

NOTE

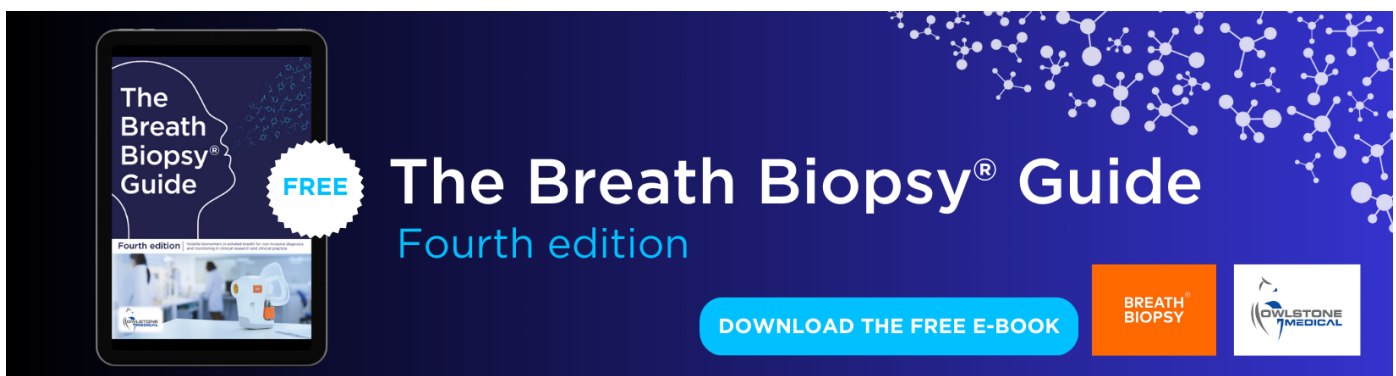
A new database with annotations of P waves in ECGs with various types of arrhythmias

To cite this article: Lucie Saclova *et al* 2022 *Physiol. Meas.* **43** 10NT01

View the [article online](#) for updates and enhancements.

You may also like

- [Simultaneous Monitoring of ECG and EDA Using a Wearable Armband for Analyzing Sympathetic Nerve Activity](#)
Farzad Mohaddes, Yilu Zhou, Jenna Pedersen et al.
- [A Complete Adaptive Method for Fetal ECG Extraction Based on Single Channel](#)
Miao Zhang and Guo Wei
- [Impact of Catalyst Ink Dispersing Solvent on PEM Fuel Cell Performance and Durability](#)
Chao Lei, Fan Yang, Natalia Macauley et al.



The Breath Biopsy® Guide
Fourth edition

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL



NOTE

A new database with annotations of P waves in ECGs with various types of arrhythmias

RECEIVED
23 November 2021REVISED
13 August 2022ACCEPTED FOR PUBLICATION
22 September 2022PUBLISHED
26 October 2022Lucie Saclova^{1,2,*} , Andrea Nemcova¹, Radovan Smisek^{1,3}, Lukas Smital¹ , Martin Vitek¹ and Marina Ronzhina¹¹ Department of Biomedical Engineering, Brno University of Technology, Technická 12, Brno 61600, Czech Republic² Department of Technical Studies, College of Polytechnics Jihlava, Tolstého 16, Jihlava 58601, Czech Republic³ Institute of Scientific Instruments, The Czech Academy of Sciences, Královopolská 147, Brno, 61264, Czech Republic

* Author to whom any correspondence should be addressed.

E-mail: Marsanova@vutbr.cz**Keywords:** annotation of P wave, ECG database, ECG pathology, ECG**Abstract**

