning three-eighths the time between the current and previous peak and five-eighths the time between the current and next peak.

B. Transform Training Sequence

Leads V2 and V5 were used to reconstruct the other precordial leads due to their preferred use in past lead reconstruction work [3,13-14] and because we observed high levels of correlation between the ICs of various sets of precordial leads with those of V2 and V5. Using ICA over the two leads generated two ICs, which represent the underlying sources as a dipole.

For training, we made use of all six precordial leads. The first valid beat was used as a training sequence to determine a transform between the ICs of leads V2 and V5 and the missing precordial leads V1, V3, V4, and V6. First, each lead had the mean removed and was normalized to unit variance. Then, the 4-by-2 mixing matrix was generated to reconstruct the four missing leads from the pair of ICs generated from leads V2 and V5. It was obtained by performing FastICA [9] on several sets of lead combinations that had previously shown to produce ICs that were highly correlated to those of leads V2 and V5. As part of the ICA solving procedure, mixing matrices are generated that relate the observations and ICs as seen in (1). The coefficients that make up the mixing matrices of (V1,V5), (V3,V5), (V2,V4), and (V2,V6) were used to create a new mixing matrix that could reconstruct V1, V3, V4, and V6, respectively. For example, the application of ICA to leads V1 and V5 resulted in the square mixing matrix as seen in (3). The elements relating to V1 were placed into a new 4-by-2 mixing matrix, which can reconstruct all missing leads from a set of ICs (4).

$$\begin{bmatrix} x_{V1} \\ x_{V5} \end{bmatrix} = \begin{bmatrix} a_{V1} & b_{V1} \\ a_{V5} & b_{V5} \end{bmatrix} \begin{bmatrix} IC_1 \\ IC_2 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x_{V1} \\ x_{V3} \\ x_{V4} \\ x_{V6} \end{bmatrix} = \begin{bmatrix} a_{V1} & b_{V1} \\ a_{V3} & b_{V3} \\ a_{V4} & b_{V4} \\ a_{V6} & b_{V6} \end{bmatrix} \begin{bmatrix} IC_1 \\ IC_2 \end{bmatrix} \quad (4)$$

Due to the non-ordered and sign-independent nature of ICA, the leads were sorted based upon the correlations of each lead combination's set of ICs with the (V2,V5) IC pair. The sorted reconstruction mixing matrix, the original ICs of V2 and V5, and the transform matrix between the ICs and (V2,V5) were saved for future use.

C. Reconstruction Sequence

For reconstruction, we used leads V2 and V5 to reconstruct all other precordial leads. Each beat after the training sequence was handled in the following manner. The mean of the beat was removed and the variance normalized. Then, ICA was performed on V2 and V5, but instead of using a random initial mixing matrix for the iterative solving procedure in the FastICA algorithm, the original transform matrix was provided as an initial guess. This helped direct the solution to a similar set of ICs and shortened the time of convergence.

Even though an initial guess was provided, the FastICA algorithm occasionally converged to a switched or negative IC pair, which was sorted in the following manner. The resulting IC pair and the original ICs that were generated during training were downsampled by a factor of 5 and compared. A correlation function was formed that was the sum of the absolute value of the correlations for the two possible configurations of ICs:

$$\underset{\text{Config. 1}}{\begin{bmatrix} |IC_1(0)| \\ |IC_2(0)| \end{bmatrix} \approx \begin{bmatrix} |IC_1(t)| \\ |IC_2(t)| \end{bmatrix}} \quad \underset{\text{Config. 2}}{\begin{bmatrix} |IC_1(0)| \\ |IC_2(0)| \end{bmatrix} \approx \begin{bmatrix} |IC_2(t)| \\ |IC_1(t)| \end{bmatrix}} \quad (5)$$

Since the ICs were expected to match each other, configuration 1 was given a preferential weight of 1.25. The two functions were compared by the above metric and the set that had a higher maximum correlation was selected. The index of the maximum and individual correlation functions were used to determine if inverting the sign of the IC was required. Using the sorted ICs and the sorted reconstruction mixing matrix from the training sequence, V1, V3, V4, and V6 were reconstructed from the ICs of (V2,V5) using (4).

