Extract Fetal ECG from Single-Lead Abdominal ECG by De-Shape Short Time Fourier Transform and Nonlocal Median

Li Su[1][†] and Hau-Tieng Wu[2,3][†]*

[1] Research Center for Information Technology Innovation, Taipei, Taiwan, [2] Department of Mathematics, University of Toronto, Toronto, ON, Canada, [3] Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan

The multiple fundamental frequency detection problem and the source separation problem from a single-channel signal containing multiple oscillatory components and a nonstationary noise are both challenging tasks. To extract the fetal electrocardiogram (ECG) from a single-lead maternal abdominal ECG, we need to solve both challenges. We propose a novel method to extract the fetal ECG from a single-lead maternal abdominal ECG, without any additional measurement. The algorithm is composed of three components. First, the maternal and fetal heart rates are estimated by the de-shape short time Fourier transform (STFT), which is a recently proposed nonlinear time-frequency analysis technique. The beat tracking technique is the second component which is applied to accurately obtain the maternal and fetal R peaks. The third component consists of establishing the maternal and fetal ECG waveforms by the nonlocal median. The algorithm is tested on two real databases with the annotation provided by experts (*adfecgdb* database and *CinC2013* database) and a simulated database (*fecgsym*), and provides the state-of-the-art results. We conclude that with the proposed algorithm, the fetal ECG waveform and the fetal heart rate could be accurately obtained from the single-lead maternal abdominal ECG.

Keywords: de-shape short time Fourier transform, fetal electrocardiogram, single-lead maternal abdominal electrocardiogram, beat tracking, nonlocal median, instantaneous frequency

1. INTRODUCTION

Electrocardiograph (ECG) is inarguably the most widely applied measurement to non-invasively study cardiac activity, since its appearance in 1901 [1]. Its waveform provides a significant amount of clinical information. In addition, the time-varying speed of heart beating, widely understood as the heart rate variability (HRV), has proven to be a portal to our physiological dynamical status. While it has been widely applied in different scenarios, its application to the intra-uterus fetus is still limited, mainly due to the lack of a direct contact measurement of the fetal ECG (fECG) signal. Like the adult ECG signal processing, there are two main purposes in the fECG signal processing. First, we want to non-invasively obtain the fetal heart rate, which is intimately related to the fetal distress [2]. Second, we would like to analyze the fECG morphology for the sake of diagnosing cardiac problems. However, the fECG morphological analysis is infrequently performed in clinics, except for the ST analysis (STAN) monitor, which detects and alerts the potential risk for fetal hypoxia (see e.g., [3] and the citations therein).

There are two main types of fECG signals. The first kind of signal is directly recorded through an electrode attached to the fetal skin. For example, the electrode could be attached to the scalp while the cervix dilates during delivery, which is considered invasive. While the recorded signal is of high quality, it can only be recorded during a specific and short period, and the instrument is not designed for long-term monitoring purposes. The infection risk is also not negligible. Therefore, it is not routinely used in clinics. This is called the *direct fECG signal*, which the STAN monitor depends on. The second kind of signal is recorded from the mother's abdomen, where the sensor is close to the fetus so that the fECG signal is strong enough compared to the maternal ECG. The recorded signal is called the *abdominal ECG* (aECG), which is composed of the maternal cardiac activity, called the *maternal abdominal ECG (maECG)*, and the fetal cardiac activity, called the *indirect fECG signal* (or noninvasive fECG signal). When there is no danger of confusion, we call the indirect fECG signal simply the fECG signal in this paper. Excellent summaries of the available measurement techniques and fECG history (as well as several other topics) are provided in Jenkins [2] and Sameni and Clifford [4].

The aECG signal is non-invasive, easy-to-obtain, and suitable for long-time monitoring purposes. However, from the signal processing prospective, it is challenging to obtain the indirect fECG signal from the aECG signal. For example, the fECG signal is always "contaminated" by or mixed with the maECG, and the signal-to-noise ratio (SNR) is generally low. These issues challenge the estimation of the fECG and hence the HRV analysis from the aECG signal. Furthermore, even if the maECG signal could be successfully decoupled from the fECG signal and perfectly denoised, interpreting the morphology of the fECG signal is still challenging. This issue originates from the individual variation among subjects, e.g., the uterus position and shape, and the fetal size and presentation. Therefore, even if we could standardize the lead system on the mother's abdomen, the application of the fECG waveform is still limited.

The challenge has attracted considerable attention in recent decades, and several algorithms have been proposed. As is summarized in Clifford et al. [5], most methods take the following four steps to study the aECG: first, pre-process the aECG; second, estimate the maECG; third, remove the maECG from the aECG; fourth, post-process the remainder to obtain the fECG and/or estimate the R peaks and hence the fIHR. In short, the maECG is removed first so that the fECG could be analyzed from the remainder. It is possible for the available algorithms to be classified in different ways based on different criteria. However, for the work presented in this paper, we summarize the existing algorithms based on the number of needed leads, and classify them into two categories—one that depends on more than one ECG channel and one that depends on only one aECG signal.

Most algorithms need multiple aECG channels and/or one maternal thoracic ECG (mtECG) signal, or at least one aECG channel and one mtECG; including: blind source separation (BSS) [6–9], semi-BSS like periodic component analysis (πCA), or πTucker decomposition, which takes the pseudo-periodic structure into account [10–12], echo state neural network [13],

least mean square (LMS) [14], recursive least square (RLS) [13], and blind adaptive filtering [15], Kalman filter [16–18], channel selection approach based on features extracted by different methods, like discrete wavelet transform [19], time-adaptive Wiener-filter like filtering [20], principal component regression [21], phase space embedding [22], to name but a few. On the other hand, fewer algorithms depend on the single-lead aECG signal; e.g., template subtraction (TS) [13, 23–26], and its variation based on singular value decomposition (SVD) or principal component analysis [27, 28], the time-frequency analysis, like wavelet transform, pseudo-smooth Wigner-Ville distribution [29–32] (in practice, three aECG channels are averaged in [30]), and S-transform [33], sequential total variation [34], adaptive neuro-fuzzy inference system and extended Kalman filter [35], particle swarm optimization and extended Kalman smoother [36] state space reconstruction via lag map [37, 38], etc. We refer the reader to, e.g., Sameni and Clifford [4] and Andreotti et al. [39] for a more detailed review.

The above-mentioned algorithms all have their own merits and disadvantages; e.g., algorithms depending on multiple leads usually provide a more accurate result, but the dependence on multiple leads render it less applicable for screening and monitoring purposes. On the other hand, the algorithms depending on the single-lead aECG signal usually have lower accuracy, although they could be applied to a wider range of situations. To simultaneously fulfill the practical need and the accuracy, in this paper, we propose a novel algorithm to extract the fetal instantaneous heart rate (fIHR) and the fECG signal from the *single-lead aECG signal* from a different viewpoint. The proposed algorithm combines a recently developed nonlinear time-frequency (TF) analysis called the de-shape short time Fourier transform (de-shape STFT) and the nonlocal median; the de-shape STFT extracts the maternal instantaneous heart rate (mIHR) from the single-lead aECG, which provides the maternal R peak information. The maECG is then extracted from the aECG by the nonlocal median algorithm. The difference between the aECG and estimated maECG serves as a rough fECG estimate. The fIHR could be estimated from the rough estimate of fECG by the de-shape STFT, and hence the fetal R peaks. The fECG is then extracted by the nonlocal median algorithm. The R peak information could be accurately estimated by the beat tracking algorithm based on the dynamic programming. While not explicitly used in the algorithm, we mention that our method has the ability to simultaneously obtain the fIHR and mIHR, and hence simultaneously the fECG and mECG.

The novelty and the main difference between our proposed method and the other algorithms based on the single-lead aECG signal are two folds. First, we use more information from the single-lead aECG signal. Note that the traditional R peak detection algorithms mainly count on the morphological landmarks (fiducial points), such as the maximal points representing the R peaks, or the maximal "energy" pattern driven by the QRS complex in the TF domain determined by, e.g., the wavelet transform. The mIHR and fIHR are then obtained by interpolating the estimated R peak locations. On the other hand, the de-shape STFT allows us to *directly* extract the mIHR and fIHR from the single-lead aECG signal, and

the fIHR from the rough fECG estimate. It is then possible to utilize the mIHR and fIHR to guide an accurate R peak detection. Unlike the traditional approach, we simultaneously use the frequency information (the mIHR and fIHR), which reflects the time-varying and nonlinear beat-to-beat relationship, and the morphological landmark information. Second, based on the nonlinear manifold model, we apply the nonlocal median algorithm [40, 41] to extract the maECG and fECG signals. For each cardiac activity candidate, we only consider those aECG segments with a similar pattern, and use the median to estimate the underlying cardiac activity. Compared with the traditional TS methods [27], where the mean, or the mean together with the first few principal components, of consecutive aECG segments containing cardiac activities is considered to be the template of the cardiac activity, the nonlocal median algorithm handles the following commonly encountered issues. The fact that the QRST complex morphology is time-varying [42] might be overlooked in the traditional TS procedure; the mean of consecutive aECG segments containing cardiac activities is well known to be sensitive to outliers; the TS algorithm is sensitive to the number of principal components and an empirical optimization is needed [28]. The nonlocal median algorithm, on the other hand, could bypass these limitations. In summary, the de-shape STFT is applied to get a better mIHR and fIHR and hence maternal and fetal R peaks, and the nonlocal median is applied to get a better mECG and fECG.

The paper is organized as follows. In Section 2, we discuss a *phenomenological model* for the aECG and the mathematical background for the de-shape STFT and nonlocal median. In Section 3, the single-lead fECG extraction algorithm is introduced. The material and results are reported in Section 4. In Section 5, the paper is summarized by a discussion, including limitations and future works.

2. MATHEMATICAL BACKGROUND

We discuss the mathematical background in this section. Readers who are only interested in the application of the theory and the practical performance of the algorithm can go directly to Section 3 and Section 4. The aECG signal contains at least two components of interest: the component associated with the maternal cardiac activity and the component associated with the fetal cardiac activity. They have different time-varying frequencies and different non-sinusoidal oscillations, and hence a wide spectrum, so the linear signal processing techniques do not work. While it is challenging enough to separate these two components, the problem becomes more challenging considering the influence of different kinds of noise in the measurement. Furthermore, due to the physiological nature of the ECG signal, the non-sinusoidal oscillation is not just impulse-like but also time-varying (see e.g., the nonlinear relationship between QT and RR intervals [42]). As a result, it will encounter a time-varying wave-shape function issue. In this section, we provide a phenomenological model [43] suitable for the aECG signal and an algorithm suitable for analyzing this kind of signal. Second, we provide a low dimensional and nonlinear geometric model to describe the maternal and fetal cardiac activities. Based on this

nonlinear model, the nonlocal median algorithm is introduced to reconstruct the time-varying wave-shape function, and hence extract the mECG and fECG from the aECG.

2.1. Adaptive Non-harmonic Model and De-Shape STFT

We propose to apply the *adaptive non-harmonic model* [43] to model the aECG signal, and apply the de-shape STFT to extract fECG from the aECG.

2.1.1. Adaptive Non-harmonic Model

We start from introducing the adaptive non-harmonic model. Take a small enough $\epsilon > 0$, a non-negative sequence $c = \{c(\ell)\}_{\ell=0}^{\infty}$, $0 < C < \infty$ and $N \in \mathbb{N}$. The set of functions $\mathcal{D}_{\epsilon}^{c,C,N} \subset C^1(\mathbb{R}) \cap L^{\infty}(\mathbb{R})$ consists of functions:

$$x(t) = \frac{1}{2}B_0(t) + \sum_{\ell=1}^{\infty} B_\ell(t)\cos(2\pi\phi_\ell(t)) \tag{1}$$

satisfying the following three conditions. First, the *regularity condition* says that $B_\ell \in C^1(\mathbb{R}) \cap L^{\infty}(\mathbb{R})$ for each $\ell = 0,\dots\infty$ and $\phi_\ell \in C^2(\mathbb{R})$ for each $\ell = 1,\dots\infty$. For all $t \in \mathbb{R}$, $B_\ell(t) \geq 0$ for all $\ell = 0,1,2,\dots,\infty$ and $\phi'_\ell(t) > 0$ for all $\ell = 1,\dots,\infty$. Second, the *time-varying wave-shape* condition says that for all $t \in \mathbb{R}$, $\left|\phi'_\ell(t) - \ell\phi'_1(t)\right| \leq \epsilon\phi'_1(t)$ for all $\ell = 1,\dots,\infty$, $B_\ell(t) \leq c(\ell)B_1(t)$ for all $\ell = 0,1,\dots,N$, $\sum_{\ell=N+1}^{\infty} B_\ell(t) \leq \epsilon\sqrt{\frac{1}{4}B_0(t)^2 + \frac{1}{2}\sum_{\ell=1}^{\infty}B_\ell(t)^2}$, and $\sum_{\ell=1}^{\infty} \ell B_\ell(t) \leq C\sqrt{\frac{1}{4}B_0(t)^2 + \frac{1}{2}\sum_{\ell=1}^{\infty}B_\ell(t)^2}$. Third, the *slowly varying* condition says that for all $t \in \mathbb{R}$, $|B'_\ell(t)| \leq \epsilon c(\ell)\phi'_1(t)$, $|\phi''_\ell(t)| \leq \epsilon\ell\phi'_1(t)$, for each $\ell = 0,\dots\infty$, and $\|\phi'_1(t)\|_{L^{\infty}} < \infty$.

We call a function x in $\mathcal{D}_{\epsilon}^{c,C,N}$ an *adaptive non-harmonic (ANH) function*, where the adjective *non-harmonic* indicates the possibly non-sinusoidal nature of the oscillation, and the adjective *adaptive* indicates the time-varying nature of the frequency, amplitude, and the non-sinusoidal oscillatory pattern. We call $B_1(t)\cos(2\pi\phi_1(t))$ the *fundamental component*, $B_1(t)$ the *fundamental amplitude*, ϕ_1 the *phase function*, and ϕ'_1 the *fundamental instantaneous frequency (IF)* of the signal x. By a slight abuse of terminology, for $\ell > 1$, we call $B_\ell\cos(2\pi\phi_\ell(t))$ the *ℓ-th multiple* of the fundamental component, which we simply call the ℓ-th multiple if no danger of confusion is possible, $B_\ell(t)$ the *amplitude of the ℓ-th multiple*, and ϕ'_ℓ the *IF of the ℓ-th multiple*, although ϕ_ℓ might not be an exact integral multiple of ϕ_1. Thus, we could view an ANH function as an oscillatory component with the *time-varying amplitude, frequency, and wave-shape function*.

