Pareto-Optimal Selection of Biosignal Acquisition Configurations for an Educational Kit Using Skyline Queries: A Clinical-Scenario Analysis of ECG/PPG Acquisition Parameters

Taeyun Ha

Eastlake High School

Abstract

When acquiring electrocardiogram (ECG) and photoplethysmogram (PPG) signals, the acquisition configuration—sampling rate, amplifier gain, band-pass filtering, notch filtering, and ADC resolution—inherently trades off diagnostic fidelity against resource cost. Selecting a single "best" configuration by a weighted sum suffers from arbitrary weighting and the inability to recover solutions on non-convex regions of the Pareto frontier. This work introduces the skyline query from databases to derive, without any weights, the set of Pareto-optimal configurations over a seven-dimensional attribute space: three signal-quality dimensions (SNR, ST/low-frequency fidelity LFF, QRS-detection F1) and four resource/noise dimensions (motion-artifact ratio MAR, power, data volume, latency). We further model three clinical scenarios—resting diagnostic ECG, ambulatory monitoring, and ICU real-time monitoring—as subspace skylines, showing that different priorities surface different optimal sets. Using an illustrative application to an Arduino-based educational kit (GT-BME), we discuss the value of jointly teaching database querying and biosignal processing.

Keyboard

skyline query, Pareto optimality, multi-objective decision, biosignal acquisition, ECG/PPG, database, biomedical engineering education

Introduction

In biomedical-engineering and biosignal-processing education, learners internalize the principles of sensors, analog front ends, and digital signal processing by acquiring and analyzing ECG, PPG, and SpO_2 signals firsthand. Arduino-based educational kits such as GT-BME—providing an AD8232 ECG module, a MAX30102 PPG/SpO_2 sensor, op-amp circuits, and a built-in ECG simulator—make such practice possible.

Yet even with the same kit and the same signal, the acquisition configuration the learner sets—sampling rate, gain, band-pass cutoffs, 50/60 Hz notch on/off, and ADC resolution—substantially changes the resulting signal. A higher sampling rate improves waveform fidelity but raises data volume and power draw. Higher gain increases amplitude but risks saturation and noise amplification. A wider passband preserves diagnostic information such as the ST segment but also admits baseline wander and motion noise. Configuration selection is therefore an inherently multi-objective decision in which clinical and resource goals conflict.

Traditionally this choice has relied on instructor experience, a single quality metric, or a weighted-sum score. A weighted sum, however, suffers from (1) arbitrariness in choosing weights, (2) collapsing incommensurable metrics into one number so the trade-off structure becomes invisible to the learner, and (3) a structural inability to select superior solutions on non-convex Pareto frontiers. This work introduces the skyline query—the standard operator for multi-criteria preference queries in databases—to this problem. Our contributions are as follows.

1. We formalize ECG/PPG acquisition-configuration selection as a skyline query over a seven-dimensional dominance relation, introducing two clinically meaningful metrics (ST/low-frequency fidelity, LFF; motion-artifact ratio, MAR) to extend the attribute space.
2. We design a relational data model storing acquisition runs, configurations, and metrics, together with a skyline query layer based on the SFS algorithm.
3. We model three clinical scenarios—resting diagnostic, ambulatory/wearable, and ICU real-time—as subspace skylines, showing that the same data yield different Pareto sets under different priorities.
4. We demonstrate the framework on an illustrative GT-BME scenario and discuss its value as a converged tool for teaching database querying and biosignal processing together.

Related Work

1. Biosignal quality and clinical metrics

Biosignal quality does not reduce to a single index; it is assessed across SNR, baseline wander, motion/EMG noise, and downstream accuracy such as QRS-detection rate. Clifford et al. fused multiple signal-quality indices to judge clinical acceptability, and Elgendi compared optimal quality indices for PPG. AHA recommendations on ECG standardization (Kligfield et al.; Bailey et al.) specify that diagnostic ECG requires preservation of a roughly 0.05–40–150 Hz band, and in particular the 0.05–1 Hz low-frequency content critical to ST-segment assessment. The classic Pan–Tompkins QRS detector and the MIT–BIH arrhythmia database are widely used as evaluation standards. We build on the observation that such multiple quality indices are inherently multi-objective.