D. Method of Comparison

In order to compare the actual lead signals to the reconstructed signals, a percent correlation (ρ) was utilized as the figure of merit, which resembles the similarity

coefficient used in [13]. The percent correlation metric was calculated per lead in the following manner:

$$\rho = \frac{\sum_{i=1}^{N} X_i \tilde{X}_i}{\sqrt{\sum_{i=1}^{N} X_i^2}\sqrt{\sum_{i=1}^{N} \tilde{X}_i^2}} \times 100\% \qquad (6)$$

where the index value i represents the samples within a beat of length N samples, X_i is the original ECG signal, and $\tilde{X}_i$ is the reconstructed ECG signal. A value of 100% represents perfect correlation between the two signals while lower values indicate a worse timing alignment. Correlation was used instead of other measures such as *Root-Mean-Square Error* (RMSE) because correlation speaks to how well the timing of the signals is aligned rather than the absolute value. In many applications of ECG, these timing features are much more critical than absolute values.

The original and reconstructed waveforms were compared at several time instances after the training sequence and the percent correlation values were plotted on a histogram. This allowed for a direct, visual comparison of the changes in the fidelity of reconstruction over time.

III. Results

We ran the precordial lead reconstruction algorithm over 548 of the 549 recordings in the PTB database from the beat immediately following the training sequence (t=0 sec) to the beat that occurred 30 seconds (t=30 sec) after the training sequence. One patient (s0377) was not used because the V1 lead was removed mid-recording. The mean and variance of the distributions can be found in Table I.

TABLE I
Statistics of correlations between actual and reconstructed leads

t		V1	V3	V4	V6
0sec	μ (σ)	91.7 (12.4)	97.5 (4.8)	95.8 (7.6)	96.8 (4.7)
30sec	μ (σ)	91.5 (12.6)	97.0 (7.4)	95.2 (8.9)	96.5 (5.0)

All leads were reconstructed with a high average correlation percentage and low variance. Leads V3 and V6 had the best reconstructions on average (all above 96.4% for both 0 and 30 sec), followed by lead V4 (above 95.2% for both 0 and 30 sec). Lead V1 was reconstructed with the lowest average correlation percentage (over 91.5% for both 0 and 30 sec) and the largest standard deviation (under 12.7% for both 0 and 30 sec). This was expected as problems in the atrium have a stronger effect on the signal recorded by lead V1 than the other precordial leads due to V1's proximity to that region of the heart [6]. V1 was the most problematic to construct in past works as well [6,14].

We found that the algorithm had difficulty accurately reconstructing irregular beats that were not present in the training sequence. This was due to either a change in ICs or an error in the sorting process and was most common in patients with dysrhythmia. When the irregular beats were reconstructed, they looked differently than the typical heartbeat of the patient but did not necessarily have the exact shape of the actual irregular beat.