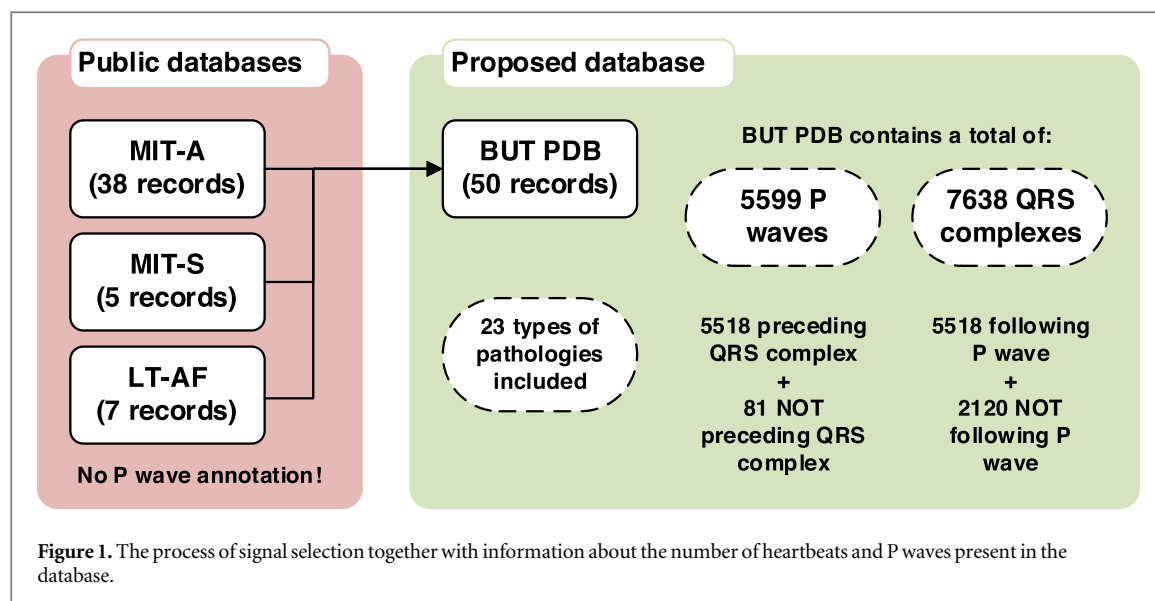
Objective. The aim of this study is to create a database for the development, evaluation and objective comparison of algorithms for P wave detection in ECG signals. **Brno University of Technology ECG Signal Database with Annotations of P-Wave (BUT PDB)** is an ECG signal database with marked peaks of P waves annotated by ECG experts. Currently, there are only a few databases of pathological ECG signals with P-wave annotations, and some are incorrect. **Approach.** The pathological ECG signals used in this work were selected from three existing databases of ECG signals: MIT-BIH Arrhythmia Database, MIT-BIH Supraventricular Arrhythmia Database and Long Term AF Database. The P-wave positions were manually annotated by two ECG experts in all selected signals. **Main results.** The final BUT PDB composed of selected signals consists of 50 two-minute, two-lead pathological ECG signal records with annotated P waves. Each record also contains a description of the diagnosis (pathology) present in the selected part of the record and information about positions and types of QRS complexes. **Significance.** The BUT PDB is created for developing new, more accurate and robust methods for P wave detection. These algorithms will be used in medical practice and will help cardiologists to evaluate ECG records, establish diagnoses and save time.

1. Introduction

Electrocardiography (ECG) is still the most available and widely used method for cardiovascular system examination ([World Health Organisation](#)). The ECG signal reflects the electrical activity of the heart and provides a significant amount of information about the heart function ([Mendis *et al* 2011](#)). Accurate detection of ECG components, such as the P wave, QRS complex and T wave, are fundamental steps in ECG analysis and subsequent cardiac pathological events detection. In practice, automated evaluation of ECG records using software is necessary ([Wagner *et al* 2009](#)). The detection of QRS complexes and T waves is usually efficient. However, methods for P-wave detection are not so successful in physiological signals and especially in pathological signals. This fact applies in both real practice and research ([Portet *et al* 2008](#), [Hossain *et al* 2019](#), [Maršánová *et al* 2019a](#), [Kalyakulina *et al* 2020](#)).

One factor that prevents progress in this field is the lack of publicly available datasets with correct P-wave annotations suitable for training and testing detection algorithms ([Leutheuser *et al* 2016](#), [Maršánová *et al* 2019a](#)).

