# Principal Component Analysis of Differential Electrocardiographic Leads – A Contribution to the Synthesis of ECGs

I. Tomašić[1], R. Trobec[1], and R. Magjarević[2]

[1] Jožef Stefan Institute / Department of Communication Systems, Ljubljana, Slovenia
[2] Faculty of Electrical Engineering and Computing / Department of Electronic Systems and Information Processing, Zagreb, Croatia

*Abstract*— **The principal component analysis (PCA) is applied on juxtaposed average beats of unipolar and differential leads (DLs), obtained from 35-lead multichannel ECGs, for assessing proportions of cardiac electrical activity variance that the leads measure. It was shown that 93% of the total 35-lead measured variance can be explained by the first three principal components (PCs). An analysis of correlations between the ECG leads and the first three PCs shows that the leads employing electrodes further to the heart pick up less of the total ECG variance. It was also demonstrated that DLs with highest correlations to the first three PCs, coincide with the universal DLs selected by our previously published ECG synthesis algorithm: BDLSA. This enables the application of the PCA in the selection of optimal DLs combinations for the ECG synthesis.**

*Keywords*— **Electrocardiography, Differential leads, Principal component analysis, Derived electrocardiograms, ECG synthesis.**

## I. Introduction

One of the most promising derived 12-lead ECG systems [1], i.e. systems with reduced numbers of leads that synthesize 12-lead electrocardiograms (ECGs), employs leads measured with wearable ECG devices called wireless body electrodes (WEs) [2, 3, 4, 5], implemented by two connected skin electrodes in a distance of about 5 cm and a device that transfers measured signal wirelessly. The WEs enable the minimal use of wires on the body and so increase the wearable comfort. They can be placed on a body surface at arbitrary positions, but to synthesize the best possible 12-lead ECG from WEs' measurements, it is necessary to position three WEs optimally [6].

In our previous study [7] the best WEs universal positions (applicable for any person) have been determined, by applying a novel "Best differential leads selection algorithm" (BDLSA) that searches for a combination of three leads, from which an ECG with the highest correlation coefficient (CC) to a target ECG, can be synthesized. The determined leads can be applied for the purpose of the 12-lead ECG or for a multichannel ECG (MECG) synthesis (Fig. 1). The BDLSA uses leads that are formed as differences of neighboring unipolar MECG leads and are therefore named differential leads (DLs). The DLs are aimed to be measured by WEs because of the short inter-electrode distance, but they can also be measured by other devices, e.g. devices that employ smart textiles.

For the universal synthesis, the DLs from various persons where concatenated to obtain all-embracing DLs which we call generic DLs. When BDLSA is supplied with the generic DLs, it yields universal combination of three DLs, which is the best combination in terms of CC among all the combinations. The algorithm also yields a universal transformation matrix, which is the best universal transformation to the 12-lead ECG (in the least-squares sense) among all transformations for a specific combination of DLs.

The BDLSA evaluations show that the synthesized ECGs are of an acceptable quality [6, 7]. This suggests that the cardiac electrical activity variance is successfully collected by the three WEs placed optimally. This research applies principal component analysis (PCA) to the generic unipolar leads and generic DLs, for the purpose of analyzing their signal variance and the amount of cardiac electrical activity variance that they pick up. The objective is also to investigate the PCA potential for selecting optimal DLs.

## II. Methods

### A. Studied Data

Twenty healthy volunteers (13 male, 7 female, mean age ± SD = 50.6±9) with no previous medical record related to heart disease and with a normal 12-lead ECG, and twenty-seven patients (22 male, 5 female, mean age ± SD = 58±10), scheduled for a coronary artery bypass surgery, were included in our study. The patients had one or multiple significant coronary artery stenoses, whereas eight of them had an old myocardial infarction. Informed consent was obtained from all the subjects before the study. A single MECG measurement was obtained from each healthy volunteer and two MECG measurements from each patient: the first, one day before the surgery, and the second, in the period from the fifth to the seventh day after the surgery. For the data recording device and procedure please refer to [8]. The measurements were obtained during our previous studies [9].