A special case deserves a discussion. When $\beta_\ell := \frac{B_\ell(t)}{B_1(t)}$ are constants for all $\ell = 0,1,\dots,\infty$ and $\phi'_\ell(t) = \ell\phi'_1(t) + \alpha_\ell$ for some $\alpha_\ell \in \mathbb{R}$ for all $\ell = 1,\dots,\infty$, (1) is reduced to the $x(t) = B_1(t)s(\phi_1(t))$, where s is a 1-periodic function with the Fourier coefficients determined by β_ℓ and α_ℓ. In this case, clearly the phase function ϕ_1 is linearly related to the deformation of the non-sinusoidal oscillation, and we say that the wave-shape function is *not* time-varying. On the contrary, for an ANH

function, the phase function ϕ_1 might be nonlinearly related to the deformation of the non-sinusoidal oscillation. If we further assume that $\epsilon = 0$, then we obtain the harmonic function.

To motivate this model, take the relationship between the RR and QT intervals of an ECG signal as an example. The nonlinear relationship between the QT interval and the RR interval has been well studied—e.g., a fully nonlinear depiction [42]. Thus, the QT interval corresponding to the ventricular response is not linearly related to the instantaneous heart rate, and we need a time-varying wave-shape function to model this physiological fact. For more detailed discussion, we refer the reader to Lin et al. [43].

In general, a signal might be composed of more than one oscillatory component; e.g., the aECG signal is composed of the fECG and maECG. Take a small enough $0 < \epsilon < 1$ and $d > 0$. We consider the set $\mathcal{D}_{\epsilon,d} \subset C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ consisting of superposition of ANH functions; that is,

$$f(t) = \sum_{k=1}^{K} f_k(t), \quad f_k(t) = \sum_{\ell=0}^{\infty} B_{k,\ell}(t) \cos(2\pi\phi_{k,\ell}(t)) \in \mathcal{D}_\epsilon^{c_k, C_k, N_k}$$

(2)

for some finite $K \in \mathbb{N}$, for some $0 \leq \epsilon_k \leq \epsilon$, non-negative sequence $c_k = \{c_k(\ell)\}_{\ell=0}^{\infty}$, $0 < C_k < \infty$ and $N_k \in \mathbb{N}$, where the fundamental IF's of all ANH functions satisfy the following two conditions if $K > 1$. First, the *frequency separation* condition says that $\phi'_{k,1}(t) - \phi'_{k-1,1}(t) \geq d$ for $k = 2, \ldots, K$. Second, the *non-multiple* condition says that for each $k = 2, \ldots, K$, $\phi'_{k,1}(t)/\phi'_{\ell,1}(t)$ is not an integer for $\ell = 1, \ldots, k-1$. We say that a signal in $\mathcal{D}_{\epsilon,d}$ satisfies the *ANH model*.

2.1.2. Model the Maternal Abdominal ECG Signal by the ANH Model
We now model a recorded aECG signal, $x(t)$, by the ANH model, whichs satisfies:

$$x(t) = x_m(t) + x_f(t) + n(t),$$

(3)

where $x_m(t) \in \mathcal{D}_\epsilon^{c_m, C_m, N_m}$ for some c_m, C_m, N_m is the maECG signal, $x_f(t) \in \mathcal{D}_\epsilon^{c_f, C_f, N_f}$ for some c_f, C_f, N_f is the fECG signal, and $n(t)$ is noise. Here, the fundamental IF of x_m (respectively x_f) is the mIHR (respectively fIHR) of maECG (respectively fECG). $n(t)$ includes different kinds of noise, ranging from the baseline wandering, power line interference, maternal electromyographic signal, to uterine contraction, so in general the noise is not stationary. We consider either smooth varying non-stationary noise model with a smooth and slowly varying covariance function for $n(t)$ to capture the possible heteroscedasticity and autocorrelation inside the noise [44, Equation 6], or a more general piecewise locally stationary model [45, Definition 1] to further capture the abrupt change inside the noise structure.

2.1.3. De-Shape STFT
There are several challenges in analyzing the aECG signal, including the time-varying amplitude, time-varying frequency, and the time-varying non-sinusoidal oscillation. To deal with this kind of signal, under the ANH model (3), we could apply the

currently proposed algorithm, the de-shape STFT. The de-shape STFT is a nonlinear TF analysis technique combining the well-known STFT and the cepstrum technique commonly applied in the signal processing field, and it is composed of the following four steps. First, with a chosen window function $h \in \mathcal{S}$, where $\mathcal{S}$ is the Schwartz space, we have the STFT of $x(t)$:

$$V_x^{(h)}(t,\xi) = \int x(\tau)h(\tau - t)e^{-i2\pi\xi(\tau-t)}\,\mathrm{d}\tau,$$

(4)

where $t \in \mathbb{R}$ indicates time and $\xi \in \mathbb{R}$ indicates frequency. Second, evaluate the *short time cepstral transform (STCT)* in order to obtain the fundamental period and its multiples:

$$C_x^{(h,\gamma)}(t,q) := \int |V_x^{(h)}(t,\xi)|^\gamma e^{-i2\pi q\xi}\,\mathrm{d}\xi,$$

(5)

where $\gamma > 0$ is sufficiently small and $q \in \mathbb{R}$ is called the quefrency (its unit is second or any feasible unit in the time domain). Third, we remove all the multiples by evaluating the de-shape STFT, which is defined on $\mathbb{R} \times \mathbb{R}^+$ as:

$$W_x^{(h,\gamma)}(t,\xi) := V_x^{(h)}(t,\xi)C_x^{(h,\gamma)}(t,1/\xi),$$

(6)

where $\xi > 0$ is interpreted as frequency. We refer the reader to Lin et al. [43, Section 3] for a discussion of the well-definedness of the de-shape STFT.

The main motivation of the de-shape STFT is to decouple the IF and the non-sinusoidal wave-shape function. Due to the non-sinusoidal oscillation, at each time t, in the STFT we could see not only the fundamental frequency but also its multiples. The existence of multiples, when there is more than one component, interfere with each other and mask the true information we are interested in. Therefore, the STCT is applied to obtain the fundamental period information of the signal. Note that the fundamental period and its multiples, after the inversion, become the fundamental frequency and its divisions, and hence, the common ingredient between $C_x^{(h,\gamma)}(t,1/\xi)$ and $V_x^{(h)}(t,\xi)$ is the fundamental frequency. Thus, after a direct element-wise product, only the fundamental frequency is preserved in $W_x^{(h,\gamma)}(t,\xi)$. This approach could be viewed as a "nonlinear masking/filtering" technique, which uses the "dual information" of the spectrum (the cepstrum), as a mask to remove the irrelevant information (the wave-shape function) and keep the relevant information (the IF).

It has been shown in Lin et al. [43, Theorem 3.6] that under the ANH model (2), the de-shape STFT could extract the fundamental IF of each component, where all multiple IF's are suppressed. In other words, the time-varying wave-shape information of maECG and fECG is decoupled from the IF and AM of maECG and fECG in the TF representation. The implementation of the de-shape STFT in the discrete-time domain will be discussed in Section 3.

2.2. Nonlocal Median
Nonlocal median algorithm [40, 41] is a variation of the nonlocal mean algorithm in the image processing, mainly for the purpose of image denoising [46, 47]. After obtaining the fundamental IF

of each component, we apply the nonlocal median algorithm to recover the time-varying wave-shape function. In this study, it allows us to extract the maECG and fECG from the aECG signal.

Without loss of generality, we assume that maECG and fECG are both with the Rs pattern in the aECG so that we could discuss the R peaks. The discussion holds for the S peaks if either maECG or fECG has the rS pattern. Take the maECG into account. Suppose the i-th cardiac activity, which could be normal or ectopic, without any other pathological arrhythmia, in the maECG starts at $s_m^{(i)} \in \mathbb{R}$ and ends at $e_m^{(i)} \in \mathbb{R}$, where s means "start" and e means "end," and suppose the i-th R peak is located at $r_m^{(i)} \in (s_m^{(i)}, e_m^{(i)})$. By the physiological property of the cardiac activity, we know that $e_m^{(i)} < s_m^{(i+1)}$ for all i, where $[e_m^{(i)}, s_m^{(i+1)}]$ are periods where the maECG is isoelectric (known as the TP interval). Thus, the ANH function x_m in Equation (3) could be written as:

$$x_m(t) = \sum_{i \in \mathbb{Z}} x_m(t) \chi_{[s_m^{(i)}, e_m^{(i)}]}(t), \tag{7}$$

where $\chi_{[s_m^{(i)}, e_m^{(i)}]}$ is an indicator function defined on $[s_m^{(i)}, e_m^{(i)}]$. Note that due to the time-varying nature of the non-sinusoidal oscillation, the amplitude of the fundamental component and each multiple in the ANH function is nonlinearly related to the waveform morphology. Consider:

$$l_m := \max_i \{r_m^{(i)} - s_m^{(i)}\} > 0 \text{ and } r_m := \max_i \{e_m^{(i)} - r_m^{(i)}\} > 0, \tag{8}$$

where l means "left" and r means "right." Denote the i-th abdominal cardiac activity mixture as $x_m^{(i)} : [0, l_m + r_m] \to \mathbb{R}$, where:

$$x_m^{(i)}(t) = \begin{cases} x(r_m^{(i)} + t - l_m) & \text{when } t \in [l_m - (r_m^{(i)} - s_m^{(i)}), l_m \\ & \qquad\qquad + (e_m^{(i)} - r_m^{(i)})] \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Clearly, $x_m^{(i)}$ is the i-th aECG segment containing the i-th maternal cardiac activities, denoted as $x_{m,m}^{(i)} : [0, l_m + r_m] \to \mathbb{R}$, where:

$$x_{m,m}^{(i)}(t) = \begin{cases} x_m(r_m^{(i)} + t - l_m) & \text{when } t \in [l_m - (r_m^{(i)} - s_m^{(i)}), \\ & \qquad\qquad l_m + (e_m^{(i)} - r_m^{(i)})] \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

and several fetal heart beats, since normally the fetal heart rate is higher. Denote:

$$\mathcal{X}_m := \{x_m^{(i)}\}_{i \in \mathbb{N}} \subset C^1([0, l_m + r_m]) \tag{11}$$

to be collection of aECG segments from $x(t)$ and

$$\mathcal{X}_{m,m} := \{x_{m,m}^{(i)}\}_{i \in \mathbb{N}} \subset C^1([0, l_m + r_m]) \tag{12}$$

to be collection of maECG segments from the underlying maternal cardiac activities. We thus view $\mathcal{X}_m$ as a "noisy" collection of maECG segments $\mathcal{X}_{m,m}$, and the mission is to recover $x_{m,m}^{(i)}$ from $x_m^{(i)}$ for all i.

It is well known that electrophysiologically, the fECG and maECG are similar, except the heart rate [4], so the above discussion could be carried over to the fECG, x_f. We could thus define $s_f^{(i)}, e_f^{(i)} \in \mathbb{R}$ and $r_f^{(i)} \in (s_f^{(i)}, e_f^{(i)})$ for the i-th cardiac activity in the fECG, where the i-th cardiac activity starts at $s_f^{(i)}$ and ends at $e_f^{(i)} \in \mathbb{R}$, and the i-th R peak is located at $r_f^{(i)}$. Again, we know that $e_f^{(i)} < s_f^{(i+1)}$ for all i, where $[e_f^{(i)}, s_f^{(i+1)}]$ are periods where the fECG is isoelectric. Similarly, we could define $\mathcal{X}_f = \{x_f^{(i)}\}_{k \in \mathbb{N}} \subset C^1([0, l_f + r_f])$ and $\mathcal{X}_{f,f} = \{x_{f,f}^{(i)}\}_{i \in \mathbb{N}} \subset C^1([0, l_f + r_f])$, where $l_f := \max_i \{r_f^{(i)} - s_f^{(i)}\} > 0$, $r_f := \max_i \{e_f^{(i)} - r_f^{(i)}\} > 0$, $x_f^{(i)}(t) = x(r_f^{(i)} + t - l_f)$ when $t \in [l_f - (r_f^{(i)} - s_f^{(i)}), l_f + (e_f^{(i)} - r_f^{(i)})]$ and $x_f^{(i)}(t) = 0$ otherwise, and $x_{f,f}^{(i)}(t) = x_f(r_f^{(i)} + t - l_f)$ when $t \in [l_f - (r_f^{(i)} - s_f^{(i)}), l_f + (e_f^{(i)} - r_f^{(i)})]$ and $x_{f,f}^{(i)}(t) = 0$ otherwise.

Physiologically, it is well known that while the underlying mechanism leading to the cardiac activities might be complicated [39], phenomenologically they are similar from beat to beat. There are two dominant parameters that quantify the similarity between cardiac activities—the scaling and the dilation, where the scaling reflects the respiratory activity and the dilation reflects the nonlinear relationship between the RR interval and QT interval. Also, the waveform representing the cardiac activity should be bounded and with a bounded differentiation. This fact could be summarized in the following:

Assumption 2.1. $\mathcal{X}_{m,m}$ is sampled from a random vector V_m, where V_m has the range supported on a bounded set inside $C^1([0, l_m + r_m])$ with a low dimensional structure. To simplify the qualitative description "low dimensional structure," we assume a low dimensional smooth and compact manifold to quantify the range of V_m.

Due to the fact that the maECG amplitude and frequency are both time-varying, under this model, two consecutive maternal cardiac activities might be far away in the manifold.

Similarly, we have the same assumption for the fetal cardiac activity; that is, the set $\mathcal{X}_{f,f}$ is sampled from a random vector V_f, with the range supported on a bounded set inside $C^1([0, l_f + r_f])$ with a low dimensional structure. A critical assumption we need to apply the nonlocal median algorithm is the following

Assumption 2.2. V_m and V_f are independent.

This assumption essentially says that for different $x^{(i)}$, while the maternal cardiac activities are similar, the fetal cardiac activities sit in random positions.

With the above setup and assumptions, we could now introduce the nonlocal median algorithm. For each $x_m^{(i)} \in \mathcal{X}_m$, find its K nearest neighbors with the L^2 norm, where $K \in \mathbb{N}$ is chosen by the user. Precisely, by ranking $d_{i,j} := \|x_m^{(i)} - x_m^{(j)}\|_{L^2}$ in the ascending order, we have the set $\mathcal{N}_m^{(i)}$ containing the first K neighbors of $x_m^{(i)}$ in $\mathcal{X}_m$ with the smallest L^2 norm. Note that we could also consider the correlation or other more sophisticated metrics, but to keep the discussion simple, we focus on the L^2

norm in this paper. Then, the i-th maternal cardiac activity is estimated by:

$$\tilde{x}_{m,m}^{(i)}(t) := \text{median}\{x_m^{(j)}(t) \mid x_m^{(j)} \in \mathcal{N}_m^{(i)}\} \quad (13)$$

for all $t \in [0, l_m + r_m]$. Based on Assumption 2.1, the neighbors we find have similar maternal cardiac activities while they are not neighbors in the temporal axis. On the other hand, it is well known that the median is robust to outliers. Note that under Assumption 2.2, for different $x^{(i)}$, the fetal cardiac activities sit in different positions in the support of $x^{(i)}$, and note that the ECG morphology associated with the ventricular activity is spiky, so the fetal cardiac activity occupies a small portion of the interval $[0, l_m + r_m]$. As a result, the median value will faithfully reflect the maternal cardiac activity at t, and hence the recovery of the maECG.