Methods are usually tested either on part of a publicly available QT database ([Laguna *et al* 1997](#), [Goldberger *et al* 2000](#)) or on a CSE database ([Willems *et al* 1990](#)) that is not publicly available, both with manual P-waves annotations. In addition to these, there are two new databases, namely the MIT-BIH Arrhythmia Database P-Wave Annotations ([Goldberger *et al* 2000](#), [Maršánová *et al* 2019a](#), [2019b](#)) and the Lobachevsky University Electrocardiography Database ([Goldberger *et al* 2000](#), [Kalyakulina *et al* 2020](#)), which is not used frequently yet. There are also two publicly available databases with P-waves annotations that contain mistakes: the P-waves



annotations of the MIT-BIH Arrhythmia Database by Elgendi *et al* (Elgendi *et al* 2016), and the automatically annotated part of the QT database (Laguna *et al* 1997). Thus, these annotations cannot be recommended for use in the testing of P-wave detection algorithms.

The most commonly used databases, the QT database and the Common Standards for Electrocardiography (CSE) database, contain predominantly physiological ECG records or pathologies that do not affect P-wave detection. However, the content of the pathologies in databases is very important for the objective testing of P-wave detection algorithms. During pathological function of the heart, information about the positions of P waves is very important in determining the diagnosis. Unfortunately, current algorithms are not able to detect P waves in pathological signals reliably.

Therefore, we fill this gap and introduce a new database of ECG signals with manually annotated P waves. The Brno University of Technology ECG Signal Database with Annotations of P Wave (BUT PDB) consists of 50 two-minute two-lead ECG signals. The records contain pathologies during which it is usually difficult to detect P waves by automatic algorithms (Portet *et al* 2008, Hossain *et al* 2019, Maršánová *et al* 2019a, Kalyakulina *et al* 2020). The P-waves positions were manually annotated by two ECG experts with seven years of practical experience evaluating Holter ECG records in cardiovascular ambulance (Maršánová *et al* 2019a).

2. Selection of ECG signals

The ECG signals were selected by two ECG experts from three existing databases: the MIT-BIH Arrhythmia (MIT-A) Database (Goldberger *et al* 2000, Moody *et al* 2001), the MIT-BIH Supraventricular Arrhythmia (MIT-S) Database (Greenwald *et al* 1991, Goldberger *et al* 2000) and the Long Term Atrial Fibrillation (LT-AF) Database (Goldberger *et al* 2000, Petrutiu *et al* 2007). All three databases contain ECG signals and annotations of positions and types of QRS complexes. The MIT-A database additionally contains annotations of types of arrhythmias present in records. All of these databases contain 2 leads of ECG signals and do not contain significant noise. Detailed information about these databases is available on Physionet (Goldberger *et al* 2000) or in articles (Greenwald *et al* 1991, Petrutiu *et al* 2007, Moody *et al* 2001).

Two ECG experts went through the databases and selected records with diagnoses which affect the P wave. From the record they selected interesting two-minute sections with higher incidence of pathologies. The signals were chosen to represent as many types of pathologies present in records as possible due to the contents of the databases used. The heartbeats influenced by pathology are not present in these signals consistently, so the ECG signals also contain physiological heartbeats. The final database consists of 50 two-minute, two-lead ECG signals with various types of pathologies. The number and type of pathologies chosen represent a real sample of data from medical practice. From MIT-A database, 38 signals were selected. From MIT-S database, 5 signals were selected. From LT-AF database, 7 signals were selected. The process of signal selection signals together with information about the number of heartbeats and P waves present in database is illustrated in figure 1.

Table 1. Types of pathologies with their abbreviations used in the BUT PDB; number of heartbeats with specific pathology (or P waves which are extra, in case of AVB II and AVB III); number of records in which the pathology is present; and IDs of the records.