I. Lacković et al. (eds.), *The International Conference on Health Informatics*, IFMBE Proceedings 42,
DOI: 10.1007/978-3-319-03005-0_66, 

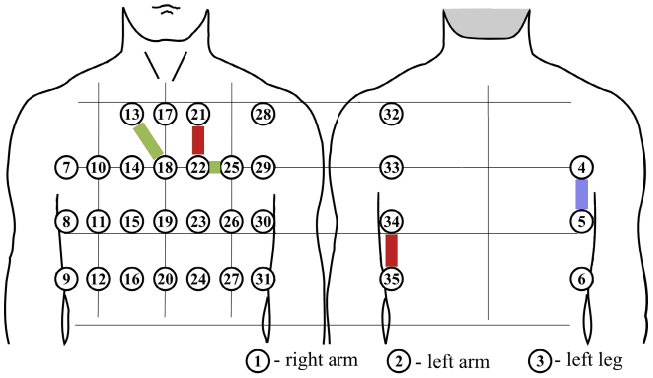


Fig. 1: Schematic locations of MECG electrodes on the chest (left) and the back (right). The leads indicated by green lines are the calculated best leads for the 12-lead ECG synthesis: {(13, 18), (22, 25)}, whereas the ones indicated by red lines: {(21, 22), (34, 35)} are the calculated best leads for the MECG synthesis. The lead indicated in blue is in both sets of best leads: {(4, 5)}.

The length of each measurement was 360 seconds. They were processed using MatLab (MathWorks, Inc.) where the average beats where obtained for each lead of the baseline corrected MECGs, by the procedure described in [7]. Only the average beats where used in all the subsequent processing. Note that the high-frequency noise is satisfactorily damped by the averaging [10].

### *B. Generic Differential Leads*

A method of juxtaposing measurements was used: the average beats from each MECG were concatenated for each lead thus forming generic MECG leads. This method was previously used in other studies, e.g., [7], [11]. The generic DLs are formed by taking differences of juxtaposed MECG lead as specified in the next section (see Eq. (2)).

### *C. ECG Principal Component Analysis – A Brief Introduction and Interpretation*

The PCA analyses a total variance of a random vector but when applied for the purpose of an ECG analyses, the random vector is replaced with the ECG measurement of interest [12].

A MECG measurement can be denoted as

$$\Gamma = \{\Gamma_1, \cdots, \Gamma_i, \cdots, \Gamma_\gamma\}, \tag{1}$$

where $\gamma$ is the total number of leads employed by a MECG. Note that in the scope of this paper lead voltages are denoted in italic to be distinguished from lead labels.

A differential lead is defined as

$$DL_{i,j} = \Gamma_i - \Gamma_j, \quad i,j \in \{1,\dots,\gamma\}, i \neq j\ , \tag{2}$$

where pairs of $i$ and $j$ are chosen such that only differences of neighboring MECG leads are taken into account [6], e.g. (13,18), (20,24), etc. Note that there are 91 DLs for the 35-lead MECG.

Total variance of $\boldsymbol{\Gamma} = [\Gamma_1 \dots \Gamma_\gamma]$ is the sum of elements on its covariance matrix diagonal: $tr(\boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}$ is diagonalizable if it has $\gamma$ different eigenvalues. The diagonalization is conducted by matrix $\boldsymbol{T}$ of its eigenvectors [13]:

$$\boldsymbol{D} = \boldsymbol{T}^{-1}\boldsymbol{\Sigma}\boldsymbol{T}, \tag{3}$$

where

$$\boldsymbol{T} = [\boldsymbol{t}_1, \dots, \boldsymbol{t}_\gamma] \in O_\gamma,\ \boldsymbol{D} = diag(\lambda_1, \dots, \lambda_\gamma), \tag{4}$$

where $O_\gamma = \{\boldsymbol{H} \in \mathbb{R}_\gamma^\gamma : \boldsymbol{H} \cdot \boldsymbol{H}^T = \boldsymbol{I}\}$ is the $\gamma$-order group of orthogonal matrices, whereas $diag(\lambda_1, \dots, \lambda_y)$ is a diagonal matrix with $\lambda_1, \dots, \lambda_y \in \mathbb{R}$ on diagonal. $\lambda_1 \geq \cdots \geq \lambda_y$ are ordered eigenvalues of matrix $\boldsymbol{\Sigma}$.

If we put

$$\boldsymbol{Y} = \boldsymbol{T}^T \cdot \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{t}_1^T\boldsymbol{\Gamma} \\ \\ \boldsymbol{t}_\gamma^T\boldsymbol{\Gamma} \end{bmatrix}, \tag{5}$$

it can be shown that $\boldsymbol{\Gamma}$ and $\boldsymbol{Y}$ have equal total variance and that variables $Y_i$ are uncorrelated [14]. By definition, variables $Y_i = \boldsymbol{t}_i^T\boldsymbol{\Gamma}, (i = 1, \dots, \gamma)$ are named the principal components (PCs) of $\boldsymbol{\Gamma}$.

