Extended Multiple Linear Regression in the Derivation of Electrocardiographic Leads

D Guldenring, DD Finlay, CD Nugent, MP Donnelly

University of Ulster, Belfast, United Kingdom

Abstract

In this study we investigate the performance of an approach for deriving electrocardiographic leads with the aim of improving derivation accuracy. We focus our attention on a limited lead system that uses leads I, II, V2 and V5 to derive the remaining precordial leads.

Our extended multiple linear regression based lead transformation (EMLRLT) approach extends the standard multiple linear regression based lead transformation (MLRLT) approach by combining the data from the recorded leads with quadratic and cross product terms from the same leads.

It was found that all missing leads were more accurately derived using an EMLRLT approach in comparison with the MLRLT approach. Using the standard MLRLT approach, the median RMSEs for the QRST were found to be 44.2μV, 42.7μV, 40.3μV and 19.3μV for leads V1, V3, V4 and V6, respectively. Using the EMLRLT approach, the median RMSEs for the QRST were found to be 28.2μV, 29.3μV, 25.1μV and 13.4μV for leads V1, V3, V4 and V6, respectively. According to the sign test, all differences were statistically significant with $p<0.05$.

In conclusion, it has been shown that alternative methods for lead transformation have the potential to improve derivation accuracy.

1. Introduction

Limited lead systems use fewer recording electrodes than standard approaches, however, they aim to provide comparable information about the heart´s electrical activity. This is achieved through lead transformations, in which, an estimate of a non recorded lead, is typically calculated or derived by a weighted sum of all recorded leads (Equation 1.).

$$dLead_i = \sum_{n=1}^{k} a_{i,n} \cdot R_n \qquad (1)$$

The weights $a_{i,n}$ in Equation 1, are applied to the recorded leads (R_n), in order to calculate derived lead number i ($dLead_i$).

In recent years, research for these systems has primarily focused on the impact of the number and spatial location of recording sites on the diagnostic performance. Very little research has been conducted on the actual lead transformation methods themselves.

In this study, we propose and assess an extended multiple linear regression based lead transformation (EMLRLT) technique, that extends the most commonly applied multiple linear regression based lead transformations (MLRLT) [1] through utilization of quadratic and cross-product terms.

2. Materials and methods

This section details the composition of the study population and outlines the methodology employed in the assessment of the derivation accuracy.

2.1. Study population

For this study, we used a subset of the "PTB Diagnostic ECG Database" [2], which is publicly available from Physionet [3]. This Database contains 15-lead electrocardiograms (ECGs). This includes 12 standard leads together with the Frank XYZ leads, each recorded with a sampling rate of 1000 samples per second and an amplitude quantization resolution of 0.5μV per least significant bit (LSB). These ECGs were obtained from a population of 294 subjects, including healthy volunteers and subjects with a number of different heart diseases. A clinical summary for 268 of these subjects is also provided [3].

Only ECGs from subjects where the clinical summary was also available were included in the study population. Several subjects had multiple ECGs recorded, however, in this study only one ECG recording per subject was considered. Subjects were excluded from the study population in instances where recordings had less than 100 QRST complexes. This minimum of 100 QRST complexes was chosen in order to ensure a sufficient amount of data for transformation design and evaluation.

Based on this selection criteria, a study population consisting of 180 ECGs was established comprised of 131 male and 49 female subjects, with an average age of 59 (range: 22-81) and 53 (range: 17-86) respectively. A composition (sex, medical diagnosis) of this study population was extracted from the clinical summary of each of the 180 ECGs utilized and is presented in Table 1.

Table 1. Study population composition (number of female / male subjects in each diagnostic class)

Diagnostic class	#female	#male
Bundle branch block	1	5
Cardiomyopathy	4	2
Dysrhythmia	0	1
Healthy control	12	31
Hypertrophy	1	3
Myocardial infarction	30	84
Myocarditis	0	3
Stable angina	0	1
Valvular heart diseases	1	1
Total	49	131

2.2. Methods

The amplitude quantization resolution of the ECGs within the study population was re-sampled from 0.5μV/LSB to a more common resolution of 5μV/LSB. This amplitude quantization resolution was recommended by the CSE Working Party in [4]. In accordance to [5], each record was (0.05–150)Hz band pass and 50Hz notch filtered. This was followed by an automatic identification of QRS onset, QRS offset and T end using ECGPUWAVE [6]. ECGPUWAVE is a QRS detector and waveform limit locator and is provided as part of the PhysioToolkit [3]. Its waveform limit locator is based on an algorithm, which is described in [7] and evaluated in [7,8]. The accuracy of this algorithm in identification of fiducial points in single-lead records was found to be robust enough to be comparable to experts [8]. A 12-lead "QRSTrecord", composed of 100 consecutive QRST complexes, was subsequently assembled for each of the original 180 ECGs. These QRSTrecords were then used for assessment of the patient-specific lead derivation accuracy.