The above procedure could be applied to extract the fECG signal from the aECG signal, if we reverse the role of the maECG and fECG, and consider $\mathcal{X}_f$ from $\mathcal{X}_{f,f}$.

3. METHODOLOGY

In this section, we introduce our single-lead fECG extraction algorithm, and provide the statistics used for the evaluation.

3.1. Single-Channel Fetal ECG Extraction Algorithm

We now introduce our algorithm to extract the fECG signal from the aECG signal. The algorithm is summarized in the flowchart in **Figure 1**. We detail the algorithm step by step below. The algorithm is implemented in Matlab, and the code can be found in https://sites.google.com/site/hautiengwu/.

The continuous aECG signal $x(t)$ is sampled with the sampling rate $f_s \in \mathbb{N}$ during the interval from the 0-th second to the T-th second, where $T > 0$. Denote:

$$\mathbf{x}_0 := [x(0), x(1/f_s), \ldots, x(\lfloor T/f_s \rfloor)]^T \in \mathbb{R}^N, \quad (14)$$

where $N = \lfloor T/f_s \rfloor + 1$, to be the collected aECG signal. Thus,

$$\mathbf{x}_0 = \mathbf{x}_m + \mathbf{x}_f + \mathbf{n}, \quad (15)$$

where $\mathbf{x}_m$, $\mathbf{x}_f$, and $\mathbf{n}$ are the discretized maECG signal, fECG signal and noise.

3.1.1. Step 0: Preprocessing

The first step of the proposed algorithm, as in most other algorithms, is signal preprocessing. For all of the analyses, we apply the following steps. First, we remove the baseline wandering by the median filter with the window size 100 ms. If needed, the power-line interference is suppressed by a zero-phase notch filter at 50 or 60 Hz. To preserve the non-stationary nature of the signal, we do not apply any other linear filtering technique to the signal. If the signal is sampled at a frequency lower than 1000 Hz, to enhance the R peak alignment needed in the nonlocal median step, the signal is upsampled to 1000 Hz [48]. To simplify

the notation, we use the same notation f_s and N to denote the resulting sampling rate and the resulting number of sampling points, and denote the resulting signal as

$$\mathbf{x} \in \mathbb{R}^N. \quad (16)$$

3.1.2. Step 1: Run De-Shape STFT on x to Estimate the Maternal Instantaneous Heart Rate

We apply the de-shape STFT algorithm to extract the IF information of the maternal and fetal cardiac activities from the preprocessed aECG signal $\mathbf{x}$. We fix the frequency resolution of the STFT by $\frac{f_s}{2M}$, where $M \in \mathbb{N}$ is the number of discretization points in the frequency axis, and the quefrency resolution by $\frac{M}{f_s M'}$, where $M' \in \mathbb{N}$ is the number of discretization points in the quefrency axis. The numerical implementation of the de-shape STFT algorithm is summarized in Algorithm 1 (Supplementary Material). Denote $\mathbf{W_x} \in \mathbb{C}^{N,M}$ to be the de-shape STFT of $\mathbf{x}$. The temporal complexity of evaluating the de-shape STFT is $O(N(M \log M + M' \log M'))$.

It has been systematically reported in Lin et al. [43] that the fundamental frequencies of the maECG and fECG are represented as dominant curves in the TFR. This fact allows us to extract the salient fundamental IF information of each oscillatory component [43]. The feature inside TFR determined by the de-shape STFT is suitable for several IF tracking methods, including dynamic programming (DP), dynamic Bayesian networks, adaptive filters, and others. However, to simplify the discussion, we apply the simple DP curve extraction algorithm to track the peaks and estimate the IFs based on the assumption that the maECG is stronger than the fECG. We perform DP on $\mathbf{W_x}$ to extract the most dominant component. The extracted curve, when adjusted with the frequency resolution, denoted as $\eta_m \in \mathbb{R}^N$, represents the mIHR. Precisely, suppose at time n, the most dominant component is located at $\mathbf{W_x}(n, j_n)$, then $\eta_m(n) = \frac{j_n f_s}{2M}$. The detailed DP algorithm is shown in Algorithm 2 (Supplementary Material)[1].

3.1.3. Step 2: Obtain the Maternal R Peaks by Beat Tracking and Dynamic Programming

Note that although theoretically the IHR is related to the R peak to R peak interval (RRI) time series, in practice, there is a discrepancy due to the time-varying nature of the wave-shape function. To obtain the exact R peak location, we apply the beat tracking technique, which has been well studied in music signal analysis[2].

We define the estimated maternal RRI as $\delta_m(n) := 1/\eta_m(n)$, which is the inverse of the mIHR. Our goal is to find an strictly increasing sequence $B_m = \{b_i\}_{i=1}^{M_m}$ of length M_m, where $M_m \in \mathbb{N}$

---

[1]Optional step: simultaneously estimate the fIHR) For every time n, multiply $\mathbf{W}(n, j)$ by θ_m, where $j \in \{\eta_m(n) - N_\Delta, \eta_m(n) - N_\Delta + 1, \ldots, \eta_m(n) + N_\Delta - 1, \eta_m(n) + N_\Delta\}$ and $N_\Delta \in \mathbb{N}$ and $0 \leq \theta_m < 1$ are chosen by the user, to suppress the maternal cardiac activity. This procedure makes the fECG be the predominant component in $\mathbf{W_x}$. Then, performing the curve extraction on $\mathbf{W_x}$ again to extract the fIHR, denoted as $\eta_f \in \mathbb{R}^N$. In practice, we could choose N_Δ so that $\frac{N_\Delta f_s}{2M} = 0.1$ Hz and $\theta_m = 10^{-4}$. This optional step is not carried out in this paper.

[2]The *beat tracking* problem in music refers to finding the instants of beats in music by analyzing the accents of music signals (e.g., drum hits).

FIGURE 1 | The flow chart of the proposed algorithm for extracting fECG from the signal lead aECG signal. The shown aECG signal is of 6 s long.

and $0 < b_i \leq N$ so that b_i is the index of i-th maternal R peak position and $(b_i - b_{i-1})/f_s$ is close to the estimated maternal RRI at time b_i/f_s, $\delta_m(b_i)$. We call B_m the *maternal beat sequence*. This problem is understood as the beat tracking problem, which

is formulated as Ellis [49]:

$$\tilde{B}_m = \underset{B_m}{\arg\max} \left[\sum_{i=1}^{M_m} \mathbf{x}(b_i) + \lambda_{\mathrm{BT}} \sum_{i=2}^{M_m} P(b_i, b_{i-1}) \right], \quad (17)$$

where $P(b_i, b_{i-1}) := -\left(\log_2\left(\frac{(b_i-b_{i-1})/f_s}{\delta_m(b_i)}\right)\right)^2$ and $\lambda_{BT} \geq 0$ is the penalty term determined by the user. Notice that Ellis [49] assumes $\delta_m(n)$ (the inter-beat interval in the music clip; in our case, the IHR) to be constant, which is clearly not suitable for an oscillatory signal with time-varying frequency. To address this issue, we modify the original formulation in Ellis [49] by allowing a time-varying function $1/\eta_m(n)$. Note that the first term in the objective function in Equation (17) reaches a maximum when b_i matches the R-peak location, and the second term penalizes the discrepancy between the estimated IHR and the IHR determined by the true RR interval, with the maximal value zero. The resulting beat sequence $\tilde{B}_m = \{\tilde{b}_i\}_{i=1}^{M_m}$ provides an estimate of R peaks location; that is, $\tilde{b}_i$ is an estimate of the i-th R peak location. In our experiments, setting λ_{BT} between 20 and 50 turns out to yield a suitable tradeoff, and the result is not sensitive to λ_{BT}. While the number of possible beat sequences in $\mathbf{x}$ grows exponentially as N grows, the optimization problem (17) can be solved effectively by the DP.

The main idea leading to the DP is that the objective function in Equation (17) is an accumulation in time. Thus, we could divide the problem into a set of optimization subproblems, each of which optimizes the objective function up to a prescribed time step, and the solution is the combination of the optima at every time step. This is implemented by introducing two additional vectors $\mathbf{C} \in \mathbb{R}^N$ and $\mathbf{D} \in \mathbb{R}^N$, where $\mathbf{C}(n)$ records the maximum of the objective function accumulated from 1 to n, where $n = 1, \ldots, N$, and $\mathbf{D}(n)$ records the estimated beat position yielding this maximum at the time step of $n-1$. $\mathbf{D}$ allows us to trace previous beat positions, and this step is called the *backtracking*. The whole procedure is sketched in Algorithm 2 (Supplementary Material). More details of this method can be found in Ellis [49], and the source code is available in http://labrosa.ee.columbia.edu/projects/beattrack/.

To avoid the possible cardiac axis deviation problem, we correct the polarity of the ECG signal. Denote $\tilde{B}_m^- = \{\tilde{b}_i^-\}_{i=1}^{M_m^-}$ to be the estimated locations of R peaks of $-\mathbf{x}$ by the beat tracking algorithm, where M_m^- is the number of estimated R peaks. The polarity of the ECG signal is determined to be positive if the median value of $\{|\mathbf{x}(\tilde{b}_i)|\}_{i=1}^{M_m}$ is greater than the median value of $\{|\mathbf{x}(\tilde{b}_i^-)|\}_{i=1}^{M_m^-}$; otherwise the polarity is negative. If the polarity of $\mathbf{x}$ is negative, multiply $\mathbf{x}$ by -1. To simplify the notation, we use the same notation $\mathbf{x}$ to denote the polarity-corrected ECG signal and $\tilde{B}_m$ to denote $\tilde{B}_m^-$.

3.1.4. Step 3: Estimate the maECG Morphology by the Nonlocal Median

With the maternal R peak location, we could extract $\mathbf{x}_m$ from the aECG signal. Based on the physiological knowledge, we choose large enough $L_m, R_m \in \mathbb{N}$ so that $[\frac{\tilde{b}_i - L_m}{f_s}, \frac{\tilde{b}_i + R_m}{f_s}]$ is long enough to cover the i-th maternal cardiac activity for $i = 1, \ldots, M_m$. Define the aECG segments $\mathbf{x}^{(i)} \in \mathbb{R}^{L_m + R_m + 1}$, $i = 1, \ldots, M_m$, as:

$$\mathbf{x}^{(i)} := \left[\mathbf{x}(\tilde{b}_i - L_m) \ldots \mathbf{x}(\tilde{b}_i) \ldots \mathbf{x}(\tilde{b}_i + R_m)\right]^t, \quad (18)$$

where the superscript t means taking the transpose. Note that the i-th R peak is located on the $(L_m + 1)$-th entry of all $\mathbf{x}^{(i)}$. Denote $\mathcal{N}_m^{(i)} = \{\mathbf{x}^{(i_1)}, \ldots, \mathbf{x}^{(i_{K_m})}\}$ to be the first K_m nearest neighbors of $\mathbf{x}^{(i)}$ with respect to the L^2 norm, where $K_m \in \mathbb{N}$ is chosen by the user. The i-th maternal cardiac activity is thus estimated by:

$$\tilde{\mathbf{x}}_m^{(i)}(l) := \text{median}\{\mathbf{x}^{(i_j)}(l)\}_{j=1}^{K_m}, \quad (19)$$

where $l = 1, \ldots, L_m + R_m + 1$.

Before estimating the maECG from $\{\tilde{\mathbf{x}}_m^{(i)}\}_{i=1}^{M_m}$, we need to take care of the possible overlapping issue. If $\tilde{\mathbf{x}}_m^{(i)}$ overlaps with its neighboring segments $\tilde{\mathbf{x}}_m^{(i+1)}$, we need to taper the overlapping regions of both segments. See Algorithm 3 (Supplementary Material) for the implementation of the tapering step. To simplify the notation, we use the same notation $\tilde{\mathbf{x}}_m^{(i)}$ to denote the tapered segment. Finally, the maECG is estimated by $\tilde{\mathbf{x}}_m \in \mathbb{R}^N$, where:

$$\tilde{\mathbf{x}}_m(\hat{b}_i + j) = \sum_{i=1}^{M_m} \tilde{\mathbf{x}}_m^{(i)}(j + L_m + 1) \quad (20)$$

for all $i = 1, \ldots, M_m$ and $j = -L_m, \ldots, R_m$, and zero otherwise. In practice, the result is stable when K_m ranges from 20 to 60. The algorithm is shown in Algorithm 3 (Supplementary Material).

3.1.5. Step 4: Get the fIHR and Obtain the fECG Signal

Denote

$$\tilde{\mathbf{x}}_{f,0} := \mathbf{x} - \tilde{\mathbf{x}}_m \quad (21)$$

to be the *rough fECG* estimate. The fIHR, fetal R peaks, and fECG could be obtained by repeating Step 1-3 on the rough fECG estimate. Precisely, the fIHR could be obtained by running the de-shape STFT on $\tilde{\mathbf{x}}_{f,0}$ and another DP curve extraction on $\mathbf{W}_{\tilde{\mathbf{x}}_{f,0}}$ in Step 1. Before running the DP curve extraction, we could re-weight $\mathbf{W}_{\tilde{\mathbf{x}}_{f,0}}$ around the band associated with the mIHR η_m by a small constant $\theta_f > 0$ to reduce to the possible impact of the remaining maECG component. Precisely, set $\mathbf{W}_{\tilde{\mathbf{x}}_{f,0}}(n, j_n)$ to be $\theta_f \mathbf{W}_{\tilde{\mathbf{x}}_{f,0}}(n, j_n)$, where $n = 1, \ldots, N$ and $j_n \in \{\eta_m(n) - N_\Delta, \eta_m(n) - N_\Delta + 1, \ldots, \eta_m(n) + N_\Delta - 1, \eta_m(n) + N_\Delta\}$. In practice, choosing $\theta_f = 1/10$ could slightly improve the result. Although the improvement is marginal, we recommend to take it into account.

Denote $\eta_f \in \mathbb{R}^N$ to be the final estimated fIHR. The fetal R peak location could be further refined by running the beat tracking technique by the DP and the polarity correction. Denote the final estimated fetal R peaks by a strictly increasing sequence $\tilde{B}_f$. The fECG signal could be reconstructed by the nonlocal median algorithm, with $L_f, R_f \in \mathbb{N}$ chosen based on the physiological fact and $K_f \in \mathbb{N}$ nearest neighbors chosen by the user. Denote the final reconstructed fECG as $\tilde{\mathbf{x}}_f \in \mathbb{R}^N$. In practice, we found that the result is stable when K_f ranges from 20 to 60.

3.1.6. Final Step

The final step takes the physiological constraint into account. If the average η_f determined from the rough fECG signal is smaller

than η_m, the final reconstructed maECG and fECG could be exchanged to respect the physiological constraint that normally the mIHR is slower than fIHR. When both the fetus and mother are healthy, this step could help when the fECG is stronger than the maECG, a case which is not commonly seen.