Let the sample CC between variable $\Gamma_i$ and principal component $Y_j$ be $r_{\Gamma_i Y_j}$. It can be shown [14] that following holds:

$$\sum_{j=1}^{\gamma} r_{\Gamma_i Y_j}^2 = 1,\ \forall i \in \{1, \dots, \gamma\} \tag{6}$$

so that $r_{\Gamma_i Y_j}^2$ can be conceived as the variance proportion of $\Gamma_i$ explained by PC $Y_j$.

In the space of the first two PCs we can draw $r_{\Gamma_i Y_1}$ and $r_{\Gamma_i Y_2}$ .The obtained graphical representation shows how much each of the input variables correlates with PCs $Y_1$ and $Y_2$. It follows from (6) that $r_{\Gamma_i Y_1}^2 + r_{\Gamma_i Y_2}^2 \leq 1$, which shows that all the points on such a figure always fall inside a unit circle. The variables represented with dots nearer to the circle are more correlated with first two PCs. For a variable more correlated with PCs (especially the first one) it may be inferred that it contributes more to the variability of the entire system.

The PCA was applied on the juxtaposed MECG leads $(\gamma = 35)$, which enables the results to be interpreted as being true in average for each person. The interpretation is justified by the fact that average beats of each person, from the studied population, are fairly contained in the juxtaposed leads. The generic DLs contributions to the MECG system variability are estimated by calculating the correla tion between the DLs and first three PCs.

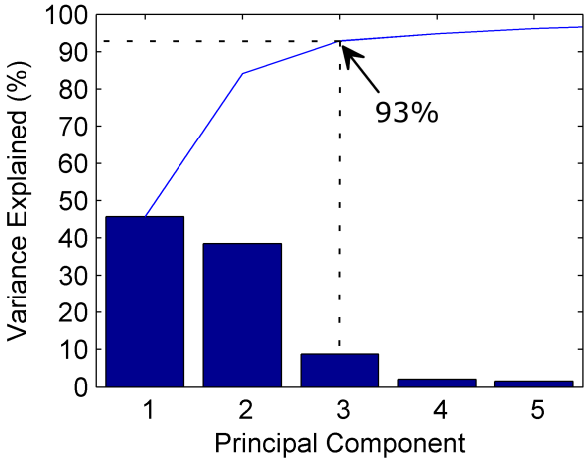


Fig. 2: Variances explained by the first five PCs. The line represents the cumulative variance explained.

## III. Results

The correlation between DLs and PCs was analyzed only for the first three PCs, as it was found that approximately 93% of the total variance in the 35-leads MECG system can be explained by the first three PCs (Fig. 2), which confirms similar result from previous investigations [15], [16]. Figure 3 shows how MECG leads correlate with two first PCs.

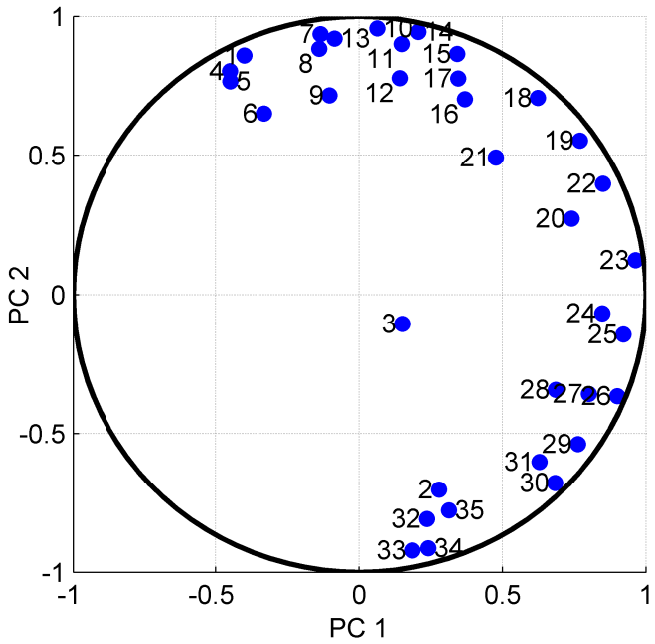


Fig. 3: Correlations between generic MECG leads and the first two PCs.