During this assessment, MLRLT and EMLRLT were used to derive precordial leads V1, V3, V4 and V6 of a QRSTrecord from leads I, II, V2 and V5 of the same record. This particular limited lead set [9] was chosen, since it is one of the most widely reported commercial limited lead systems.

The derivation accuracy of both transformations was assessed via twofold cross-validation, in order to account for potential over-fitting. Each QRSTrecord was therefore divided into two partitions, each of which was 50 QRST complexes in length. Patient-specific weights $a_{i,n}$ of MLRLT (Equation 2.) and EMLRLT (Equation 3.) were calculated from the first 50 QRST complexes of a QRSTrecord via multiple linear regression analysis.

In both, MLRLT and EMLRLT, the patient-specific weights $a_{i,n}$ are applied to the recorded leads (I, II, V2 and V5 and their combinations) in order to calculate derived lead number i ($dLead_i$).

$$dLead_i = a_{i,1} \cdot I + a_{i,2} \cdot II + a_{i,3} \cdot V2 + a_{i,4} \cdot V5 \quad (2)$$

$$\begin{aligned} dLead_i = {} & a_{i,1} \cdot I + a_{i,2} \cdot II + a_{i,3} \cdot V2 + a_{i,4} \cdot V5 \\ & + a_{i,5} \cdot I^2 + a_{i,6} \cdot II^2 + a_{i,7} \cdot V2^2 + a_{i,8} \cdot V5^2 \\ & + a_{i,9} \cdot I \cdot II + a_{i,10} \cdot I \cdot V2 + a_{i,11} \cdot I \cdot V5 \\ & + a_{i,12} \cdot II \cdot V2 + a_{i,13} \cdot II \cdot V5 \\ & + a_{i,14} \cdot V2 \cdot V5 \end{aligned} \quad (3)$$

Following the calculation of the weights $a_{i,n}$, missing precordial leads dV1, dV3, dV4 and dV6 were derived by application of MLRLT and EMLRLT to the remaining 50 QRST complexes.

Similarity between derived complexes and those actually measured was used to assess the derivation accuracy of MLRLT and EMLRLT. For consistency with other studies [9-11], we used the root mean squared error (RMSE) to assess the derivation accuracy of MLRLT and EMLRLT during different waves of the ECG complex. The RMSE was thereby calculated for the QRS complex ($RMSE_{QRS}$), STT segment ($RMSE_{STT}$) and the entire QRST ($RMSE_{QRST}$). Such a detailed analysis is required, since previous research [12] has revealed different derivation accuracies for the QRS complex and the STT segment.

The derivation accuracy for MLRLT and EMLRLT was determined and compared in the following steps:

1. The $RMSE_{QRS}$ (Figure 1) was calculated separately for every derived QRST complex (#50) of each derived lead (dV1, dV3, dV4, dV6) for all QRSTrecords (#180).
2. The median $RMSE_{QRS}$ was calculated over the 50 QRST complexes of every derived lead thus providing one subject specific median $RMSE_{QRS}$ for each derived lead. The median instead of the mean $RMSE_{QRS}$ was chosen to reduce the effect of outliers.
3. The distribution of the subject specific median $RMSE_{QRS}$ values (#180) of each derived lead was tested for normality using the Lilliefors test ($p<0.05$). This test revealed a non normal distribution for each derived lead.

4. The skewness of the distribution of the subject specific median $RMSE_{QRS}$ values (#180) of each derived lead was determined. This test revealed an asymmetric probability distribution for each lead.
5. The interquartile range (IQR) and median across all subject specific median $RMSE_{QRS}$ values (#180) of each derived lead were calculated. Thus one global IQR and one global median $RMSE_{QRS}$ value were obtained for each derived lead.
6. The statistical significance of the differences between the global median $RMSE_{QRS}$ of MLRLT and EMLRLT for a derived lead was tested using the paired two-sided sign test (p<0.05). This particular test was chosen, since it does require neither a normal (it is a non-parametric test) nor symmetric distribution of the assessed data. This test revealed the statistical significance of the difference between the global median $RMSE_{QRS}$ values of MLRLT and EMLRLT.

