

## Reproducibility and reliability of fetal cardiac time intervals using magnetocardiography

P van Leeuwen<sup>1</sup>, S Lange<sup>1</sup>, A Klein<sup>1</sup>, D Geue<sup>1</sup>, Y Zhang<sup>2</sup>, H-J Krause<sup>2</sup>  
and D Grönemeyer<sup>3</sup>

<sup>1</sup> Department of Biomagnetism, EFMT, Bochum, Germany

<sup>2</sup> Institute of Thin Films and Interfaces (ISG), Research Centre Jülich, Germany

<sup>3</sup> Institute of Microtherapy, University Witten/Herdecke, Bochum, Germany

E-mail: [petervl@microtherapy.de](mailto:petervl@microtherapy.de)

Received 29 September 2003, accepted for publication 19 February 2004

Published 25 March 2004

Online at [stacks.iop.org/PM/25/539](http://stacks.iop.org/PM/25/539) (DOI: 10.1088/0967-3334/25/2/011)

### Abstract

We investigated several factors which may affect the accuracy of fetal cardiac time intervals (CTI) determined in magnetocardiographic (MCG) recordings: observer differences, the number of available recording sites and the type of sensor used in acquisition. In 253 fetal MCG recordings, acquired using different biomagnetometer devices between the 15th and 42nd weeks of gestation, P-wave, QRS complex and T-wave onsets and ends were identified in signal averaged data sets independently by different observers. Using a defined procedure for setting signal events, interobserver reliability was high. Increasing the number of registration sites led to more accurate identification of the events. The differences in wave morphology between magnetometer and gradiometer configurations led to deviations in timing whereas the differences between low and high temperature devices seemed to be primarily due to noise. Signal-to-noise ratio played an important overall role in the accurate determination of CTI and changes in signal amplitude associated with fetal maturation may largely explain the effects of gestational age on reproducibility. As fetal CTI may be of value in the identification of pathologies such as intrauterine growth retardation or fetal cardiac hypertrophy, their reliable estimation will be enhanced by strategies which take these factors into account.

Keywords: cardiac time intervals, reproducibility, fetal magnetocardiography

## 1. Introduction

Magnetocardiography is a non-invasive and reliable technique to detect the fetal cardiac electrophysiological signal in the second and third trimester of pregnancy. Using low temperature (LTS) superconducting quantum interference devices (SQUID) as sensors, signal quality is generally sufficient to identify fetal QRS complexes at the 20th week of gestation (van Leeuwen *et al* 1999a) and fetal ventricular depolarization has been detected as early as the 13th week (Dunajski and Peters 1995). Due to lower source strength and variation in fetal position, atrial depolarization is only recognizable in about one third to one half of raw fetal magnetocardiogram (FMCG) signals, more so later in gestation. Although high temperature (HTS) SQUID devices are not as sensitive as LTS sensors, recent design improvements have allowed the identification of fetal QRS complexes using such sensors (van Leeuwen *et al* 2003).

In order to study FMCG signal morphology dependably, signal quality is often improved through signal processing, one mainstay being the identification of QRS complexes, their alignment and subsequent averaging of a sufficient number of signal segments containing the PQRS course. This results in a signal in which the onsets and ends of atrial and ventricular depolarizations can be determined with high precision.

A number of studies have been published in which the duration of cardiac time intervals (CTI) such as the P-wave and the QRS complex have been determined on the basis of FMCG. Using a single-channel biomagnetometer, Quinn and co-workers were among the first to demonstrate the increase of QRS duration between the 20th and 42nd weeks (Quinn *et al* 1994). Subsequent studies with different multichannel systems confirmed that, over a similar gestational period, the duration of QRS increased from around 35 ms to around 50 ms (van Leeuwen 1997, Menendez *et al* 1998, Horigome *et al* 2000, Kähler *et al* 2002). Similarly, it has been shown that the P-wave increases from about 35 ms to about 60 ms (van Leeuwen 1997, Menendez *et al* 1998, Kähler *et al* 2002) and the PR interval from about 90 ms to about 100 ms (Quinn *et al* 1994, Menendez *et al* 1998, Horigome *et al* 2000). Previous work has also shown that the QT interval varies over a wide range of values between about 100 and 300 ms with an average increase of approximately 40 ms over the last 20 weeks of gestation (Quinn *et al* 1994, van Leeuwen *et al* 1999b, Horigome *et al* 2000, Kähler *et al* 2002).