2. Skyline queries

Since Börzönyi et al. introduced the skyline operator into database query languages, it has become a core tool for multi-criteria preference queries. Core algorithms include block-nested-loop (BNL), Sort–Filter–Skyline (SFS) combining presorting with filtering, and Branch–and–Bound Skyline (BBS) over R-trees. To mitigate the rapid growth of skyline size in high dimensions, subspace skylines that consider only a subset of dimensions have been proposed. We adopt SFS as the computation engine and subspace skylines as a per-scenario analysis tool.

Background: Physiology and Clinical Metrics

1. ECG morphology and the ST segment

The ECG comprises the P–QRS–T complex: the P wave reflects atrial depolarization, the QRS complex ventricular depolarization, and the T wave ventricular repolarization. Subtle ST-segment elevation or depression is decisive for diagnosing myocardial ischemia and infarction, and is sensitive to preservation of low-frequency content near 0.05–1 Hz. Raising the high-pass cutoff above 0.5 Hz reduces baseline wander but can distort the ST segment. We define LFF (ST/low-frequency fidelity) to quantify preservation of this diagnostic band; values near 1 indicate faithful preservation.

2. PPG and motion artifact

Photoplethysmography optically measures volumetric changes in peripheral microcirculation and is used to estimate heart rate and oxygen saturation. PPG is highly sensitive to motion, so in ambulatory and wearable settings motion artifact severely corrupts the signal. We define MAR (motion-artifact ratio) as the relative proportion of in-band motion-artifact energy; smaller values indicate greater robustness under motion.

Attribute summary. Benefit attributes (higher is better) are SNR, LFF, and F1 (QRS-detection F1); cost attributes (lower is better) are MAR, power, data volume, and latency. Because the two directions coexist, a simple sum is inappropriate and a dominance-based skyline becomes the natural solution.

Problem Definition and Method

1. Formalization

Let the configuration set be $C=\{c_1,\dots,c_n\}$ and the attribute set A={SNR, LFF, F1, MAR, Power, Data, Latency}. For benefit attributes larger is preferred; for cost attributes smaller is preferred. A configuration c_i dominates c_j if c_i is no worse than c_j on every attribute and strictly better on at least one. The skyline is the set of non-dominated configurations—the Pareto-optimal set.

2. Relational data model and query layer

We model the data with three relations: Configuration (config_id, sampling_rate, gain, HPF/LPF cutoff, notch, ADC_resolution), Run (run_id, config_id, subject, timestamp), and Metric (run_id, SNR, LFF, F1, MAR, power, data, latency). The query layer aggregates Metric per configuration and computes the skyline via SFS, which presorts by a monotone score so that each candidate need only be compared against already-emitted non-dominated configurations.

3. Subspace skylines

The relevant attribute subset depends on clinical purpose. The resting-diagnostic scenario prioritizes diagnostic fidelity (SNR, LFF, F1); ambulatory monitoring prioritizes resource efficiency (power, data, latency); and ICU real-time monitoring prioritizes timeliness and accuracy (latency, F1, SNR). Computing the skyline within each subspace yields a per-scenario Pareto-optimal set from the same data.

Illustrative Application (GT-BME)

All values in the tables below are illustrative data meant to demonstrate the framework; in practice they must be replaced with measured values from the GT-BME kit. Table 1 lists nine candidate acquisition configurations with seven attributes.

Config	SNR (dB)	LFF	F1	MAR	Power (mW)	Data (KB/s)	Latency (ms)
C1	27	0.97	0.96	0.3	46	12	42
C2	24	0.9	0.95	0.22	32	8	26
C3	20	0.8	0.92	0.2	16	2.5	15
C4	22	0.88	0.93	0.24	34	9	28
C5	21	0.85	0.93	0.18	28	7	22
C6	23	0.87	0.94	0.15	26	6	20
C7	30	0.93	0.98	0.38	55	16	50
C8	17	0.66	0.86	0.28	11	3	9
C9	19	0.78	0.9	0.22	17	5	17

Table 1. *Candidate acquisition configurations with illustrative seven-dimensional attribute values.*

Table 2 reports the skyline over the full seven-dimensional space. Six of nine configurations (C1, C2, C3, C6, C7, C8) remain non-dominated; C4 is dominated by C2, C5 by C6, and C9 by C3 on all dimensions. Thus, three clearly inferior configurations are filtered out objectively, without any weighting.