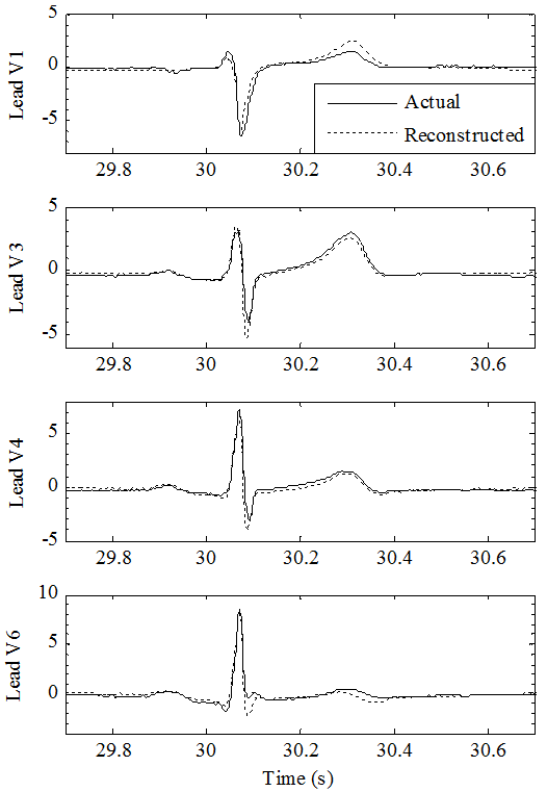

Fig. 1. Comparisons of actual precordial leads and those reconstructed from the ICs of leads V2 and V5. This patient (s0055) had the closest correlation values across all precordial leads to the mean of the database at 30 seconds after the training period

Fig. 1 presents an illustrative case of reconstruction of a patient 30 seconds after the training sequence occurred. For presenting an unbiased sample, we chose to present a pulse with reconstruction quality that most closely matches the mean correlation percentages, rather than the best reconstructed pulse. Due to the locations of V2 and V5, the interpolated leads, V3 and V4, were reconstructed the best and the extrapolated leads, V1 and V6, had larger reconstruction error [6]. If the patient had been diagnosed with a problem in the atrium rather than an anterior myocardial infarction, the reconstruction of lead V1 would have been much worse.

IV. Discussion

The algorithm was tested up to 30 seconds after the training period, but even in that short amount of time, we were able to observe the adaptive nature of ICA. The reconstruction mixing matrix from the ICs to the precordial leads remained constant but the transform from (V2,V5) to their ICs changed with each beat. This feature results in a time-adapting and patient-specific transform from (V2,V5) to the other precordials that can effectively adapt to changing beat patterns. This is a clear advantage over static linear transformations.

Although the precordial reconstruction worked on the majority of patients with a high level of fidelity, the system

is susceptible to errors when patients have atrial conditions, which are most strongly detected in lead V1 [6]. Since the reconstruction is based on the signals recorded at V2 and V5, an issue that is present in lead V1 may not be accurately reconstructed. On the other hand, an error in lead V2 or V5 will cause an error to be propagated across all reconstructed leads. For patients who have displayed signs of dysrhyhmias or have atrial conditions, it is suggested that lead V1 is not removed from the patient since it cannot be reliably reconstructed and is important for accurate diagnosis. This can be checked in the initial training phase when all six precordial leads are attached. The electrodes for lead V1, V3, V4, and V6 are removed after the transforms have been developed in the training phase.

Due to a lack in consistent methods and metrics in previously published works, our reconstruction results are difficult to compare to other studies. Nelwan et al. [3] performed reconstruction of four missing precordial leads from V2 and V5 and found a median correlation of 0.964 for a general set of coefficients and 0.994 for a patient-specific set. These numbers are high for several reasons: first, they present the median correlation instead of the average, reducing the effect of outliers that were present in our results; second, the ECG waveforms that were used were the median complexes about a certain time; and third, all patients were of the same diagnostic class and the most difficult to reconstruct, patients with arrhymias and left bundle branch blocks, were excluded from the reconstruction test. In other work, the improved EASI coefficients as derived by [4] were used to reconstruct the precordials with varying degrees of success. The average correlations across the precordials with the improved coefficients range from 0.919 to 0.941, which fall below our calculated average correlation of 0.950. Also, our correlation calculations do not include leads V2 or V5, which would have perfect correlation because they did not need to be reconstructed.

Inaccuracies in our proposed reconstruction can be reduced by using multiple training sequences rather than the single pulse used in this work. Elaborated training is an issue of further investigation.

The spatial and temporal stationarity of the ICs for the precordials is to be explored in more depth. Body surface potential maps will be used to explore the effect of electrode misplacement on the ICs, and long-term precordial ECG recordings will enable us to view how the sources change over longer times.