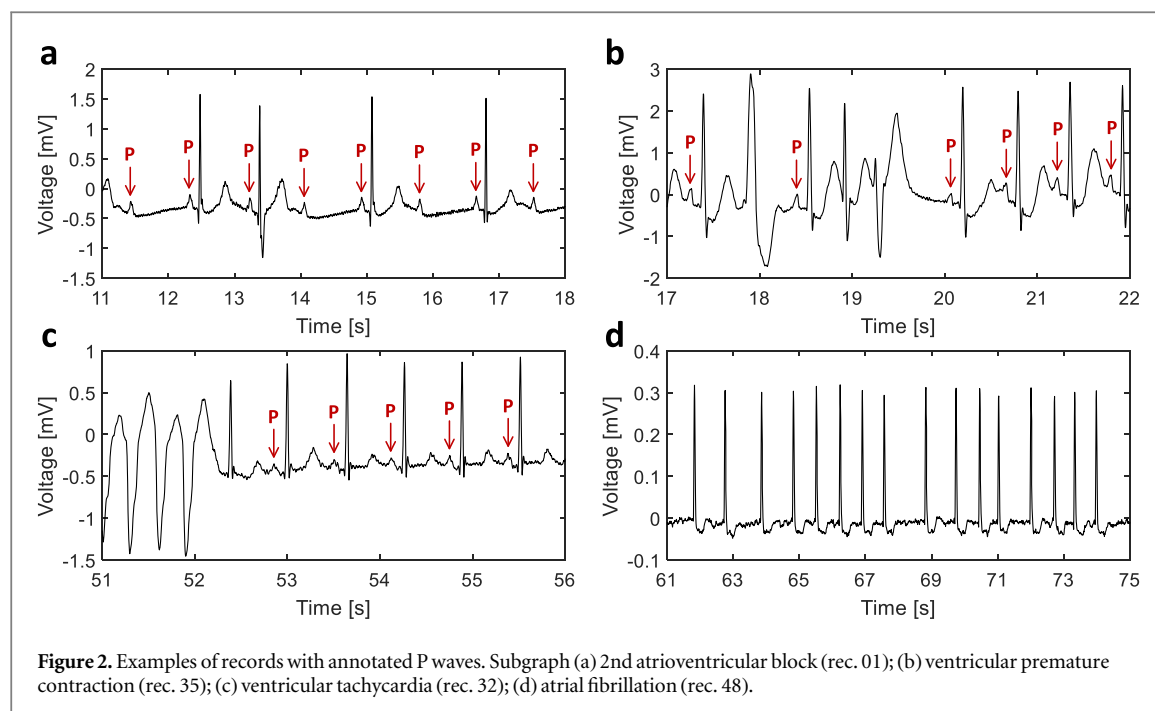
Abb.	Type of pathology	N. of heartbeats with pathology	N. of records	IDs of the records with the pathology
A	Atrial premature beat	142	21	01,04,05,09,16,17, 18,26,28,31,32,35, 38,39,40,41,42,43, 46,49,50
SVTA	Supraventricular tachyarrhythmia	included in A	3	09,11,43
AFIB	Atrial fibrillation	1079	9	07,08,44,45,46,47, 48,49, 50
AFL	Atrial flutter	86	1	38
BI	1st degree atrioventricular block	included in L(140)	1	22
BII	2nd degree atrioventricular block	<i>extra 80 P wave</i>	2	1,13
BIII	3rd degree atrioventricular block	<i>extra 61 P wave</i>	1	3
E	Ventricular escape beat	99	1	9
F	Fusion beat	76	7	06,10,14,19,32,35,36
J	Nodal beat	26	2	7,38
L	Left bundle branch block beat	448	4	21,22,36,41
NA	Sinus arrhythmia	129	1	24
NOD	Nodal premature beat	76	2	6,15
P	Paced rhythm	236	2	3,19
PREX	Pre-excitation	130	1	12
R	Right bundle branch block beat	717	6	01,06,13,26,33,34
V	Ventricular premature beat	547	27	02,03,05,08,10,14, 20,21,22,25,26,27, 28,29,30,31,32,33, 35,36,37,39,40,41, 42,45,47,50
B	Ventricular bigeminy	included in V	3	02,14,27
T	Ventricular trigeminy	included in V	2	27,29
IVR	Idioventricular rhythm	included in V	1	30
VP	Ventricular pair	included in V	1	25
VFL	Ventricular flutter	66	1	33
a	Aberrated atrial pre-mature beat	9	1	23
N	Normal beat	3772		