Config	On skyline?	Note (dominance)
C1	Skyline (non-dominated)	Pareto-optimal configuration
C2	Skyline (non-dominated)	Pareto-optimal configuration
C3	Skyline (non-dominated)	Pareto-optimal configuration
C4	Dominated (excluded)	Dominated on all dimensions by C2
C5	Dominated (excluded)	Dominated on all dimensions by C6
C6	Skyline (non-dominated)	Pareto-optimal configuration
C7	Skyline (non-dominated)	Pareto-optimal configuration
C8	Skyline (non-dominated)	Pareto-optimal configuration
C9	Dominated (excluded)	Dominated on all dimensions by C3

Table 2. *Full seven-dimensional skyline result and dominance relations.*

Table 3 gives the subspace skylines for the three clinical scenarios. Under resting diagnosis

(SNR·LFF·F1), only the highest-fidelity C1 and C7 remain; under ambulatory monitoring (power·data·latency), only the most efficient C3 and C8 remain. Under ICU real-time monitoring (latency·F1·SNR), by contrast, latency and accuracy conflict strongly and the skyline barely narrows—indicating that no single optimum exists and that an operational policy must guide the final choice.

Clinical scenario	Priority attributes (subspace)	Subspace skyline
Resting diagnostic ECG	SNR, LFF, F1 (diagnostic fidelity)	C1, C7
Ambulatory / wearable	Power, Data, Latency (resource)	C3, C8
ICU real-time monitoring	Latency, F1, SNR (timeliness+accuracy)	C1, C2, C3, C6, C7, C8

Table 3. *Subspace skylines per clinical scenario.*

Discussion

The framework offers three educational benefits. First, learners distinguish "clearly inferior" from "genuinely trade-off" configurations without setting arbitrary weights. Second, through subspace skylines they experience how the same data are interpreted differently across clinical purposes, simultaneously acquiring signal-processing knowledge and database-query concepts. Third, designing the relational model and implementing SFS can be turned into hands-on assignments that strengthen informatics skills.

Limitations include (1) all reported values being illustrative, (2) not modeling measurement uncertainty or correlation among attributes, and (3) the potential for skyline size to grow in high dimensions, which may call for relaxations such as k-dominant skylines. Future work will calibrate the LFF and MAR metrics on measured GT-BME data and extend the tool so learners can interactively explore scenario weightings.

Conclusion

We formalized acquisition-configuration selection for an educational biosignal kit as a seven-dimensional skyline query, combining clinical metrics (LFF, MAR) with clinical-scenario subspace skylines. The illustrative application derived Pareto-optimal configurations without weights and showed that different scenarios surface different optimal sets, suggesting the approach's promise as a converged tool for teaching database querying alongside biosignal processing.

References

1. Kligfield P, et al. Recommendations for the standardization and interpretation of the electrocardiogram. Circulation. 2007; 115(10): 1306–1324.
2. Bailey JJ, et al. Recommendations for standardization and specifications in automated electrocardiography. Circulation. 1990; 81(2): 730–739.
3. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007;28(3): R1–R39.
4. Sörnmo L, Laguna P. Bioelectrical Signal Processing in Cardiac and Neurological Applications. Elsevier; 2005.
5. Pan J, Tompkins WJ. A real-time QRS

detection algorithm. IEEE Trans Biomed Eng. 1985; BME-32(3):230–236.
6. Moody GB, Mark RG. The impact of the MIT–BIH arrhythmia database. IEEE Eng Med Biol Mag. 2001;20(3):45–50.
7. Clifford GD, et al. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiol Meas. 2012;33(9):1419–1433.
8. Elgendi M. Optimal signal quality index for photoplethysmogram signals. Bioengineering. 2016;3(4):21.
9. Börzönyi S, Kossmann D, Stocker K. The skyline operator. Proc. IEEE ICDE. 2001:421–430.
10. Papadias D, Tao Y, Fu G, Seeger B. Progressive skyline computation in database systems. ACM Trans. Database Syst. 2005;30(1):41–82.
11. Chomicki J, Godfrey P, Gryz J, Liang D. Skyline with presorting (SFS). Proc. IEEE ICDE. 2003:717–719.
12. Yuan Y, et al. Efficient computation of the skyline cube. Proc. VLDB. 2005:241–252.