We performed our analysis on an offline database. When ICA is to be performed online, computational complexity should be considered as well. Considering available computational power on a host computer for processing the ECG signals and the fact that ICA has to be performed once for every heartbeat, we do not identify a practical limit on performing ICA online for lead reconstruction.

Currently, the algorithm works by performing ICA in a blind source separation approach, which is moderately inefficient because in reality we know a substantial amount about the sources. Another path of work for the future will focus on applying a priori knowledge of the heart and chest.

V. Conclusion

In this work we presented a novel ICA-based patient-specific method for reconstructing precordial leads from the ICs of V2 and V5. The reduced number of leads increases patient comfort and accessibility to the chest for diagnostic procedures, and unlike some other reduced lead systems, it utilizes electrode placements that caregivers already know.

Our results show that patient-specific transform generation using ICA is an improvement over the static linear transforms that are currently under research. It possesses the ability to reconstruct all of the precordial leads from leads V2 and V5 and can adapt its transforms over time.

References

[1] P. Kligfield et al., "Recommendations for the standardization and interpretation of the electrocardiogram, Part I: The electrocardiogram and its technology," *Heart Rhythm*, vol. 3, no. 3, pp. 394-412, March 2007.

[2] G. E. Dower, H. B. Machado, and J. A. Osborne, "On Deriving the Electrocardiogram from Vectorcardiographic Leads," *Clin Cardiology*, vol. 3, no. 2, pp. 87-95, 1980.

[3] S. P. Newlan, J. A. Kors, S. H. Meij, J. H. van Bemmel, and M. L. Simoons, "Reconstruction of the 12-lead Electrocardiogram From Reduced Lead Sets," *Journal of Electrocardiology*, vol. 37, no. 1, 2004.

[4] D. Q. Feild, C. L. Feldman, and B. M. Horáček, "Improved EASI Coefficients: Their Derivation, Values, and Performance," *Journal of Electrocardiology*, vol. 35, suppl. pp. 23-33, 2002.

[5] R. J. A. Schijvenaars, "Intra-individual Variability of the Electrocardiogram: Assessment and exploitation in computerized ECG analysis," Ph.D. dissertation, Erasmus University, Rotterdam, Netherlands.

[6] D. Q. Feild, S. H. Zhou, E. D. Helfenbein, R. E. Gregg, and J. M Lindauer, "Technical challenges and future directions in lead reconstruction for reduced-lead systems," *Journal of Electrocardiology*, vol. 41, pp. 466-473, 2008.

[7] R.E. Gregg, S. H. Zhou, J. M. Lindauer, E. D. Helfenbein, and D. Q. Feild, "Limitations on the Re-Use of Patient Specific Coefficients for 12-lead ECG Reconstruction," *Comp. in Cardiology*, vol. 35, pp. 23-33, 2002.

[8] C. J. James and C. W. Hesse, "Independent Component Analysis for Biomedical Signals," *Physiological Measurement*, vol.26, pp. R15-R39, 2005.

[9] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 2000.

[10] T. He, G. Clifford, and L. Tarassenko, "Application of independent component analysis in removing artefacts from the electrocardiogram," *Neural Comput. And Applic*, vol. 15, pp. 105-116, 2006.

[11] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, June 2000.

[12] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230-236, March 1985.

[13] B. M. Horáček, J. W. Warren and J. J. Wang, "On designing and testing transformations for derivation of standard 12-lead/18-lead electrocardiograms and vectorcardiograms from reduced lead sets of predictor leads," *Journal of Electrocardiology*, vol. 41, pp. 220-229, January 2008.

[14] R. E. Gregg, S. H. Zhou, J. M. Lindauer, D. Q. Feild, and E. D. Helfenbein, "Where do derived precordial leads fail?" *Journal of Electrocardiology*, vol. 41, pp. 546-552, July 2008.