Although the data in the published studies demonstrate similar values and trends, discrepancies have also been noted. In a multicentre study examining the duration of the CTIs during pregnancy, systematic differences could be noted between the values obtained by the participating centres (Stinstra *et al* 2002). Various reasons may be forwarded to explain these differences. The biomagnetometer used in acquiring the FMCG may differ with respect to the sensor type (magnetometer, first- or second-order gradiometers), the number of available channels as well as system noise and the use of shielding. Furthermore, the gestational age, the presentation and lie of the fetus and its distance to the sensors will affect signal strength and morphology. A role will also be played by the procedure used in signal processing: filter settings, the elimination of the maternal signal and the selection procedure used to identify fetal beats for averaging. Finally the definition of the waveform onsets and ends and interobserver differences may affect the determination of CTIs.

The aim of this work was to investigate the role of some of these factors which may influence the estimation of CTIs. In a large number of FMCG tracings, we have examined the effect of the observer as well as the effect of the number of available registration sites. Furthermore the influence of sensor type was investigated by examining magnetometer and gradiometer signal traces as well as by comparing several FMCG acquired using both HTS

and LTS systems. The CTIs examined were the P-wave, the PR and PQ interval, the QRS complex, the ST segment, the T-wave and the QT interval.

## 2. Methods

### 2.1. Subjects

We recruited 54 women in 57 healthy pregnancies ( $32 \pm 3$  years of age), 27 were para 0, 26 para I and 4 para II. Between the 15th and 42nd weeks of gestation, we recorded 253 FMCGs, on average  $4.4 \pm 4.1$  per fetus (range 1–14). The number of recordings per week of gestation lay between 7 and 16 ( $11.0 \pm 2.7$ ) in the period from the 18th to the 39th week, with 1–3 recordings in the earlier and later weeks.

### 2.2. Data acquisition

Several biomagnetometer systems were used in data acquisition. One hundred and sixty-six FMCGs were performed using a 61-channel system (Magnes 1300C, 4D Neuroimaging, San Diego). The sensing channels have a diameter of 28 mm and are arranged as four concentric rings around a central channel with an overall diameter of 32.4 cm (area of coverage: about 800 cm<sup>2</sup>). A configuration of 11 reference coils at a distance of about 18 cm to the plane of measurement is used to detect ambient noise. Intrinsic system noise is given as being less than 10 fT Hz<sup>-1/2</sup> (fT: femtotesla) for frequencies >5 Hz.

A further 87 FMCGs were recorded using a 37-channel system (Krenikon, Siemens, Erlangen). This system consists of sensing coils with a diameter of 24.5 mm configured as first-order gradiometers (baseline 7 cm) in a hexagonal grid with an overall diameter of 19 cm (area of coverage: about 283 cm<sup>2</sup>). System noise is given as less than 30 fT Hz<sup>-1/2</sup> for frequencies >2 Hz and 10 fT Hz<sup>-1/2</sup> for frequencies >10 Hz.

Two subjects, in whom data were obtained using the 61-channel system, were also examined with a single-channel HTS sensor designed and constructed in the Forschungszentrum Jülich. This system consists of 2 YBaCuO rf SQUIDs configured as an axial gradiometer (baseline 18 cm). One acquisition was done with a pickup area of 10 × 10 mm, the second acquisition with an increased pickup area (1 inch diameter using an additional flux concentrator (van Leeuwen *et al* 2003)). Noise were about 2 and 1 pT<sub>pp</sub>, (picotesla peak-to-peak) respectively (bandwidth of 90 Hz). The acquired data were compared with that obtained with the 61-channel LTS system (noise about 0.3 pT<sub>pp</sub>).