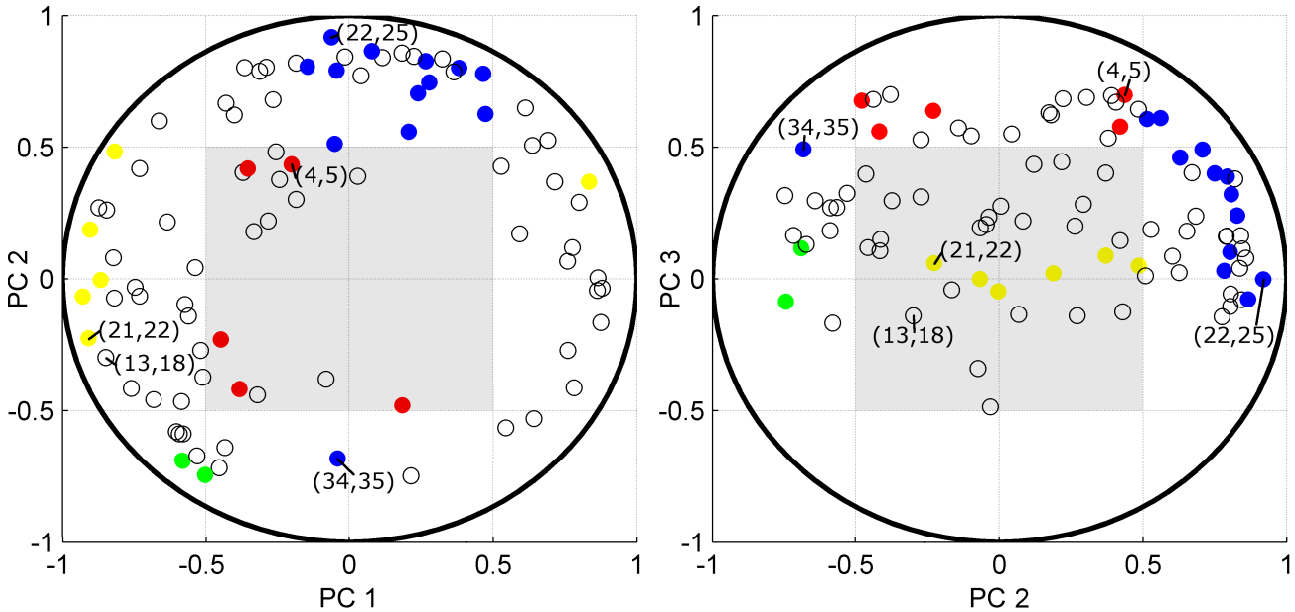


Fig. 4: CCs between generic DLs and PCs one and two (left), two and three (right). Colored circles represent all DLs that may be found in top thirty optimal DLs combinations provided by the BDLSA.: blue - abs. CCs to PC1 less than 0.5 ($|r_{DL_iY_1}| < 0.5$), yellow - abs. CCs to PC2 less than 0.5 ($|r_{DL_iY_2}| < 0.5$), red - abs. CCs to both PC1 and PC2 less than 0.5 ($|r_{DL_iY_1}| < 0.5$ *and* $|r_{DL_iY_2}| < 0.5$), green: remaining optimal DLs, empty circles: DLs not present in the top thirty optimal DLs combinations. The best DLs selected by the BDLSA for the MECG and 12-lead synthesis are indicated by the electrode pairs in parentheses (see Fig. 1). The shaded rectangle covers the areas in which the CCs with both PCs are less than 0.5.

In Fig. 4 the correlations between generic DLs and first three PCs are shown.

## IV. DISCUSSION

Almost the complete variance in the 35-leads MECG can be explained by the first three PCs. This implies that for each person there exist three leads that can collect 93% of a hearth's electrical activity variance (Fig. 2). It is therefore sufficient to analyze DLs' CCs to only the first three PC.

Fig. 4 shows 91 DLs' CCs to the first PCs. Marked with color are the DLs selected by the BDLSA in the thirty best combinations of DLs. It is evident from the left graph that the algorithm selects DLs with a high CCs to the first two PCs. The red circles however indicate DLs that have CCs smaller than 0.5 with both first two PCs. Nevertheless, the right graph reveals that these DLs are among those with the highest CCs to the third PC. The figure shows that the BDLSA selects DLs with highest CCs to the first three PCs. Note that our additional inspection of the thirty best DLs combinations, showed that each DL in every optimal triplet of DLs, was highly correlated to a different PC. It follows that for the purpose of selecting three optimal DL for the ECG synthesis, one can select three DLs that each has a high CC with a different of the first three PCs. The quality of the synthesis still needs to be confirmed by the standard evaluation measures [1].