3.2. Evaluation Statistics

The R peak detection result is evaluated by beat-to-beat comparisons between the detected beats and the annotation provided by the experts. We follow the suggestion [50] and choose a matching window of 50 ms. Denote TP, FP, TN, and FN to be true positive, false positive, true negative, and false negative. We report the sensitivity (SE),

$$SE = TP/(TP + FN), \tag{22}$$

the positive predictive value (PPV),

$$PPV = TP/(TP + FP), \tag{23}$$

and the F_1 score, which is the harmonic mean of PPV and SE,

$$F_1 = 2TP/(2TP + FN + FP). \tag{24}$$

In addition, the mean absolute error (MAE) of the estimated R peak locations is also reported. To make the evaluation independent of the detection accuracy, we follow the suggestion in Andreotti et al. [39] to calculate the MAE only on TP annotations.

To evaluate the fECG morphology recovery, we consider the correlation between the estimated fECG signal and the true fECG signal, over the period ranging from 80 ms before the annotated R peak and 120 ms after the annotated R peak. We again evaluate the correlation only on TP annotations.

4. MATERIAL AND RESULT

In this section we show the performance of the proposed algorithm. For a fair comparison, the parameters for the algorithm are set to be the same for all signals throughout the paper, unless otherwise stated. The window h is chosen to be the Hamming window of length 5 s, the up-sampling rate $\alpha = 10$, the frequency resolution in STFT and de-shape STFT is set to 0.02 Hz, the quefrency resolution is set to 10 ms, $\gamma = 0.3$ for the STCT, and $\upsilon = 10^{-4}\%$ of the root mean square energy of the signal under analysis for the de-shape STFT. In the beat tracking, λ_{BT} is set to 50. In the nonlocal median, we choose $K_m = K_f = 40$.

4.1. Material

We evaluate the proposed algorithm on three databases of aECG signals. The first database is the simulated fECG signal database (*fecgsym*) [39]. The publicly available simulator generates simultaneously the maECG and fECG signals in 34 channels, which model a number of realistic non-stationary physiological phenomena that affect the morphology and dynamics of the aECG, including different kinds of noise. A total of seven physiological events are introduced in the simulator:

- (Baseline) the baseline abdominal mixture with a constant fIHR and mIHR without noise or events;
- (Case 0) baseline signal contaminated by noise;
- (Case 1) Case 0 is complicated by the fetal movement noise;
- (Case 2) the fIHR and mIHR are time-varying with noise;
- (Case 3) Case 2 is complicated by the uterine contraction noise;
- (Case 4) Case 2 is complicated by ectopic beats in both fECG and maECG;
- (Case 5) twin pregnancy contaminated by noise.

The simulator also generates five different levels of additive noise, ranging from 0, 3, 6, 9, to 12 dB. Each physiological event and noise level were simulated independently five rounds to mimic the realistic situation. A generated benchmark is available in https://physionet.org/physiobank/database/fecgsyndb/ [51], which contains ten subjects, five rounds, and five SNR's for each case. Each simulation is of 5-min long and is discretized at the sampling rate 250 Hz. For each subject, case, round, and SNR, there are one maECG, fECG, two noise realizations, and one extra uterine contraction noise in Case 3 in the database. We sum all these time series together to get the simulated signal for analysis. We test our algorithm on all Cases, all levels of noise, and all five simulations, except Case 5. We consider the twin pregnancy case as an independent project, and the result will be reported in the other work.

The second database is the Abdominal and Direct Fetal Electrocardiogram Database (*adfecgdb*), where the aECG signals with the annotation provided by experts is publicly available https://www.physionet.org/physiobank/database/adfecgdb/ [38, 51]. The reference fetal beat annotations are determined from the direct fECG recorded from the fetal scalp lead. There are five pregnant women between 38 and 40 weeks of pregnancy in this database. Each has 4 aECG channels and one direct fECG signal recorded from the Komporel system (ITAM Institute, Zabrze, Poland http://www.itam.zabrze. pl/developments-english-version-233/665-komporel). The four abdominal leads are placed around the navel, a reference lead is placed above the pubic symphysis, and a common mode reference electrode with active-ground signal is placed on the left leg. The signal is of 5 min long and is sampled at 1000 Hz with 16 bit resolution.

The third database is the 2013 PhysioNet/Computing in Cardiology Challenge (https://physionet.org/challenge/2013/# data-sets), abbreviated as CinC2013, which is composed of three sets, learning (training) set A, open test set B, and hidden test set C, where there are 447 number of records in total. Each recording comes with four aECG channels of length 1-min long sampled at 1,000 Hz at 16-bit resolution. To the best of our knowledge, the lead placement is not available in public. There are 75 recorded segments in set A. Since only set A comes with the fetal beat annotations, we thus used set A for an assessment of our proposed algorithm. Among 447 records, 25 records are from the adfecgdb [52], so there might be overlapping between the CinC2013 and adfecgdb databases. More details about these two databases can be found on the website.

4.2. Results of the First Database: Simulation Fecgsym Database

In this simulated database, since we a priori know that the noise could be as big as 0 dB, we choose a longer window of length 10 s. The other parameters are fixed for all the cases.

Following the recommended report method in Andreotti et al. [39], **Tables 1**, **2** give an overview of the performance of the proposed algorithm. In addition to reporting the 1 min result, we also report the whole 5 min result. For the 1 min (respectively 5 min) result, for each subject, noise level, case, and round, we find the channel and the 1-min subset out of the 5-min signal (respectively the channel and the whole 5-min signal) that gives the highest F_1. For each noise level and case, the median and interquartile (IQR) of those highest F_1 values, and the associated MAE of all subjects and rounds are shown in **Table 1**. To better understand the performance of the proposed algorithm, the whole gross statistics with the *tenth* highest F_1 is reported in **Table 2**.

From these two tables, we could see that as SNR decreases, the performance decreases, which is consistent with the results of different TS techniques reported in Andreotti et al. [39, Figure 4]. It is also clear that the overall accuracy of the proposed algorithm is higher than those of TS techniques and adaptive filtering techniques reported in Andreotti et al. [39, Table 3]. We

also notice that the IQR in our report is higher than those of different TS techniques reported in Andreotti et al. [39, Table 3]. This is caused by the fact that we do not use the ground truth maternal R peak information but estimate it from the aECG signal[3]. We should notice that when SNR decreases, the lost information could not be recovered by only one channel signal. Clearly, in Case 3 where the signal is contaminated by the uterine contraction noise, the results on the 5-min long signal is much lower. This comes from the fact that the energy of the uterine contraction is much larger than both the maECG and fECG, and almost no information could be extracted when the uterine contraction happens. Note that except Case 4, the MAE outperforms the reported results in Andreotti et al. [39, Table 3]. This indicates the strength of beat tracking and nonlocal median. In Case 4, although the MAE is larger than the other cases, it is still on the same level of that reported in Andreotti et al. [39, Table 3].

On the other hand, we see that as compared with the adfecgdb database, the MAE is smaller in this simulation database. Note that this smaller MAE is reasonable. In the real databases, we

---

[3]Note that the purpose of Andreotti et al. [39] is comparing different methods, ranging from different BSS algorithms to single-lead TS algorithms, instead of evaluating simply a specific TS algorithm, but in this paper we only focus on evaluating our proposed algorithm.

TABLE 1 | Results of fetal R peaks estimation—median F_1 and MAE and their IQRs for different cases and noise levels among 10 subjects and 5 rounds from the simulated database *fecgsym*.

		Case 0	Case 1	Case 2	Case 3	Case 4
F_1 (%)						
12 dB	1 min	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
	5 min	99.93 (0.38)	99.92 (0.21)	100 (0.87)	72.96 (10.32)	99.85 (0.47)
9 dB	1 min	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
	5 min	99.9 (0.61)	99.91 (0.29)	99.93 (0.27)	73.6 (9.44)	99.58 (3.75)
6 dB	1 min	100 (0.73)	100 (0.88)	100 (0.67)	100 (0.43)	100 (7.83)
	5 min	99.89 (1.38)	99.82 (11.01)	99.91 (9.56)	73.72 (13.58)	99.31 (22.52)
3 dB	1 min	100 (71.53)	100 (0.88)	100 (69.14)	100 (6.11)	100 (19.82)
	5 min	99.12 (76.49)	98.72 (6.17)	99.44 (78.63)	72.11 (28.53)	98.6 (45.75)
0 dB	1 min	95.63 (56.14)	97.99 (19.31)	97.9 (64.75)	99.61 (37)	93.76 (64.04)
	5 min	88.24 (74.69)	83.46 (38.1)	81.17 (73.67)	64.25 (38.3)	85.84 (74.18)
MAE (ms)						
12 dB	1 min	0.85 (1.54)	0.5 (0.77)	0.68 (1.2)	0.53 (0.52)	2.15 (1.46)
	5 min	0.82 (1.82)	0.58 (0.78)	0.63 (1.59)	3.07 (1.54)	2.32 (1.63)
9 dB	1 min	1.01 (2.83)	0.75 (0.65)	0.7 (0.49)	0.51 (0.58)	3.21 (1.77)
	5 min	0.96 (2.96)	0.71 (0.7)	0.77 (0.37)	3.15 (1.62)	3.2 (1.77)
6 dB	1 min	1.21 (3.04)	0.92 (2.7)	1.06 (2.16)	0.85 (1.01)	4.03 (5.56)
	5 min	1.15 (3.17)	0.93 (2.48)	1 (3.11)	3.43 (1.99)	4.01 (5.52)
3 dB	1 min	2.23 (10.64)	1.43 (3.18)	1.41 (4.15)	1.11 (6.67)	5.13 (11.54)
	5 min	1.9 (12.18)	1.76 (4.16)	1.38 (4.13)	3.77 (8.61)	5.07 (11.89)
0 dB	1 min	5.65 (8.06)	4.36 (8.4)	3.5 (13.46)	2.13 (11.66)	6.6 (13.47)
	5 min	5.27 (7.85)	5.59 (9.47)	4.56 (10.93)	5.53 (11.47)	6.26 (13.34)

The 1 min result is reported in the following way. For each subject, noise level, case, and round, we take the channel and the 1-min subset out of the 5-min signal that leads to the highest F_1 to report the result. Then, for each noise level and each case, the median and IQR of all subjects and rounds are shown as the final result, as is suggested in Andreotti et al. [39]. For the 5 min result, for each subject, noise level, case, and round, we take the channel that leads to the highest F_1, and then for each noise level and each case, we report the median and IQR of all subjects and rounds.

TABLE 2 | Results of fetal R peaks estimation for the *tenth* best channel—median F_1 and MAE and their IQRs for different cases and noise levels among 10 subjects and 5 rounds from the simulated database *fecgsym*.

		Case 0	Case 1	Case 2	Case 3	Case 4
F_1 (%)						
12 dB	1 min	100 (0)	100 (0)	100 (0.36)	100 (0.4)	100 (1.74)
	5 min	99.82 (9.28)	99.88 (0.35)	99.39 (14.41)	64.17 (19.45)	99.02 (16.1)
9 dB	1 min	100 (1.25)	100 (74.63)	100 (4.62)	100 (2.08)	99.2 (3.13)
	5 min	99.48 (5.48)	99.53 (78.69)	99.48 (27.05)	64.61 (17.64)	97.71 (10.31)
6 dB	1 min	100 (10.37)	100 (1.64)	99.23 (64.58)	99.6 (14.44)	98.94 (12.78)
	5 min	96.02 (43.21)	98.06 (6.49)	94.48 (74.91)	60.57 (21.72)	93.71 (28.81)
3 dB	1 min	97.89 (59.78)	99.18 (9.54)	99.8 (80.33)	99.7 (66.55)	94.79 (29.41)
	5 min	87.54 (74.19)	95.27 (21.42)	78.4 (90.88)	60.31 (47.41)	87.48 (48.56)
0 dB	1 min	51.55 (71.07)	80.53 (80.79)	46.34 (67.88)	83.73 (68.65)	83.69 (38.35)
	5 min	43.7 (69.08)	49.29 (90.35)	32.07 (49.34)	45.65 (41.85)	66.02 (57.91)
MAE (ms)						
12 dB	1 min	1.09 (1.8)	0.84 (1.23)	1.29 (2.66)	1.03 (2.26)	3.87 (2.97)
	5 min	1.16 (2.14)	0.75 (1.07)	1.17 (2.61)	4.41 (3.72)	3.86 (2.94)
9 dB	1 min	1.36 (5.53)	0.72 (1.5)	1.37 (5.99)	1.51 (2.31)	4.74 (3.76)
	5 min	1.34 (6.02)	0.64 (1.71)	1.48 (5.58)	5.18 (3.51)	4.74 (4.15)
6 dB	1 min	2.87 (7.36)	1.36 (3.68)	4.4 (13.72)	1.36 (5.73)	6.75 (5.03)
	5 min	3.17 (7.5)	2.52 (4.39)	4.33 (17.84)	5.47 (5.92)	6.74 (6.27)
3 dB	1 min	4.44 (11.48)	2.98 (4.53)	2.18 (13.44)	2.09 (19.65)	8.13 (8.8)
	5 min	6 (11.63)	3.54 (5.34)	3.14 (14.91)	5.49 (15.82)	8.8 (8.56)
0 dB	1 min	11.46 (12.85)	6.18 (16.45)	14.41 (12.98)	8.01 (19.81)	9.06 (9.89)
	5 min	13.29 (13.67)	6.22 (17.23)	15.57 (13.19)	12.11 (16.83)	9.04 (10.82)

The 1 min result is reported in the following way. For each subject, noise level, case, and round, we take the channel and the 1-min subset out of the 5-min signal that leads to the tenth highest F_1 to report the result. Then, for each noise level and each case, the median and IQR of all subjects and rounds are shown as the final result, as is suggested in Andreotti et al. [39]. For the 5min result, for each subject, noise level, case, and round, we take the channel that leads to the tenth highest F_1, and then for each noise level and each case, we report the median and IQR of all subjects and rounds.

use the R peaks of the direct fECG as our ground truth, but the direct fECG has a different projection direction compared with the fECG recorded from the maternal abdomen. This difference leads to the slightly larger MAE in the adfecgdb database.

For the computational time over the 5-min signal, it takes about 12 s to finish a round in MacBook Pro (Retina, 15-inch, Mid 2014) with Processor 2.5 GHz Intel Core i7, Memory 16 GB 1,600 MHz DDR3, OS X Yosemite (Version 10.10.5), and Matlab R2014b without implementing the parallel computation.

The fECG morphology estimation results are shown in **Table 3** and **Figure 2**. The high correlation between the estimated fECG and the ground truth indicates that the nonlocal median algorithm leads to an estimated fECG morphology with low distortion. Note that in this evaluation, we only evaluate the correlation on the correctly detected fECG beats (true positive fECG beats). This suggests that if we could accurately estimate the fetal R peaks (the higher F_1 score), then we could have an accurate fECG reconstruction.