3. Annotation of ECG signals

The P-waves positions were manually annotated by two ECG experts with seven years of practical experience evaluating Holter ECG records in cardiovascular ambulance (Smíšek *et al* 2017, Maršánová *et al* 2019a). The ECG data was not pre-processed. The first expert provided manual annotations of peaks of P waves, and the second manually checked them. Unclear parts of the records were discussed by both experts until a consensus was reached. Everything was conducted manually without the use of automated annotation software. To facilitate the work of the ECG experts, a free software tool—SignalPlant (Plesinger *et al* 2016)—was used for manual marking of peaks of P waves.

Each record also contains annotation of the diagnosis (pathology) and types of QRS complexes (from the original databases) (Goldberger *et al* 2000, Moody *et al* 2001). The information about pathologies present in records was checked by ECG experts—the original annotations were found correct—then taken from the original databases and supplemented by the experts in cases where information was missing (all signals from MIT-S). The information about types of QRS complexes was taken from the original databases.

Types of pathologies with their abbreviations used in the BUT PDB, number of cases (records), number of heartbeats with specific pathology and IDs of the records in which the pathology is present are listed in table 1. The column ‘Number of heartbeats with specific pathology’ also contains the number of P waves which are present extra - without QRS complex. This is in case of AVB II and AVB III. The column provides also information that numbers of SVTA is included in number of heartbeats A, as same as B, T, IVR, VP, VFL in V - this is due the fact that these pathologies have same origin and morphologic shape of QRS but are present multipletimes (e.g. in pair, trigeminy, tachycardia). BI is included in L, because BI was present in same beats as L (diagnose is left bundle branch block with atrioventricular block 1st degree).



4. Summary of the database

The BUT PDB database contains annotation of P waves in ECG signal with 23 different types of pathologies. The database includes 7,638 QRS complexes, of which 2,120 are without P wave (e.g. atrial fibrillation, ventricular beats or during nodal rhythm) and 5518 following P wave (normal timing). In BUT BDP is 5,599 P waves of which 81 are present without QRS complex (e.g. AVB II).

On Physionet (Goldberger *et al* 2000, Maršánová *et al* 2021), where database is published, information about each record from BUT PDB is listed. The information contains: ID of the record, types of pathologies present in each record, types of heart beats, number of heart beats, number of P waves, name of original database, ID of the record from the original database, the start and the end (in samples) of the selected segment and the sampling frequency. The column ‘Types of pathologies’ contains information about pathological beats as same as pathological rhythm. The column ‘Types of heartbeats’ contains informations about all types of heartbeats present in record, these annotation are taken over from original database and supplemented by missing information by ECG experts (e.g. AFIB, VFL, PREX beats).

The examples of records with various pathologies (i.e. second atrioventricular block (rec. 01), ventricular premature contraction, ventricular tachycardia, atrial fibrillation (rec. 48) and newly annotated P waves are shown in figure 2.

The BUT PDB is available on Physionet (Goldberger *et al* 2000, Maršánová *et al* 2021) (<https://physionet.org/content/but-pdb/1.0.0/>). All data is provided in the WaveForm Database (WFDB) format, which is supported by the WFDB Software Package (Goldberger *et al* 2000). The IDs of the recordings are numbers from 01 to 50. The ECG signals are stored in files with suffix *.dat; the annotations of P waves are stored in files with suffix *.pwave; the positions of QRS complexes, their types and the sampling frequency of each ECG signal are stored in files with suffix *.qrs. The exact types of pathologies present in each signal are described in the text file with the name README.txt.

The codes needed for loading the data are available on Physionet (Goldberger *et al* 2000) (<https://archive.physionet.org/physiotools/wfdb.shtml>). In the case of using Matlab, to load files *.dat with ECG records, use function ‘rdsamp’: e.g. [signal, fs, tm] = rdsamp(‘17.dat’). To load files with annotations of P waves or QRS complexes and their types, use function ‘rdann’: e.g. [P] = rdann(‘17’, ‘pwave’) or [QRS, typeQRS] = rdann(‘17’, ‘qrs’).