Fig. 3 reveals that the same type of graph can be used for assessing unipolar leads. The leads that are further to the heart have smaller CCs to the first two PCs so it shows that they pick a smaller amount of the total system's variance. The lead 3 employs an electrode at a left foot which is the furthest electrode; hence it has the smallest CCs with both PCs.

## V. CONCLUSIONS

We have demonstrated that the PCA can be used for assessing both unipolar leads and DLs. It was additionally demonstrated that DLs with highest correlations with the first three PCs, coincide with the DLs selected by the BDLSA. This is an asset to the application of the PCA in selecting optimal combinations of DLs for the ECG synthesis, instead of applying BDLSA, which is an algorithm with a significantly higher computational complexity.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## REFERENCES


1. Tomasic I, Trobec R (2014) Electrocardiographic systems with reduced numbers of leads - synthesis of the 12-lead ECG. IEEE Rev Biomed Eng 7
2. Trobec R, Depolli M, Avbelj V (2010) Wireless network of bipolar body electrodes. In: WONS. Kranjska Gora, pp 145-150
3. Valchinov ES, Pallikarakis NE (2007) Wearable wireless biopotential electrode for ECG monitoring. In: IFMBE Proc. pp 373-376
4. Nemati E, Deen MJ, Mondal T (2012) A Wireless Wearable ECG Sensor for Long-Term Applications. IEEE Commun Mag 50:36-43
5. Kang TH, Merritt CR, Grant E, Pourdeyhimi B, Nagle HT (2008) Nonwoven fabric active electrodes for biopotential measurement during normal daily activity. IEEE Trans Biomed Eng 55:188-195
6. Trobec R, Tomašić I (2011) Synthesis of the 12-Lead Electrocardiogram From Differential Leads. IEEE Trans Inf Technol Biomed 15:615 - 621
7. Tomasic I, Frljak S, Trobec R (2013) Estimating the Universal Positions of Wireless Body Electrodes for Measuring Cardiac Electrical Activity. IEEE Trans Biomed Eng
8. Avbelj V, Trobec R, Gersak B, Vokac D (1997) Multichannel ECG measurement system. In: Kokol PS, B.; et al. (ed) Proc. of the 10th IEEE Symposium on Computer-Based Medical Systems. New York: IEEE Computer Society Press, pp 81-84
9. Frljak S, Avbelj V, Trobec R, Meglic B, Ujiie T, Gersak B (2003) Beat-to-beat QT interval variability before and after cardiac surgery. Comput Biol Med 33:267-276
10. Jager F (2006) Introduction to Feature Extraction. In: Clifford GD, Azuaje F, McSharry PE (eds) Advanced Methods and Tools for ECG Data Analysis. Artech House, Inc., pp 245-267
11. Atoui H, Fayn J, Rubel P (2010) A novel neural-network model for deriving standard 12-lead ECGs from serial three-lead ECGs: application to self-care. IEEE Trans Inf Technol Biomed 14:883-900
12. Castells F, Laguna P, Sörnmo L, Bollmann A, Roig JM (2007) Principal component analysis in ECG signal processing. Eurasip J Adv Sig Pr 2007
13. Hoffman K, Kunze R (1971) Elementary Canonical Forms. In: Linear Algebra. Pretince-Hall, Inc., Englewood Cliffy, New Jersey, pp 181-226
14. Härdle W, Simar L (2007) Principal Components Analysis. In: Applied Multivariate Statistical Analysis. Springer Berlin Heidelberg New York, pp 215-250
15. Finlay DD, Nugent CD, Donnelly MP, Lux RL (2010) Eigenleads: ECG leads for maximizing information capture and improving SNR. IEEE Trans Inf Technol Biomed 14:69-78
16. Tomasic I, Skala K, Trobec R (2008) Principal component analysis and visualization in optimization and personalization of lead's set for generation of standard 12-lead ECG. In: MIPRO 2008, Proceedings of the 27th International Convention. pp 307-313



Corresponding author:
Author: Ivan Tomašić
Institute: Jožef Stefan Institute
Street: Jamova cesta 39
City: Ljubljana
Country: Slovenia
Email: ivan.tomasic@ijs.si