To look deeper into the algorithm and its performance, we take subject 1, round 1, case 4, and channel 21 into account, and show the result without noise in **Figure 2**. In this example, over the 5-min, the F_1 is 1 and the MAE is 0.78 ms. In addition to the de-shape STFT of the maECG signal, the estimated R peaks by the beat tracking, the decomposed maternal ECG, the rough fECG estimate $\mathbf{x}_{f,0}$ and its TFR, and the final fECG estimate are

shown. We see that by the de-shape STFT, the information of non-sinusoidal oscillation; that is, the time-varying wave-shape function, and the IHR information are decoupled, and only the IHR information is shown in the TF representation. Note that we could see both the mIHR and fIHR in the TFR, and as expected, fIHR has a weaker intensity than mIHR, meaning that the energy of fECG is smaller. The intensity level could be seen in the colorbar. For a comparison, we could see that in the STFT of aECG, both the mIHR and fIHR could be seen, while the fIHR is much weaker compared with that in the de-shape STFT. Furthermore, the multiples of the maternal fundamental component could mask both the mIHR and fIHR information and even interfere with each other. To show how the fECG could be reconstructed, the estimated fECG and the ground truth fECG signal are put side by side for a comparison. On the other hand, we could see that the nonlocal nature of the nonlocal median algorithm does help us to recover ECG morphology, even the ectopic beats, and the median and IQR of the correlation of all TP beats are 0.997 and 0.004.

To show the result when noise exists, we take another signal, subject 3, round 1, case 4, and channel 23 into account, and with SNR 6dB, as an example, and show the result in **Figure 3**. We choose this signal as an example since it has a smaller fECG amplitude. As can be clearly seen, even when the noise is 6 dB, the de-shape STFT still gives a reasonable TFR with the

TABLE 3 | Results of fECG waveform estimation over the 5-min signal—median correlation and its IQR for different cases and noise levels among 10 subjects and 5 rounds from the simulated database *fecgsym*.

		Case 0	Case 1	Case 2	Case 3	Case 4
CORRELATION						
12 dB	highest	0.962 (0.067)	0.977 (0.044)	0.984 (0.052)	0.983 (0.017)	0.97 (0.032)
	10th highest	0.953 (0.097)	0.96 (0.047)	0.959 (0.062)	0.973 (0.052)	0.963 (0.108)
9 dB	highest	0.962 (0.076)	0.968 (0.05)	0.977 (0.021)	0.978 (0.034)	0.962 (0.039)
	10th highest	0.93 (0.222)	0.943 (0.753)	0.953 (0.12)	0.97 (0.075)	0.942 (0.164)
6 dB	highest	0.954 (0.076)	0.958 (0.083)	0.949 (0.069)	0.977 (0.053)	0.957 (0.141)
	10th highest	0.912 (0.256)	0.919 (0.108)	0.897 (0.374)	0.948 (0.118)	0.926 (0.191)
3 dB	highest	0.919 (0.48)	0.94 (0.103)	0.939 (0.276)	0.965 (0.133)	0.95 (0.216)
	10th highest	0.859 (0.348)	0.925 (0.29)	0.923 (0.428)	0.945 (0.47)	0.863 (0.37)
0 dB	highest	0.892 (0.404)	0.894 (0.289)	0.9 (0.371)	0.942 (0.215)	0.875 (0.386)
	10th highest	0.656 (0.378)	0.846 (0.518)	0.768 (0.3)	0.886 (0.345)	0.778 (0.375)

For each subject, noise level, case, and round, we take the channel that leads to the highest F_1 (and the thenth highest F_1), and then for each noise level and each case, we report the median and IQR of all subjects and rounds.

IHR information, although there are several "speckles" in the background, which come from the noise. In this example, since fECG is smaller compared with the maECG, the fIHR could not be clearly seen in the de-shape STFT of the aECG. However, it could be seen clearly in the de-shape STFT of the rough fECG estimate. The intensity level could be seen in the colorbar. Also note that since the fECG is smaller and the noise is big, and no denoise technique is combined into the algorithm, the de-shape STFT of the rough fECG estimate has a "scattered" background. For the morphology reconstruction, due to the large noise, the nonlocal median failed to recover two ectopic beats, as are indicated in the red arrows. Note that although the nonlocal nature of the nonlocal median has the power to handle ectopic beats, it is not designed for this purpose. To have a better recovery of the ectopic beats, more features specifically designed for ectopic beats should be taken into account. In this case, the F_1 is 0.996, the MAE is 3.476 ms, and the median and IQR of correlations of all TP beats are 0.968 and 0.034. We could see that the nonlocal median provides a convincing potential in recovering the fECG morphology, specially when noise exists.

We should mention that although the *fecgsyn* database provides a universal comparison platform with many interesting examples, the model at this stage is still oversimplified, and that might explain the high accuracy of our proposed algorithm. Many other limitations have been discussed in Andreotti et al. [39]. A better "ground truth" should be considered for evaluation purposes. For example, a well established sheep model [53] could provide a gold standard test bed for the algorithm.

4.3. Results of the Second Database: *adfecgdb* Database

Table 4 shows the F_1, MAE, PPV and SE of all channels and all subjects. Notice that in the r10 record, the direct fECG measurement was lost between 187 and 191 s and between 203 and 211 s. Therefore, these two segments were neglected in the evaluation. To avoid the boundary effect introduced by the

window function, the first and last 0.5 s in every recording are also not evaluated. In terms of F_1, compared with the state-of-art result reported in the field, Castillo et al. [31, Table 5], our result is better overall. Even if compared with the methods based on ICA [54, Table 1], our result is in general compatible. The MAE, which is less reported in the literature, is as small as 10 ms, which indicates the potential to carry out the fetal HRV analysis from the single-lead aECG. However, this topic is out of the scope of this paper.

To further explore the proposed algorithm, in **Figure 4**, we show the result of STFT and de-shape STFT of the third channel of the case r07, denoted as $\mathbf{x}_{11}$. We also show the result of STFT and de-shape STFT of the rough fECG estimate from $\mathbf{x}_{11}$, denoted as $\mathbf{x}_{11,f0}$ (the signal used in the flowchart in **Figure 1** is the channel 1 of the case r01). Clearly, in the STFT of $\mathbf{x}_{11}$, we not only see the fundamental IF of maECG around 1.2 Hz, but also its multiples, due to the nature of non-sinusoidal oscillation of the ECG signal. We could also see a relatively vague component around 2 Hz in STFT, which turns out to be the fIHR. On the other hand, after the de-shape process, in the de-shape STFT of $\mathbf{x}_{11}$, only two dominant curves associated with the mIHR and fIHR are left.

An example of the estimated fECG waveform, $\mathbf{x}_{11,f0}$, is shown in **Figure 5**. The R peaks are clearly well reconstructed and match those in the direct fECG signal recorded from the fetal scalp. By comparing the rough fECG estimate and the final fECG estimate, we could see the effectiveness of nonlocal median. Furthermore, some additional fiducial points could be observed, as shown in **Figure 6**, which is the zoomed in of **Figure 5**. Although not all critical fiducial points could be extracted, this result indicates the potential for studying the morphology of fECG, particularly taking into account the fact that we only count on a single-lead aECG signal. We mention that, in some cases, the P wave and T wave could be reconstructed. However, in general, they are buried in the noise. Therefore, we do not consider it to be an achievement of the proposed algorithm. More work is needed to recover these morphological features.

FIGURE 2 | The result of subject 1, round 1, case 4, and channel 21 without noise from the simulated database (*fecgsym*). In the top subplot, all relevant signals are shown together for a visual comparison. The clean aECG signal is shown in black (shifted up by 8 units) with the detected R peaks in round red circles on the top row; the clean maECG is shown in light gray (shifted up by 5 units) superimposed with the estimated maECG signal in blue on the second row; the rough fetal ECG is shown in dark gray (shifted up by 2.5 units) with the detected R peaks in round red circles on the third row; the clean fECG is shown in light gray superimposed with the estimated fECG signal in red on the bottom row. In the middle subplot, the de-shape STFT of the maECG is shown on the left and the de-shape STFT of the rough fECG is shown on the right. In the bottom subplot, the STFT of the maECG is shown on the left and the STFT of the rough fECG is shown on the right. Note that the discrepancy between the clean maECG and the estimated maECG comes from the median filter, which in general is needed to remove the baseline wandering.

FIGURE 3 | The result of subject 3, round 1, case 4, and channel 23 with the SNR 6dB from the simulated database (*fecgsym*). In the top subplot, all relevant signals are shown together for a visual comparison. The clean maECG signal is shown in black (shifted up by 8 units) with the detected R peaks in round red circles on the top row; the clean maECG is shown in light gray superimposed with the estimated maECG signal in blue (shifted up by 5 units) on the second row; the rough fetal ECG is shown in dark gray (shifted up by 2.5 units) with the detected R peaks in round red circles on the third row; the clean fECG is shown in light gray superimposed with the estimated fECG signal in red on the bottom row. The two red arrows indicate two ectopic beats that are not recovered by the nonlocal median algorithm. In the middle subplot, the de-shape STFT of the maECG is shown on the left and the de-shape STFT of the rough fECG is shown on the right. In the bottom subplot, the STFT of the maECG is shown on the left and the STFT of the rough fECG is shown on the right. Note that the discrepancy between the clean maECG and the estimated maECG comes from the median filter, which in general is needed to remove the baseline wandering.

TABLE 4 | F_1 score, MAE, PPV, and SE of all channels and all subjects over the whole 5 min signals in the adfecgdb database.

Subject	Channel	F_1 (%)	MAE (ms)	PPV (%)	SE (%)
r01	1	**99.53**	**1.5227**	**99.38**	**99.69**
	2	75.21	4.4004	72.33	78.32
	3	87.35	4.4046	86.41	88.30
	4	78.83	4.9358	77.53	80.19
r04	1	82.15	9.0635	81.63	82.67
	2	**98.81**	7.6447	**98.73**	**98.89**
	3	97.54	7.2622	97.46	97.62
	4	98.41	**6.9065**	98.26	98.57
r07	1	84.81	10.5816	84.20	85.42
	2	98.48	8.3182	98.25	98.72
	3	**99.44**	7.6329	**99.36**	**99.52**
	4	99.28	**7.3048**	99.20	99.36
r08	1	**97.85**	**2.0219**	**97.26**	**98.46**
	2	84.44	3.7191	82.69	86.27
	3	71.42	5.3655	69.49	73.46
	4	69.69	5.2397	71.92	67.59
r10	1	**98.65**	**2.8119**	**98.26**	**99.04**
	2	98.49	3.5919	98.26	98.73
	3	78.52	8.5177	76.28	80.89
	4	95.25	4.1681	94.79	95.70

For each subject, the best result among 4 channels is marked in the boldface.

4.4. Results of the Third Database: CinC2013 Database

In this database, since the signal is only 1-min long, we take $K_m = K_f = 10$ to enhance the local nature of the nonlocal median algorithm. The result is slightly worse if $K_m = K_f = 40$.

We now follow the suggested way in Andreotti et al. [39] to report the summary statistics, including the mean and the median of all subjects—for each subject, we choose the channel with the highest F_1 score as the channel for this subject. We then report the gross statistics of the obtained F_1, PPV, SE, and MAE over 75 recorded segments. Over 75 recorded segments, the mean and standard deviation of the F_1 score (respectively, PPV and SE) are 86.37 and 22.9% (respectively 85.77 and 23.42%, and 87.5 and 22.05%), and the mean and standard deviation of MAE are 6.12 ms and 5.56 ms[4]. Our F_1 result is compatible with the reported results in Ghaffari et al. [19, Table 2], [34, Table 5] [8, Section 3.2, where a detected R-peak was labeled as TP if within 100 ms of a reference R-peak], [52, Table 3], while the MAE is better. On the other hand, over 75 recorded segments, the median and IQR of the F_1 score (respectively PPV and SE) are 98.28 and 14% (respectively 98.25 and 14.72%, and 98.59%, and 22.05%), and the median and IQR of MAE are 3.81 ms and 5.74 ms.

---

[4]If we follow Behar et al. [52] and remove 7 cases, including a33, a38, a47, a52, a54, a71, and a74, then the mean and standard deviation of F_1 score (respectively PPV and SE) are 88.73 and 20.96% (respectively 88.39 and 21.08%, and 89.12% and 20.77%), and the mean and standard deviation of MAE over 75 recorded segments are 5.88 ms and 5.14 ms.

The discrepancy between the mean and median indicates the existence of outliers in the database. We thus took a closer look into the database. We found that the fECG signal could be hardly visualized in 10 recorded segments in the database, including a27, a32, a43, a50, a59, a60, a63, a64, a68, and a75, even when the signal is clean, and these segments are considered difficult for our algorithm. If we remove these segments, the mean and standard deviation of the F_1 score (respectively PPV and SE) become 94.54% and 11.63% (respectively 93.87% and 13.05%, and 95.78%, and 8.79%), and the mean and standard deviation of MAE become 4.35 msec and 3.04 msec; the median and IQR of the F_1 score (respectively PPV and SE) become 98.84% and 4.91% (respectively 98.53% and 6%, and 99.21%, and 4.16%), and the mean and standard deviation of MAE become 3.35 ms and 4.33 ms.

In order to examine this problem in more detail, we show one example, a59, which is considered difficult for our algorithm, in **Figure 7**. In **Figure 7**, the second channel of a59 is shown as an illustration. It is clear that the signal is quite clean, and it is not easy to identify if the fECG exists, even with the help of the provided annotation. Not surprisingly, the fetal R peaks are all detected incorrectly. By a direct visual inspection, we could see that those locations that are erratically identified as fetal R peaks are only the residue coming from the incomplete maECG removal. This fact could also be observed in the de-shape STFT of the rough fECG estimate—the only "dominant component" in the de-shape STFT coincides with that of the aECG, as is indicated by the red arrow. We mention that the other three channels all share the same result—the fECG is too small to even be sensed from this relatively clean aECG signal. However, if we take the vectocardiogram (VCG) notion into account, then it is possible to enhance the result. Precisely, by linearly combining different channels, we have a chance to obtain a stronger fECG signal relative to the maECG signal, and hence the result could be better. This idea is shown in **Figure 8**, where the difference between Channel 2 and Channel 3 is analyzed. Clearly, we could see that now the fetal R peaks could be almost perfectly recovered. In this specific case, the F_1 is 99.3% and the MAE is 0.993 ms. We mention that if we do the same linear combination trick between channel 1 and channel 3 (respectively channel 2 and channel 3, channel 2 and channel 3, channel 1 and channel 2, channel 1 and channel 2, channel 2 and channel 3, channel 2 and channel 3, channel 2 and channel 3, and channel 1 and channel 2), the F_1's of a27 (respectively a32, a43, a50, a60, a63, a64, a68, and a75) become 79.3% (respectively 100, 100, 97.9, 84.7, 91, 98.5, 91.4, and 86.2%). This is a naive way to handle the problem and it works well in this preliminary analysis. However, we need at least two channels. Since we focus on the single-lead aECG analysis, this direction will be explored in the other work. We mention that this idea is also considered in Andreotti et al. [18].