5. Conclusions

BUT PDB was created for the development, testing and objective comparison of algorithms for P-waves detection. For objective comparison of algorithms, the prerequisite is that the researchers will use the entire database for testing their algorithms and that they will not select and/or shorten the signals. The BUT PDB

includes a representative sample of pathologies. The BUT PDB will help develop new, more accurate and robust methods for processing and analysing ECG records in the sense of P wave detection. These algorithms will be usable for implementation to the software for ECG signals analysis in real medical practice and thus will help cardiologists to evaluate ECG records.

Acknowledgments

The authors wish to thank LCDR Joshua Swift from Office of Naval Research (ONR) Code 342 and Dr Stephen O'Regan from ONR Global Central and Eastern European Office for their support. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of interest

There are no conflicts of interest.

Funding statement

This work has been funded by the United States Office of Naval Research (ONR) Global, award number N62909-19-1-2006. The authors wish to thank LCDR Joshua Swift from ONR Code 342 and Dr Martina Siwek from ONR Global Central and Eastern European Office for their support.

ORCID iDs

Lucie Saclova  <https://orcid.org/0000-0003-1638-4814>

Lukas Smital  <https://orcid.org/0000-0003-1526-4626>

References

- Elgendi M *et al* 2016 A proof-of-concept study: simple and effective detection of P and T waves in arrhythmic ECG signals *Bioengineering* **3** 1–14
- Goldberger A L *et al* 2000 PhysioBank, PhysioToolkit, and PhysioNet *Circulation* **13** 15–20
- Greenwald S D *et al* 1991 Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information *Proc. Computers in Cardiology* pp 461–4
- Hossain M B *et al* 2019 An accurate QRS complex and P wave detection in ECG signals using complete ensemble empirical mode decomposition with adaptive noise approach *IEEE Access* **7** 128869–80
- Kalyakulina A I *et al* 2020 LUDB: a new open-access validation tool for electrocardiogram delineation algorithms *IEEE Access* **8** 186181–90
- Laguna P *et al* 1997 A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG *Computers in Cardiology* **1997** 673–6
- Leutheuser H *et al* 2016 Instantaneous P- and T-wave detection: assessment of three ECG fiducial points detection algorithms 2016 *IEEE 13th Int. Conf. on Wearable and Implantable Body Sensor Networks* **13**, 329–34
- Moody G B *et al* 2001 The impact of the MIT-BIH arrhythmia database *IEEE Eng. Med. Biol. Mag.* **20** 45–50
- Maršánová L *et al* 2019a Advanced P wave detection in ECG signals during pathology: evaluation in different arrhythmia contexts *Sci. Rep.* **9** 1–11
- Maršánová L *et al* 2019b Automatic detection of P Wave in ECG during ventricular extrasystoles *World Congress on Medical Physics and Biomedical Engineering* 2018 **7** 381–5
- Maršánová L *et al* 2021 Brno university of technology ECG signal database with annotations of P Wave (BUT PDB) *PhysioNet*
- Mendis S *et al* 2011 *Global Atlas on Cardiovascular Disease Prevention and Control* (France: World Health Organisation) 9789241564373
- Petrutiu S *et al* 2007 Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans *EP Europace* **9** 466–70
- Plesinger F *et al* 2016 SignalPlant: an open signal processing software platform *Physiol. Meas.* **37** 38–48
- Portet F P *et al* 2008 P wave detector with PP rhythm tracking: evaluation in different arrhythmia contexts *Physiol. Meas.* **29** 141–55
- Smišek R *et al* 2017 CSE database: extended annotations and new recommendations for ECG software testing *Med. Biol. Eng. Comput.* **8** 1473–82
- Wagner G S *et al* 2009 AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram *J. Am. Coll. Cardiol.* **53** 1003–11
- Willems J L *et al* 1990 Common standards for quantitative electrocardiography *Journal of Electrocardiology* **21** 209–17
- World Health Organisation 2022 Cardiovascular Diseases.