The same assessment (step 1 to 6) was also conducted for the STT segment ($RMSE_{STT}$) and the entire QRST ($RMSE_{QRST}$). The difference in the associated median RMSEs that were achieved by MLRLT and EMLRLT were also found to be statistical significant with p<0.05.

3. Results

Figure 1, Figure 2 and Figure 3 present the global medians and global IQRs of the $RMSE_{QRS}$, $RMSE_{STT}$ and $RMSE_{QRST}$ for each of the derived leads. The global medians and global IQRs of $RMSE_{QRS}$, $RMSE_{STT}$ and $RMSE_{QRST}$ for all MLRLT derived leads show higher values than those of the EMLRLT derived ones. Indicating a superior derivation accuracy of the EMLRLT over MLRLT (for every derived lead and during the entire QRST, the QRS complex and the STT segment).

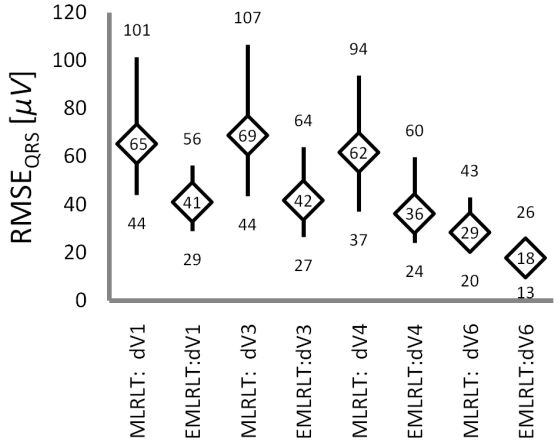

Figure 1: Global median $RMSE_{QRS}$ and global IQR $RMSE_{QRS}$ of MLRLT and EMLRLT derived precordial leads dV1, dV3, dV4 and dV6.

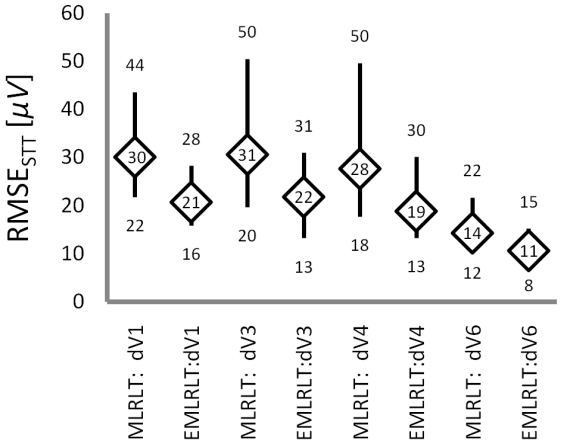

Figure 2: Global median $RMSE_{STT}$ and global IQR $RMSE_{STT}$ of MLRLT and EMLRLT derived precordial leads dV1, dV3, dV4 and dV6.

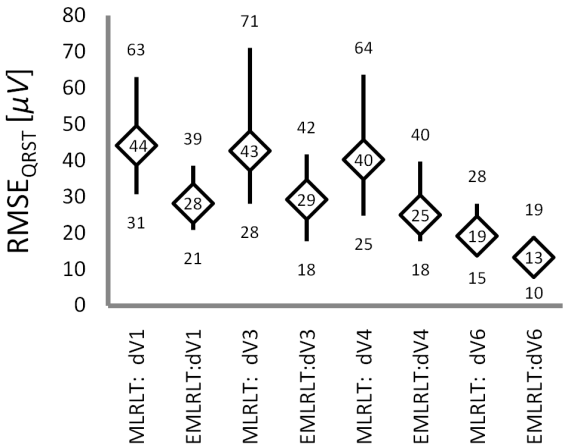

Figure 3: Global median $RMSE_{QRST}$ and global IQR $RMSE_{QRST}$ of MLRLT and EMLRLT derived precordial leads dV1, dV3, dV4 and dV6.

4. Discussion

This study compared the derivation accuracy of MLRLT and EMLRLT in patient-specific lead derivation. This comparison was based upon global medians and global IQRs of $RMSE_{QRS}$, $RMSE_{STT}$ and $RMSE_{QRST}$ for each derived lead and transformation method (MLRLT or EMLRLT).

The results presented within Figure 1, Figure 2 and Figure 3 illustrate the decreased global IQRs of EMLRLT derived leads, suggesting decreased variability within the derivation accuracy, if compared to that achieved by MLRLT.