Finally, note that the suggested way to report the summary statistics in Andreotti et al. [39] assumes that all leads are available, which is not true in the case of a single-lead recording. Therefore, while the reported summary statistics looks good, we should interpret the reported summary statistics carefully. To better understand the performance of the proposed method, **Tables 5, 6** show the F_1, MAE, PPV, and SE of all channels

FIGURE 4 | Upper left and **upper right:** the STFT and de-shape STFT of the third channel of the case r07 from the *adfecgdb* database; **bottom left** and **bottom right:** the STFT and de-shape STFT of the rough fECG estimate. The red arrow indicates the fIHR, and the blue arrow indicates the mIHR.

and all recorded segments. It is obvious that the algorithm does not always perform well. The performance is strictly connected to an important issue regarding single-lead fECG extraction—the electrode placement on the abdomen. Based on the ECG theory, a change of electrode positioning leads to a change of the recorded fECG, and hence the performance of the algorithm. Thus, the reported summary statistics could be interpreted in the following way—if the electrode is placed in the correct position, the proposed algorithm could provide an accurate fECG information. Since the database is not collected specifically for the single-lead aECG analysis problem and the electrode placement information is not available, a well-designed study is needed to further confirm the performance of the proposed algorithm.

5. DISCUSSION AND FUTURE WORK

In this section, we discuss different findings of the proposed algorithm and areas for future work.

5.1. General Technical Difficulty

In the analysis of multi-component oscillatory signals, there are two chicken-and-egg problems of fundamental importance. The first is the *detection problem*; that is, how to determine

the number of components and how to find the fundamental frequency or equivalently the fundamental period of each component? The second one is the *separation problem*; that is, how to separate all components from a recorded signal? These two problems coexist in many kinds of data, ranging from physiological signals to polyphonic music signals, where each component in the mixture contains information different from others.

Previously, these two problems are usually discussed separately, probably because only discussing one of them is challenging enough. However, recently some research works start to consider these two problems as a single one by viewing the separation problem as an extension of the detection problem. For example, we could simultaneously estimate the IF of each component, and then extract the wave-shape function as well as each component.

The methods could be classified into two classes, *iterative* or *joint*. The iterative approach extracts the most prominent IF/component in each iteration, until no additional IF/component can be found. Although iterative models are usually computationally inexpensive, they have a main drawback: iterative models tend to accumulate errors at each iteration step if the feature representation is not robust

FIGURE 5 | The aECG signal of the third channel of the case r07 from the *adfecgdb* database is plotted in black on the top row. The rough fECG signal estimate (shifted down by 90 units) is plotted in gray on the second row. The final fECG estimation (shifted down by 120 units) is plotted in red superimposed on the rough fECG signal (shifted down by 120 units) plotted in gray for the comparison purpose on the third row. The direct fECG recorded from the fetal scalp (shifted down by 240 units) is shown in blue on the bottom row.

FIGURE 6 | The small segment fECG estimation is plotted in red with the rough fECG signal plotted in gray for the comparison purpose. The data is the third channel of the case r07 from the *adfecgdb* database.

enough. To handle this limitation, we could consider the "joint" approach. Vast majority of recent approaches in multi-pitch estimation and source separation in music now falls within the "joint" category [55], and more and more studies on source separation started to consider pitch information to improve the source separation algorithms [56, 57]. Note that although it is intuitive to utilize the IF information to handle the detection and separation problem, this kind of approach has been less studied until recent years, probably because the task of finding IF is by no means easy, especially when there are multiple components. Compared with the iterative method, the joint methods lead to more accurate estimates but with more involved mathematical tool and increased computational cost.

Our proposed algorithm falls in the "iterative" category—estimate and remove the maternal component first, and get the fIHR and fECG from the left. We do have the accumulated error issue when we estimate the fECG, and we count on the median filter to alleviate this error. One natural question is to ask if it is possible to generate a single-lead fECG algorithm in the joint category, in order to alleviate this problem, and the answer is positive. Precisely, the mIHR and fIHR could be estimated simultaneously by the de-shape STFT, as is shown in **Figure 2**. The estimated R peaks of each component could then be applied to estimate the maECG and fECG by the nonlocal median algorithm. In brief, we could consider an algorithm falling in the joint category, which is composed of two steps: (1) jointly estimate the mIHR and the fIHR by the de-shape STFT;

FIGURE 7 | The result of channel 2 of a59 in the CinC2013 database. In the top subplot, all relevant signals are shown together for a visual comparison. The clean maECG signal is shown in black (shifted up by 1600 units) with the detected maternal R peaks in round red circles and the annotated fetal R peaks in blue diamonds on the top row; the estimated maECG signal is shown in blue (shifted up by 600 units) on the second row; the rough fetal ECG is shown in dark gray (shifted up by 300 units) with the detected R peaks in red upper triangles and the annotated R peaks in blue diamonds on the third row; the fECG estimation is shown in red on the bottom row. It is clear that the fetal R peaks are all detected incorrectly. In the bottom subplot, the de-shape STFT of the maECG is shown on the left and the de-shape STFT of the rough fECG is shown on the right. The red arrow indicated the "dominant curve" in the de-shape STFT, which comes from the incomplete removal of the maECG signal.

that is, carry out the optional step in Step 1 of the algorithm shown in Section 3.1.2; (2) jointly approximate the maECG and fECG signal by the nonlocal mean. However, a preliminary study showed that without a more sophisticated curve extraction algorithm, this approach does not outperform what we propose in the current paper. Its potential will be explored in the future work.

5.2. Several Theoretical and Algorithmic Topics

We discuss more about each step in the proposed algorithm. First, the proposed algorithm is specifically designed to handle a nonstationary signal, which is commonly encountered in physiology. The de-shape STFT respects the local information hidden in a nonstationary time series, and decouples the non-sinusoidal oscillatory pattern from the IF. The nonlocal

median takes the nonlinear relationship between the phase and the oscillation (the nonlinear relationship between the RR interval and QT interval) into account. Thus, the proposed algorithm could be applied to analyze a long time series without modification, and the direct truncation of the long time series into pieces is not needed.

We have seen the potential of extracting more information hidden inside the single-lead aECG signal by the ANH model and the de-shape STFT. In the past few years, different ideas in the TF analysis field have been proposed, regarding the model and the algorithm. It is possible to incorporate them into the current algorithm to achieve a better result. Some examples are: the Blaschke decomposition and unwinding based on the analytic signal analysis technique [58], the synchrosqueezing transform, or the concentration of frequency and time (ConceFT) [59]. However, to simplify the discussion, we do not carry out this

FIGURE 8 | The result of the difference between channel 2 and channel 3 of a59 in the CinC2013 database. In the top subplot, all relevant signals are shown together for a visual comparison. The clean maECG signal is shown in black (shifted up by 55 units) with the detected maternal R peaks in round red circles and the annotated fetal R peaks in blue diamonds on the top row; the estimated maECG signal is shown in blue (shifted up by 35 units) on the second row; the rough fetal ECG is shown in dark gray (shifted up by 20 units) with the detected fetal R peaks in round red circles and the annotated fetal R peaks on red upper triangles on the third row; the fECG estimation is shown in red on the bottom row. In the bottom subplot, the de-shape STFT of the maECG is shown on the left and the de-shape STFT of the rough fECG is shown on the right. Clearly, the fetal R peaks is perfectly recovered.

combination in this paper. While de-shape STFT works well, we cannot ignore the limitation in the curve extraction step for the IHR estimation. Specifically, when the energy of the fECG signal is greater then that of the maECG, then the curve extraction algorithm could easily go astray. When the fECG has an almost equal or larger energy than the maECG, we could see two dominant curves, which may or may not intersect, and the curve extraction algorithm could get confused. While we could apply the physiological constraint to distinguish which curve belongs to the maECG, e.g., the fIHR is, in general, faster than mIHR, it is not universal, and will limit the method to normal subjects. We thus need more tools to handle this issue.

The ability to obtain an accurate IHR estimation is the key step to an accurate R peak location estimation by the beat tracking algorithm. The technical limitation in obtaining good IF information might explain why the intuitive and powerful beat tracking technique has not been widely applied until now,

although it was introduced in the music signal analysis field a long time ago [49]. In sum, the combination of the de-shape STFT, or any other technique providing an accurate IF estimation, is essential. This combination could offer an alternative way to estimate the R peaks or other fiducial points.

Furthermore, to the best of our knowledge, this work is the first one combining the nonlocal median and the nonlinear manifold model to analyze the biomedical time-series analysis, and particularly reconstructing the time-varying wave-shape function. Using the nonlinear manifold structure to analyze a time series is certainly not a new idea [22, 37, 38]. However, in the past, the focus was on decoupling maECG and fECG by the locally linear projection on the sequential beats. In this work, the nonlinear geometric structure is further explored to apply the nonlocal median. In general, we could apply other manifold learning techniques to further improve the algorithm. From a data analysis viewpoint, it is well known that a manifold

TABLE 5 | F_1 score, MAE, PPV, and SE of all channels and all subjects over the whole 1 min signals in the CinC2013 database, part 1.

Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE
1	1	98.9	1.49	99.3	98.6	2	1	95.2	1.85	94.9	95.5	3	1	99.2	3.67	99.2	99.2
	2	50.6	18.3	55.5	46.5		2	21.4	24.7	24.4	19.1		2	99.2	3.88	99.2	99.2
	3	26.2	10.15	28.8	23.9		3	39.1	8.31	38.8	39.5		3	78.1	9.28	76.3	80
	4	97.2	0.82	97.9	96.5		4	40.8	23.42	40.1	41.4		4	98.4	4.2	98.4	98.4
4	1	99.2	1.69	98.4	100	5	1	100	1.86	100	100	6	1	92.7	2.83	92.4	93
	2	83.7	2.97	81.8	85.7		2	23.3	16.94	21.5	25.4		2	50.3	10.77	49.1	51.6
	3	92.1	3.26	91.4	92.9		3	93.7	5.19	93.7	93.7		3	24.6	25.42	25	24.2
	4	89.4	4.99	88.4	90.5		4	98.8	4	98.4	99.2		4	86.9	1.99	87.2	86.6
7	1	92.2	3.3	91.5	92.9	8	1	100	2.53	100	100	9	1	98.8	1.67	98.4	99.2
	2	46.3	10.75	43.4	49.6		2	93.6	4.03	93.6	93.6		2	25.2	19.42	22.6	28.3
	3	30.2	24.93	29	31.5		3	26	11.46	24.3	28		3	33.5	5.94	30.5	37
	4	75.1	4.2	75.4	74.8		4	87.6	3.91	87.3	88		4	34.3	11.61	28.7	42.5
10	1	98	1.78	98.2	97.7	11	1	94.2	1.37	93.5	94.9	12	1	98.9	3.49	98.5	99.3
	2	20.7	19.52	26.9	16.9		2	27.7	23.28	26.3	29.2		2	99.3	3.73	99.3	99.3
	3	21.4	20.53	27.5	17.4		3	31.3	13.62	27.3	36.5		3	76.5	12.58	74.6	78.5
	4	93.8	1.53	94.7	93		4	28.8	25.06	31.9	26.3		4	99.3	4.71	99.3	99.3
13	1	77.9	9.43	77	78.9	14	1	98.8	4.41	98.3	99.2	15	1	98.9	2.39	98.5	99.2
	2	100	7.9	100	100		2	96.7	5.5	95.9	97.5		2	88.1	3.63	86.1	90.1
	3	100	8.19	100	100		3	79.4	10.59	77.2	81.7		3	93.6	3.8	92.5	94.7
	4	100	7.11	100	100		4	95.4	5.65	95	95.8		4	79.6	4.4	77.5	81.7
16	1	86.7	2.58	86	87.4	17	1	97.7	3.34	97.7	97.7	18	1	76.9	3.81	75.7	78.2
	2	24.5	25.38	26.4	22.8		2	90	5.45	89.3	90.7		2	31.2	27.28	27.1	36.7
	3	31.9	17.8	31.5	32.3		3	96.1	6.66	96.1	96.1		3	18	36.04	21.3	15.6
	4	33.2	23.3	31.9	34.6		4	71.9	4.61	72.4	71.3		4	17.3	35.45	20.4	15
19	1	78.7	10.18	78.4	79	20	1	78.1	8.24	79.7	76.6	21	1	98.6	2.22	98.6	98.6
	2	99.2	7.47	99.2	99.2		2	99.6	8.36	99.2	100		2	65.5	21.12	64.8	66.2
	3	100	7.64	100	100		3	92.5	7.05	92.9	92.2		3	22.3	9.59	24.6	20.4
	4	100	6.37	100	100		4	97.3	7.22	96.9	97.7		4	97.2	2.36	97.2	97.2
22	1	100	2.09	100	100	23	1	75.9	11.25	76.2	75.6	24	1	93.3	9.71	93.3	93.3
	2	89.4	4.52	89.4	89.4		2	100	7.64	100	100		2	100	9.04	100	100
	3	99.2	4.42	99.2	99.2		3	100	6.9	100	100		3	97.5	9.07	97.5	97.5
	4	99.2	3.63	99.2	99.2		4	100	6.85	100	100		4	99.2	9.31	99.2	99.2
25	1	82.9	8.62	82.3	83.6	26	1	24.2	21.68	25.4	23.1	27	1	28.9	19.18	25.3	33.6
	2	98.8	6.92	98.4	99.2		2	90	4.71	89.6	90.3		2	27.6	16.74	24.3	32.1
	3	98.8	6.36	98.4	99.2		3	82.2	5.69	81.6	82.8		3	30.3	17.07	26.6	35.1
	4	98.8	6.65	98.4	99.2		4	95.9	3.98	95.6	96.3		4	29.8	16	26.3	34.4
28	1	99.4	2.37	98.8	100	29	1	23.4	24.25	21.6	25.4	30	1	97.5	6.01	97.1	97.8
	2	98.2	2.52	97.6	98.8		2	89.8	7.12	88.5	91.3		2	89.5	19.02	89.1	89.8
	3	99.4	3.1	98.8	100		3	87.5	5.75	86.2	88.9		3	90.9	19.7	90.6	91.2
	4	99.4	2.93	98.8	100		4	58	12.59	57.4	58.7		4	55.7	20.67	55.9	55.5
31	1	19	22.15	18.6	19.4	32	1	40.4	18.5	39.2	41.6	33	1	72.3	11.69	70.3	74.5
	2	99.3	5.55	99.3	99.3		2	42.7	18.58	41.3	44.3		2	94.6	3.73	93	96.4
	3	98.5	6.23	98.5	98.5		3	31.3	27.02	30.4	32.2		3	90	1.71	88.1	92
	4	97.8	7.41	97.8	97.8		4	29.9	22.55	31.8	28.2		4	92.1	2.63	90.2	94.2

(Continued)

TABLE 5 | Continued

Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE
34	1	29.3	24.71	29.2	29.5	35	1	99.7	6.41	99.4	100	36	1	100	3.76	100	100
	2	95.4	5.94	94.7	96.1		2	99.1	6.9	98.8	99.4		2	100	3.13	100	100
	3	95.4	6.19	94.7	96.1		3	99.7	6.02	99.4	100		3	100	3.08	100	100
	4	83.5	10.3	82.6	84.5		4	99.7	6.17	99.4	100		4	99.7	3.49	100	99.4
37	1	98.9	1.78	98.6	99.3	38	1	24.4	23.88	28.7	21.3	39	1	100	1.59	100	100
	2	99.6	1.72	99.3	100		2	32.6	26.74	30.7	34.8		2	100	1.43	100	100
	3	20	20.66	19.1	21		3	35.1	13.22	32.6	38.1		3	17.5	21.55	18.5	16.7
	4	99.6	3.2	99.3	100		4	39	14.23	36.5	41.9		4	99.6	1.79	99.2	100

Sbj, Subject; Ch, Channel; the unit for F_1, PPV and SE is %; the unit for MAE is ms.

structure might still be too restrictive to be the best model for this dataset. However, it could be low dimensional and nonlinear with nontrivial topology, and could serve as a good model to understand different algorithms. Thus, finding a more flexible model to include more physiological facts is a critical issue in future studies. While the L^1 norm is associated with the nonlocal median algorithm, to enhance the sparsity feature, it is possible to consider the L^p quasi-norm, where $0 < p < 1$, which has been applied in the nonlocal patch regression for image denoising [60].