In each acquisition, data were recorded for 5 min at a sampling rate of 1 kHz and with a bandpass of 1–200 Hz. To reduce the effects of external noise, all measurements were performed in a standard shielded room (AK3b, Vacuumschmelze, Hanau). In the 37-channel system and the single-channel HTS system, further noise reduction is achieved by the gradiometer configuration of the sensors. In the 61-channel system, the signal recorded by the magnetometers may be corrected for noise and artefacts using the data from the set of reference channels which quantify field gradients. Thus the system performs essentially as a first order gradiometer when noise cancellation is applied.

### 2.3. Data analysis

In each set of data, maternal heart beats were automatically identified by correlation to a maternal QRS signal template, constructed on the basis of an average of 10–20 manually identified QRS signal traces. Using appropriate software the maternal PQRST signal

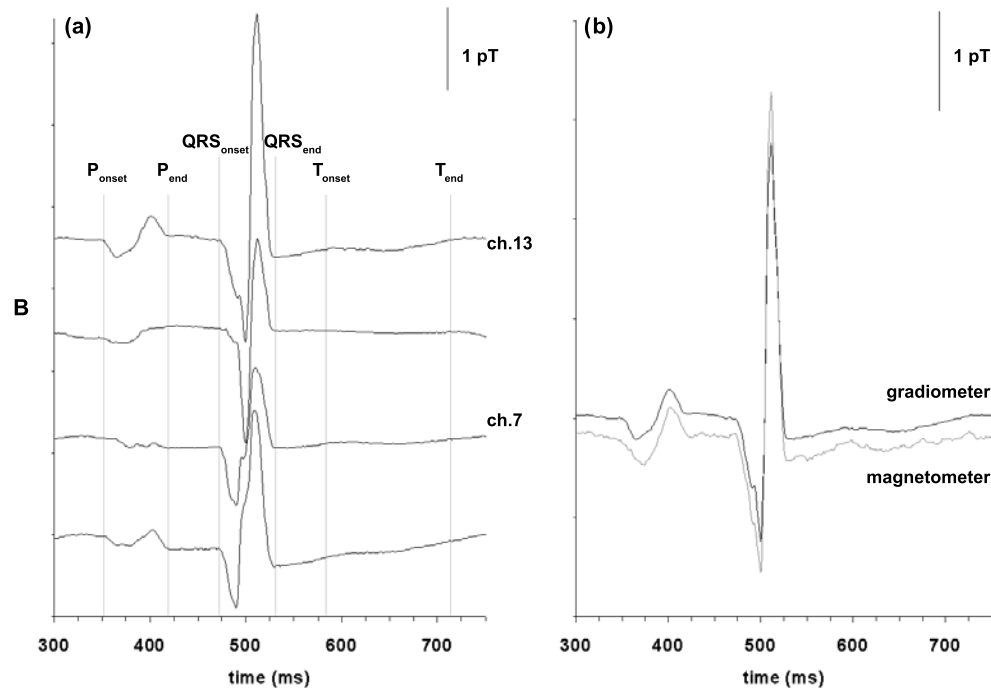
components were then digitally averaged and this average was subsequently subtracted from the signal traces. In the resulting traces, fetal beats were similarly identified on the basis of a representative fetal QRS signal template constructed using 20–30 manually selected fetal QRS traces of similar morphology. Fetal PQRST courses to be averaged were chosen on the basis of a good correlation ( $r \geq 0.90$ ) to the template and thus excluded those beats in which the signal was altered due to fetal movements or other factors strongly affecting signal shape. Heart rate dependencies can be largely excluded as one would expect them to only affect ventricular depolarization times in data sets in which heart rates vary over substantial ranges. We thus presumed that the 600–800 fetal signal traces included in the average were generally constant within any one data set.