Decreased global medians $RMSE_{QRS}$, $RMSE_{STT}$ and $RMSE_{QRST}$ for each EMLRLT derived lead (dV1, dV3, dV4 and dV6) suggest a reduction of 30% (circa) in global RMSE when compared to MLRLT derived leads.

4.1. Limitations

This study compares the derivation accuracy of EMLRLT and MLRLT over a relatively short period of time (50 QRST complexes only).

Also, the effect of variability within electrode placement (especially of the precordial electrodes V2 and V5) on the derivation accuracy of EMLRLT and MLRLT has not been taken into consideration.

The pathological condition of each subject was the same during calculation of the weights $a_{i,n}$ and assessment of the derivation accuracy. Further research is required to compare the derivation accuracy of EMLRLT and MLRLT in the presence of changes in the pathological condition (between calculation of the patient-specific weights $a_{i,n}$ and the assessment of the derivation accuracy).

Furthermore, no test has been conducted to consider whether or not the increase in derivation accuracy translates into an improved diagnostic accuracy.

5. Conclusion

This study has demonstrated that alternative methods for lead transformation have the potential to improve lead derivation accuracy. Initial experimentation has revealed that the EMLRLT approach reduces global RMSEs by 30% (circa) when compared to MLRLT. Further research is required in order to assess whether or not the improved derivation accuracy translates into an improved diagnostic accuracy.

Acknowledgements

This work has been supported by the Centre for Intelligent Point of Care Sensors, funded by the Department for Education and Learning, Northern Ireland.

References

[1] Feild DQ, Zhou SH, Helfenbein ED, Gregg RE, Lindauer JM. Technical challenges and future directions in lead reconstruction for reduced-lead systems. J Electrocardiol 2008; 41(6):466-473. doi:10.1016/j.jelectrocard.2008.07.019.

[2] Koch H, Bousseljot R, Kreiseler D, Schmitz L. The PTB diagnostic ECG database. [document on the Internet]. PhysioNet; n.d. [cited 2010 Jun 11]. Available from: http://www.physionet.org/physiobank/database/ptbdb/

[3] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000; 101(23):e215-e220.

[4] The CSE Working Party. Recommendations for measurement standards in quantitative electrocardiography. Eur.Heart J. 1985; 6(10):815-825.

[5] Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, et al. Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology: A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. Circulation 2007; 115(10):1306-1324. doi:10.1161/CIRCULATIONAHA.106.180200.

[6] Laguna P, Jané R, Bogatell E, Anglada DV. QRS detection and waveform boundary recognition using ecgpuwave [document on the Internet].PhysioNet; n.d. [updated 2008 Jun 12; cited 2010 May 10]. Available from: http://www.physionet.org/physiotools/ecgpuwave/

[7] Laguna P, Jané R, Caminal P. Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database. Comput. Biomed. Res. 1994; 27(1):45-60. doi:10.1006/cbmr.1994.1006.

[8] Jané R, Blasi A, García J, Laguna P. Evaluation of an automatic threshold based detector of waveform limits in holter ECG with the QT database. Comput Cardiol 1997; 24:295-298. doi:10.1109/CIC.1997.647889.

[9] Nelwan SP, Kors JA, Meij SH, van Bemmel JH, Simoons ML. Reconstruction of the 12-lead electrocardiogram from reduced lead sets. J Electrocardiol 2004; 37(1):11-18. doi:10.1016/j.jelectrocard.2003.10.004.

[10] Nelwan SP, Crater SW, Meij SH, van Dam TB, Kors JA, Simoons ML, et al. Simultaneous comparison of 3 derived 12-lead electrocardiograms with standard electrocardiogram at rest and during percutaneous coronary occlusion. J Electrocardiol 2004; 41(3):230-237. doi:10.1016/j.jelectrocard.2008.01.011.

[11] Nelwan SP, Meij SH, van Dam TB, Kors JA. Correction of ECG variations due to non-standard electrode positions. Comput Cardiol 2001; 28:317-319. doi:10.1109/CIC.2001.977656.

[12] Gregg RE, Zhou SH, Lindauer JM, Helfenbein ED, Feild DQ. Limitations on the re-use of patient specific coefficients for 12-lead ECG reconstruction. Comput Cardiol 2008; 35:209-212. doi:0.1109/CIC.2008.4749014.

Address for correspondence.

Daniel Guldenring (Room 16J26)
School of Computing and Mathematics
University of Ulster, Shore Road, Newtownabbey,
Co. Antrim, BT37 0QB, guldenring-d2@email.ulster.ac.uk