When there is more than one channel, it could be beneficial to combine results from different channels, or to integrate the proposed algorithm with existing multiple-lead algorithms. The benefit of doing so has been shown in **Figure 8**. Another direct benefit of using multiple-lead algorithms is guiding the mIHR estimation by the curve extraction, when the fECG has a stronger energy than the maECG. In practice, we can combine other kinds of signals, like the contact photoplethysmogram signal commonly equipped in modern wearable devices, to guide the mIHR estimation. The result of this kind of combination and the analysis of multiple channel signals will be reported in future work.

5.3. Several Clinical Topics

In clinics, the fECG signal could provide at least two different kinds of physiological information—the HRV and the electrophysiological dynamics. For the HRV analysis, the most important ingredient is having an accurate R peak detection algorithm. One of the main strength of the propose algorithm is providing the state-of-the-art MAE in the field. For example, in the real database, the averaged MAE is about 5 ms, which is roughly equivalent to the R peak information from an fECG signal sampled at about 200 Hz, and its clinical value does deserve a further evaluation. However, note that when we evaluate the MAE, we focus only on the TP beats, which means that we could only get a good HRV analysis when F_1 is high enough. See **Figure 9** for an illustration of the relationship between the HRV analysis and the F_1. This figure is generated in the following way. For each case in CinC2013 (75 in total), we evaluate the root mean square error (RMSE) to quantify how accurate we could recover the fIHR for the HRV analysis. Based on the Task Force [61], we get the fIHR's from the estimated fetal R peaks and from the annotated fetal R peaks, and RMSE between these

two fIHR is evaluated. Note that we take all estimated fetal R peaks to simulate the real scenario. We could see that when F_1 is slightly worse than 95%, the RMSE is as high as 0.1 beats/s. This preliminary result emphasizes the importance of getting an accurate F_1. An extensive study of this important topic will be reported in the future work.

For the fECG morphology reconstruction, although the result is convincing, note that we could not infer too much electrophysiological information from the single-lead fECG signal, even if the reconstruction is perfect. Indeed, due to the variation of relative cardiac axes between the mother and fetus, the projection direction of fECG is in general unknown, even if the abdominal lead system on the abdominal surface is standardized. To explore this direction, we need to design or choose the optimal single abdominal lead location under the complicated time-varying dynamics, which is adaptive to the mother, or we may take multiple leads and proceed with an adaptive way to reconstruct the VCG signal.

Uterine contraction is in general a difficult problem and big challenge in the fECG signal analysis, since it behaves like a huge noise in the aECG signal. In general, to handle such a huge noise, some a priori knowledge of the noise is needed. While the current method could provide a reasonable result in Case 3 of the simulated database, it does not have the ability to handle the uterine contraction, and that is why we always get an F_1 less then 80% in the 5 min result. While it is out of the scope of this paper, we mention that one possible approach is to incorporate the physiological behaviors underlying cardiac signals, a realistic model of an individual fetus' ECG, or the statistical behavior of the noise associated with the uterine contraction to improve the result [62–64].

It is not a reasonable assumption that the mother or the fetus are both healthy in the real life. Sometimes we might encounter subjects with a problematic heart. In Case 4 of the simulated database, the situation when the fetus and mother both have frequent ectopic beats, like ventricular pre-contraction (VPC), is considered. We see that the nonlocal median algorithm gives us a reasonable cardiac activity estimation. Indeed, the distance between an ectopic beat and a normal beat is in general larger than the distance between two ectopic beats and two normal betas. Therefore, the nonlocal nature of the nonlocal median algorithm could accurately estimate the cardiac activity, since

TABLE 6 | F_1 **score, MAE, PPV, and SE of all channels and all subjects over the whole 1 min signals in the CinC2013 database, part 2.**

Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE
40	1	27.5	28.7	29.8	25.5	41	1	99.2	4.42	99.2	99.2	42	1	24.8	25.37	23.8	25.9
	2	97.2	4.3	97.2	97.2		2	99.6	2.37	99.3	100		2	99.7	1.57	99.3	100
	3	17.3	25.5	20.2	15.2		3	100	2.53	100	100		3	23.7	24.39	27	21.1
	4	100	2.25	100	100		4	98.1	3.07	97.8	98.5		4	99.7	1.7	99.3	100
43	1	50.5	16.05	50	51	44	1	99.7	6.78	100	99.4	45	1	14.8	21	15.2	14.5
	2	30.4	24.06	29.8	31		2	89.3	6.58	89.9	88.8		2	98.9	6.51	98.6	99.3
	3	35.6	19.66	37.1	34.2		3	99.7	6.54	100	99.4		3	99.6	6.57	99.3	100
	4	33.3	19.75	33.1	33.5		4	99.7	6.96	100	99.4		4	98.2	7.18	97.8	98.6
46	1	92.7	6.22	91.6	93.8	47	1	23.8	25.84	26.1	22	48	1	27	26.55	25.2	29
	2	94.6	4.61	93.8	95.3		2	74.5	4.4	72.5	76.6		2	85.3	3.78	84.3	86.3
	3	51.9	11.7	50	53.9		3	82.9	5.09	81.5	84.4		3	96.2	2.26	95.5	96.9
	4	31.2	6.67	35	28.1		4	85.3	3.34	84.1	86.5		4	80	4.81	79.1	80.9
49	1	97.2	10.15	97.2	97.2	50	1	27.3	16.29	29.9	25.2	51	1	83.6	7.56	82.1	85.2
	2	98.3	1.77	97.9	98.6		2	37	12.45	38.9	35.3		2	92.4	6.83	90.7	94.1
	3	97.6	1.65	97.3	97.9		3	37.5	15.17	37.7	37.4		3	92.7	7.9	91.4	94.1
	4	95.9	3.8	95.2	96.6		4	23.5	28.32	24.8	22.3		4	94.2	13.1	92.8	95.6
52	1	83.9	5.07	81.8	86.2	53	1	24.4	24.84	29.8	20.7	54	1	10.9	27.25	6.5	34.3
	2	78.4	6.14	76.1	80.8		2	98	2.52	98	98		2	38.8	1.74	24.8	88.6
	3	77.3	5.52	74.8	80		3	20.7	25.75	23.1	18.7		3	38.8	0.71	24.8	88.6
	4	77.6	4.91	75.4	80		4	99.7	1.32	99.3	100		4	38.8	0.65	24.8	88.6
55	1	21.9	20.41	23.2	20.7	56	1	21	29.2	23.1	19.2	57	1	77.1	12.58	76.4	77.9
	2	98.6	8.08	98.6	98.6		2	78.8	6.23	77.6	80		2	73.2	11.65	72	74.5
	3	97.9	7.45	97.9	97.9		3	25.7	27.72	22.4	30		3	72.4	10.3	71.6	73.1
	4	97.1	7	97.1	97.1		4	25.1	14.46	23.5	26.9		4	63.5	10.34	62.8	64.1
58	1	100	3.35	100	100	59	1	32.1	17.87	33.6	30.7	60	1	22.1	28.3	27.6	18.5
	2	18.9	28.3	21.1	17.2		2	30.5	15.12	34.5	27.3		2	24.4	23.7	30	20.5
	3	100	2.17	100	100		3	23	31.19	25.8	20.7		3	20.4	23.52	25.3	17.1
	4	99.6	2.43	99.3	100		4	25.8	16.6	28.9	23.3		4	43.9	15.97	42.6	45.2
61	1	39.7	22.02	39.3	40.1	62	1	28	22.87	24.9	31.9	63	1	20.7	7.68	23.4	18.5
	2	96.4	2.12	95	97.8		2	98.6	2.42	98.6	98.6		2	30.6	6.81	34.6	27.4
	3	25.5	25.23	25.5	25.5		3	99.3	1.99	99.3	99.3		3	44.4	8.39	47.9	41.5
	4	97.8	2.1	96.5	99.3		4	97.9	2.14	97.9	97.9		4	24.2	24.32	29.2	20.7
64	1	57	11.24	57.7	56.4	65	1	27.1	26.72	26.4	27.9	66	1	20.5	23	19.9	21.3
	2	59.1	10.03	59.5	58.6		2	100	1.34	100	100		2	99.6	2.06	99.2	100
	3	51.3	6.88	52.3	50.4		3	24.2	24.73	24.8	23.6		3	99.6	1.56	99.2	100
	4	57	8.14	56.2	57.9		4	100	1.45	100	100		4	99.6	2.06	99.2	100
67	1	93.4	2.49	92.2	94.7	68	1	28	26.9	26.1	30.1	69	1	97.9	1.51	97.9	97.9
	2	93.7	1.39	92.8	94.7		2	31.6	25.19	32.3	30.9		2	99.3	0.78	99.3	99.3
	3	95.4	1.7	94.2	96.7		3	24.5	26.97	24.8	24.3		3	100	1.46	100	100
	4	95.4	1.65	94.2	96.7		4	23.3	22.66	21.2	25.7		4	97.9	1.22	97.9	97.9
70	1	100	9.57	100	100	71	1	20.1	26.46	21.7	18.8	72	1	100	5.44	100	100
	2	99.3	9.18	99.3	99.3		2	18.9	22.58	20.6	17.4		2	100	5.45	100	100
	3	99.3	10	99.3	99.3		3	17	28.32	20	14.8		3	100	6.01	100	100
	4	100	10.15	100	100		4	12.5	26.31	15.1	10.7		4	99.7	5.85	100	99.4

(Continued)

TABLE 6 | Continued

Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE	Sbj	Ch	F_1	MAE	PPV	SE
73	1	97.8	6.27	97.1	98.5	74	1	82.1	4.65	80.4	83.9	75	1	28.4	11.59	25.9	31.3
	2	82.3	20.37	81.4	83.2		2	76.4	6.25	74.8	78.1		2	35.1	10.31	31.9	38.9
	3	33.7	21.06	31.3	36.5		3	52.3	10.24	50.7	54		3	24.8	24.68	29.5	21.4
	4	32.6	19.12	34.7	30.7		4	59.2	11.49	57.1	61.3		4	32.8	10.71	29.6	36.6

Sbj, Subject; Ch, Channel; the unit for F_1, PPV and SE is %; the unit for MAE is ms.

VPC and normal beats are clustered into two groups. Note that if VPC is polymorphic and/or the VPC number is small, the problem becomes more challenging and a different approach is needed. Another important issue is when the maECG and the fECG are close in the absolute amplitude. In this case, the algorithm might be confused and outputs erroneous maECG and fECG. If both fetus and mother are healthy, we could count on the heart rate relationship to correct the output (the final step of the algorithm is based on this assumption). However, in general it is not true; for example, the fetal heart rate might be as low as 80 beat per minute during umbilical cord occlusions and the maternal heart rate might be as high as over 100 beat per minute during the labor process. Without more information, it is in general a challenging issue for any single-lead fECG extraction algorithm.

This is true in general case, but . So this is problematic as a default setting). One way around is to add a temporal criterion for how long such swappable condition is allowed to persist and monitor it prior to deciding to swap; if FHR deceleration is identified, no swap would be proposed;

The twin pregnancy problem was relatively rarely discussed, except in Niknazar et al. [17] and Taylor et al. [65]. In Case 5 of the simulated database, although we do not specifically study in this paper, we mention that the proposed algorithm has the potential to handle even multiple pregnancy problem. Indeed, the mIHR and fIHR's of different fetuses could be simultaneously revealed in the de-shape STFT. However, it is still challenging to utilize this single channel information at this stage, since there is no a priori information about the relationship between the IHR's and signal strengths of those fetuses. This limits the curve extraction step in the de-shape STFT, as the IHR's and energy of two fetus might be similar. Thus, although we could see all the fIHR's, to extract the fIHR from the TF representation based on the current curve extraction algorithm is challenging, and we need a more sophisticated curve extraction algorithm to do it. The study of the above-mentioned clinical topics will be explored in the future work.

5.4. Several Practical Topics

In this study, we do not carry out any optimization to choose the best parameters. We just choose them in an *ad hoc* fashion, and our pilot study showed that the algorithm is generally stable to the choice of parameters. However, some of the parameters are critical due to physical limitation. For example, due to the inevitable uncertainty principle of the STFT, in general the window should be long enough to capture enough spectrum

FIGURE 9 | The RMSE and F_1 of the 75 subjects in the CinC2013 database is plotted.

information. However, it should not be too long or the local information could be missed. In practice, we found that a window that could cover about 5-10 oscillations is a good choice. On the other hand, when the signal is noisy, we could get more stable results if the window is longer. Therefore, we choose a 10 s long window for the very noisy simulated signal, and a 5 s window for the less noisy real databases. Since our current result is based on such an *ad hoc* choice of parameters, the result could be improved by running a systematic parameter optimization for a specific mission.

The proposed fECG extraction algorithm is just one among many successful algorithms. Each algorithm has its own strengths and weaknesses, and it is generally believed that there might not exist a single algorithm that could handle all different situations and different signals. Thus, it would be of great interest to consider the possibility of combining the proposed algorithm with other methods, like the combination proposed in Behar et al. [52]. More practically, while in real life, the noise could be highly nonstationary, we may want to combine the notion of a signal quality index (SQI) to help guide us in selecting the "good" signal for analysis [13, 66, 67]. In general, if there is too much noise and it bypasses the information limitation, then there is no hope to recover anything. However, if the noise is not that big, then some denoising techniques could certainly help. In this work, to simplify the discussion, we do not consider any denoising technique before running de-shape STFT

or beat tracking. However, it could be of great help if we could incorporate a good de-noise scheme into the algorithm. All these interesting practical issues will be explored in the near future.

Last but not the least, while the strength of the proposed algorithm is confirmed on two real databases, these databases are not big enough and not designed specifically for our algorithm. An important and relevant issue regarding the single channel fECG extraction is represented by where the electrode placement on the abdomen, and a small change of electrode positioning might cause a drastic change of the recorded fECG, and hence strongly affect the success of any single channel fECG extraction algorithm. However, since the dataset is not collected by us, this very sensitive element cannot be taken in consideration in this study. A well controlled larger scale clinical study is needed to further confirm the usefulness of the proposed algorithm.

6. CONCLUSION

In this paper, we propose a novel fECG extraction algorithm depending mainly on a single-lead aECG based on a nonlinear TF analysis technique, de-shape STFT, and the nonlocal median. In addition to providing the theoretical model and mathematical guarantee, the algorithm is tested on two real databases and one simulated database. The main novelty in our algorithm is an extensive utilization of hidden information inside the aECG signal, that is, both the frequency and energy information and the nonlinear relationship between consecutive cardiac activities.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fams.2017.00002/full#supplementary-material

REFERENCES

1. AlGhatrif M, Lindsay J. A brief review: history to understand fundamentals of electrocardiography. *J Commun Hosp Intern Med Perspect.* (2012) **2**:10.3402. doi: 10.3402/jchimp.v2i1.14383

2. Jenkins HM. Thirty years of electronic intrapartum fetal heart rate monitoring: discussion paper. *J R Soc Med.* (1989) **82**:210–4. doi: 10.1177/014107688908200410

3. Belfort MA, Saade GR, Thom E, Blackwell SC, Reddy UM, Thorp JM, et al. A randomized trial of intrapartum fetal ECG ST-segment analysis. *N Engl J Med.* (2015) **373**:632–41. doi: 10.1056/NEJMoa1500600

4. Sameni R, Clifford GD. A review of fetal ECG signal processing; issues and promising directions. *Open Pacing Electrophysiol Ther J.* (2010) **3**:4–20. doi: 10.2174/1876536x01003010004

5. Clifford GD, Silva I, Behar J, Moody GB. Non-invasive fetal ECG analysis. *Physiol Meas.* (2014) **35**:1521. doi: 10.1088/0967-3334/35/8/1521

6. De Lathauwer L, De Moor B, Vandewalle J. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans Biomed Eng.* (2000) **47**:567–72. doi: 10.1109/10.841326

7. Akhbari M, Niknazar M, Jutten C, Shamsollahi MB, Rivet B. Fetal electrocardiogram R-peak detection using robust tensor decomposition and extended Kalman filtering. In: *Computing in Cardiology Conference (CinC)* (2013). p.189–92.

8. Di Maria C, Liu C, Zheng D, Murray A, Langley P. Extracting fetal heart beats from maternal abdominal recordings: selection of the optimal principal components. *Physiol Meas.* (2014) **35**:1649–64. doi: 10.1088/0967-3334/35/8/1649

9. Varanini M, Tartarisco G, Billeci L, Macerata A, Pioggia G, Balocchi R. An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG. *Physiol Meas.* (2014) **35**:1607–19. doi: 10.1088/0967-3334/35/8/1607

10. Sameni R, Jutten C, Shamsollahi MB. Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE Trans Biomed Eng.* (2008) **55**:1935–40. doi: 10.1109/TBME.2008.919714

11. Haghpanahi M, Borkholder DA. Fetal ECG extraction from abdominal recordings using array signal processing. *Comput. Cardiol.* (2013) **40**:173–6.

12. Akbari H, Shamsollahi MB, Phlypo R. Fetal ECG extraction using π tucker decomposition. In: *International Conference on Systems, Signals and Image Processing.* IEEE (2015).

13. Behar J, Johnson A, Clifford GD, Oster J. A comparison of single channel fetal ECG extraction methods. *Ann Biomed Eng.* (2014) **42**:1340–53. doi: 10.1007/s10439-014-0993-9

14. Widrow B, Williams CS, Glover JR, McCool JM, Hearn RH, Zeidler JR, et al. Adaptive noise cancelling: principles and applications. *Proc IEEE* (1975) **63**:1692–716. doi: 10.1109/PROC.1975.10036

15. Graupe D, Graupe MH, Zhong Y, Jackson RK. Blind adaptive filtering for non-invasive extraction of the fetal electrocardiogram and its non-stationarities. *Proc Inst Mech Eng H J Eng Med.* (2008) **222**:1221–34. doi: 10.1243/09544119JEIM417

16. Sameni R. *Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings.* Sharif University of Technology, Institut National Polytechnique de Grenoble (2008).

17. Niknazar M, Rivet B, Jutten C. Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings. *IEEE Trans Biomed Eng.* (2013) **60**:1345–52. doi: 10.1109/TBME.2012.2234456

18. Andreotti F, Riedl M, Himmelsbach T, Wedekind D, Wessel N, Stepan H, et al. Robust fetal ECG extraction and detection from abdominal leads. *Physiol Meas.* (2014) **35**:1551–67. doi: 10.1088/0967-3334/35/8/1551

19. Ghaffari A, Mollakazemi MJ, Atyabi SA, Niknazar M. Robust fetal QRS detection from noninvasive abdominal electrocardiogram based on channel selection and simultaneous multichannel processing. *Aust Phys Eng Sci Med.* (2015) **38**:581–92. doi: 10.1007/s13246-015-0381-2

20. Rodrigues R. Fetal beat detection in abdominal ECG recordings: global and time adaptive approaches. *Physiol Meas.* (2014) **35**:1699–711. doi: 10.1088/0967-3334/35/8/1699

21. Lipponen JA, Tarvainen MP. Advanced maternal ECG removal and noise reduction for application of fetal QRS detection. *Comput. Cardiol.* (2013) **40**:161–4.

22. Karvounis EC, Tsipouras MG. Detection of fetal heart rate through 3-D phase recordings. *IEEE Trans Biomed Eng.* (2009) **56**:1394–406. doi: 10.1109/TBME.2009.2014691

23. van Bemmel JH, van der Weide H. Detection procedure to represent the foetal heart rate and electrocardiogram. *IEEE Trans Biomed Eng.* (1966) **13**:175–82. doi: 10.1109/TBME.1966.4502434

24. Cerutti S, Baselli GB, Civardi S, Ferrazzi E, Marconi AM, Pagani M, et al. Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings. *J Perinat Med.* (1986) **14**:445–52. doi: 10.1515/jpme.1986.14.6.445

25. Martens SMM, Rabotti C, Mischi M, Sluijter RJ. A robust fetal ECG detection method for abdominal recordings. *Physiol Meas.* (2007) **28**:373–88. doi: 10.1088/0967-3334/28/4/004

26. Ungureanu M, Bergmans JWM, Oei SG, Strungaru R. Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique. *Biomedizinische Technik.* (2007) **52**:56–60. doi: 10.1515/BMT.2007.011

27. Kanjilal PP, Palit S, Saha G. Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Trans Biomed Eng.* (1997) **44**:51–59. doi: 10.1109/10.553712

28. Christov I, Simova I, Abächerli R. Extraction of the fetal ECG in noninvasive recordings by signal decompositions. *Physiol Meas.* (2014) **35**:1713–21. doi: 10.1088/0967-3334/35/8/1713

29. Khamene AH, Negahdaripoure S. A new method for the extraction of fetal ECG from the composite abdominal signal. *IEEE Trans Biomed Eng.* (2000) **47**:507–16. doi: 10.1109/10.828150

30. Karvounis EC, Tsipouras MG, Fotiadis DI, Naka KK. An automated methodology for fetal heart rate extraction from the abdominal electrocardiogram. *IEEE Trans Inf Technol Biomed.* (2007) **11**:628–38. doi: 10.1109/TITB.2006.888698

31. Castillo E, Morales DP, Botella G, Garcia A, Parrilla L, Palma AJ. Efficient wavelet-based ECG processing for single-lead FHR extraction. *Digital Signal Process.* (2013) **23**:1897–909. doi: 10.1016/j.dsp.2013.07.010

32. Almeida R, Gonçalves H, Bernardes J, Rocha AP. Fetal QRS detection and heart rate estimation: a wavelet-based approach. *Physiol Meas.* (2014) **35**:1723–35. doi: 10.1088/0967-3334/35/8/1723

33. Lamesgin G, Kassaw Y, Assefa D. Extraction of fetal ECG from abdominal ECG and heart rate variability analysis. *Adv Intell Syst Comput.* (2015) **334**:147–61. doi: 10.1007/978-3-319-13572-4_5

34. Lee K, Lee B. Sequential total variation denoising for the extraction of fetal ECG from single-channel maternal abdominal ECG. *Sensors* (2016) **16**:1020. doi: 10.3390/s16071020

35. Panigraphy D, Sahu PK. Extraction of fetal ECG by extended state Kalman filtering and adaptive neuro-fuzzy inference system based on single channel. *Sadhana* (2015) **40**:1091–104. doi: 10.1007/s12046-015-0381-7

36. Panigrahy D, Rakshit M, Sahu PK. An efficient method for fetal ECG extraction from single channel abdominal ECG. *ICIC* (2015) **2015**:1083–8. doi: 10.1109/iic.2015.7150908

37. Richter M, Schreiber T, Kaplan DT. Fetal ECG extraction with nonlinear state-space projections. *IEEE Trans Biomed Eng.* (1998) **45**:133–7. doi: 10.1109/10.650369

38. Kotas M, Jezewski J, Matonia A, Kupka T. Towards noise immune detection of fetal QRS complexes. *Comput Methods Programs Biomed.* (2010) **97**:241–56. doi: 10.1016/j.cmpb.2009.09.005

39. Andreotti F, Behar J, Zaunseder S, Oster J, Clifford GD. An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms. *Physiol Meas.* (2016) **37**:627–48. doi: 10.1088/0967-3334/37/5/627

40. Chaudhury KN, Singer A. Non-local euclidean medians. *IEEE Signal Process. Lett.* (2012) **19**:745–8. doi: 10.1109/LSP.2012.2217329

41. Lin CY, Minasian A, Qi J, Wu HT. Vector nonlocal euclidean median: principal bundle captures the nature of patch space. (2016) arxiv:1611.05073.

42. Malik M, Farbom P, Batchvarov V, Hnatkova K, Camm AJ. Relation between QT and RR intervals is highly individual among healthy subjects: implications for heart rate correction of the QT interval. *Heart* (2002) **87**:220–8. doi: 10.1136/heart.87.3.220

43. Lin CY, Li S, Wu HT. Wave-shape function analysis–when cepstrum meets time-frequency analysis. *J Four Anal Appl.* (in press). arXiv:1605.01805

44. Chen YC, Cheng MY, Wu HT. Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *J R Stat Soc B.* (2014) **76**:651–82. doi: 10.1111/rssb.12039

45. Zhou Z. Heteroscedasticity and autocorrelation robust structural change detection. *J Am Stat Assoc.* (2013) **108**:726–40. doi: 10.1080/01621459.2013.787184

46. Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. *Multi Model Simul.* (2005) **4**:490–530. doi: 10.1137/040616024

47. Singer A, Shkolnisky Y, Nadler B. Diffusion interpretation of nonlocal neighborhood filters for signal denoising. *SIAM J Imaging Sci.* (2009) **2**:118. doi: 10.1137/070712146

48. Laguna P, Sörnmo L. Sampling rate and the estimation of ensemble variability for repetitive signals. *Med Biol Eng Comput.* (2000) **38**:540–6. doi: 10.1007/BF02345750

49. Ellis DPW. Beat tracking by dynamic programming. *J New Music Res.* (2007) **36**:51–60. doi: 10.1080/09298210701653344

50. Guerrero-Martinez JF, Martinez-Sober M, Bataller-Mompean M, Magdalena-Benedito JR. New algorithm for fetal QRS detection in surface abdominal records. In: *2006 Computers in Cardiology* (2006). p. 441–4.

51. Goldberger AL, Amaral LAN, 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**:e215–20. doi: 10.1161/01.CIR.101.23.e215

52. Behar J, Oster J, Clifford GD. Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data. *Physiol Meas.* (2014) **35**:1569–89. doi: 10.1088/0967-3334/35/8/1569

53. Durosier LD, Green G, Batkin I, Seely AJ, Ross MG, Richardson BS, et al. Sampling rate of heart rate variability impacts the ability to detect acidemia in ovine fetuses near-term. *Front Pediatr.* (2014) **2**:38. doi: 10.3389/fped.2014.00038

54. Poian GD, Bernardini R, Rinaldo R. Sparse representation for Fetal QRS detection in abdominal ECG recordings. In: *The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015.* Vol. 1 (2015). p. 1–4.

55. Benetos E, Dixon S, Giannoulis D, Kirchhoff H, Klapuri A. Automatic music transcription: challenges and future directions. *J Intell Inform Syst.* (2013) **41**:407–34. doi: 10.1007/s10844-013-0258-3

56. Chan TS, Yeh TC, Fan ZC, Chen HW, Su L, Yang YH, et al. Vocal activity informed singing voice separation with the iKala dataset. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE (2015). p. 718–22.

57. Ikemiya Y, Yoshii K, Itoyama K. MIREX2014: singing voice separation. In: *The 10th Annual Music Information Retrieval Evaluation eXchange (MIREX)* (2014).

58. Coifman RR, Steinerberger S. Nonlinear phase unwinding of functions. *J Fourier Anal. Appl.* (2016). arXiv:1508.01241.

59. Daubechies I, Wang Y, Wu HT. ConceFT: concentration of frequency and time via a multitapered synchrosqueezing transform. *Philos Trans A Math Phys Eng Sci.* (2016) **374**:20150193. doi: 10.1098/rsta.2015.0193

60. Chaudhury KN, Singer A. Non-local patch regression: robust image denoising in patch space. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2013). p. 1345–9.

61. Task Force. Heart rate variability : standards of measurement, physiological interpretation, and clinical use. *Circulation* (1996) **93**:1043–65. doi: 10.1161/01.CIR.93.5.1043

62. Clifford GD, Shoeb A, McSharry PE, Janz BA. Model-based filtering, compression and classification of the ECG. *Proceedings of Bioelectromagnetism and 5th International Symposium on Noninvasive Functional Source Imaging within the Human Brain and Heart (BEM & NFSI).* (2005). p. 158–161.

63. Sameni R, Clifford GD, Jutten C, Shamsollahi MB. Multichannel ECG and noise modeling: application to maternal and fetal ECG signals. *Eur J Adv Signal Process.* (2007) **2007**:1–14.

64. Sameni R, Shamsollahi MB, Jutten C, Clifford GD. A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Trans Biomed Eng.* (2007) **54**:2172–185. doi: 10.1109/TBME.2007.897817

65. Taylor MJO, Smith MJ, Thomas M, Green AR, Cheng F, Oseku-Afful S, et al. Non-invasive fetal electrocardiography in singleton and multiple pregnancies. *BJOG* (2003) **110**:668–78. doi: 10.1046/j.1471-0528.2003.02005.x

66. Liu C, Li P, Di Maria C, Zhao L, Zhang H, Chen Z. A multi-step method with signal quality assessment and fine-tuning procedure to locate maternal and fetal QRS complexes from abdominal ECG recordings. *Physiol Meas.* (2014) **35**:1665–83. doi: 10.1088/0967-3334/35/8/1665

67. Gambarotta N, Aletti F, Baselli G, Ferrario M. A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. *Med Biol Eng Comput.* (2016) **54**:1025–35. doi: 10.1007/s11517-016-1453-5