In the averaged data, the onsets (or endpoints) of the P-wave, QRS complex and T-wave were defined as the visually identifiable first (or last) deviation from the signal's baseline in any one channel out of all channels available for evaluation. The channels were displayed individually avoiding overlap and within a fixed time window ( $20 \text{ ms cm}^{-1}$ , monitor resolution about  $40 \text{ pixel cm}^{-1}$ ) and a defined range of amplitudes ( $0.1\text{--}0.5 \text{ pT cm}^{-1}$  for the P- and T-wave,  $0.5\text{--}1.0 \text{ pT cm}^{-1}$  for the QRS complex). The onsets and ends were independently identified, manually by cursor placement, by two experienced investigators and their results were compared. We considered a difference  $\leq 2 \text{ ms}$  in these results as essentially 'identical,' these were averaged and all other results were re-examined. As each observer located the onset or end point according to accepted criteria (Grimm *et al* 2003), a difference between observers would be due to a different interpretation of 'deviation from the signal baseline.' The correction process consisted of both observers exchanging their perceptions of this deviation and agreeing on its timing. The identified time points were then used to calculate the duration of the CTI as follows: P-wave =  $P_{\text{end}} - P_{\text{onset}}$ , PQ =  $QRS_{\text{onset}} - P_{\text{end}}$ , QRS complex =  $QRS_{\text{end}} - QRS_{\text{onset}}$ , ST segment =  $T_{\text{onset}} - QRS_{\text{end}}$  and T-wave =  $T_{\text{end}} - T_{\text{onset}}$  (see figure 1(a)). Furthermore the PR interval was determined as (P-wave + PQ interval) and QT interval as (QRS complex + ST segment + T-wave). In some traces signal quality did not permit the unequivocal identification of the timing of an event and the corresponding CTI were not determined.

Following approaches were used to examine reproducibility whereby the effect of gestational age was considered where appropriate.

*2.3.1. Interobserver reproducibility.* The absolute difference between both observers, before final correction, in the identification of the signal onsets and ends in the 61-channel and 37-channel data sets was calculated and evaluated.

*2.3.2. Dependency on the number of channels.* The effect of the number of channels available for evaluation was examined in two ways. First, we selected 21 FMCGs recorded with the 61-channel system: seven around the 20th, 30th and 40th weeks of gestation, respectively. In each of the averaged PQRST courses we determined the onsets and ends three times: (1) using one channel (the channel with the maximum peak-to-peak QRS amplitude), (2) using seven channels (the channel with maximum QRS-amplitude and its six neighbouring channels) and (3) using all of the 61 available channels. Here we compared both the timing of onsets and ends as well as the duration of the P-wave and the QRS complex. Because the identification of the T-wave was often dubious, their evaluation in this limited number of data sets was not undertaken. Second, the duration of the CTI were determined in FMCGs recorded with the 37-channel system ( $n = 87$ ) and were compared to those recorded with the 61-channel system ( $n = 166$ ). There were no fetuses recorded with both systems.



**Figure 1.** (a) FMCG signal traces from four selected channels of the averaged data of a recording performed in the 39th week of gestation showing the setting of signal onsets and ends used to determine the cardiac time intervals. Note the variation in the timing of the wave deflections and, in the T-wave, the low frequency, low amplitude drift. (b) Channel 13 of the same data as in (a): the upper trace (black) shows the signal as recorded by the gradiometer configuration, the lower trace (grey), by the magnetometer configuration. In the magnetometer data, note not only the increase in noise but also the differences in wave morphology.

**2.3.3. Effect of sensor type.** In the 61-channel system, noise reduction is achieved by using weighting factors in combination with the data from the reference channels to correct for the effect of more distant signal sources, making it effectively a gradiometer system. The application of noise reduction will thus cancel some of the external signal components but it may also directly affect the signal of interest (Burghoff *et al* 1997). In order to investigate this effect, we determined the durations of the P-wave, PQ interval and QRS complex in seven data sets (used in section 2.3.2, acquired around the 40th week of gestation) in two ways: one based on the standard data processing procedure including noise reduction (gradiometer configuration) and one based on the raw, magnetometer data (magnetometer configuration). The timing of onsets and ends as well as the duration of these CTI were then compared between the magnetometer and the gradiometer configurations.

Furthermore, the duration of the CTI were examined in the two cases in which both LTS and HTS acquisitions had been performed. Here we compared the durations of the P-wave, PQ interval and QRS complex determined in the single-channel HTS data to those determined in that channel of the 61-channel system which was positioned at approximately the same site as the HTS channel. Due to the uncertainties involved in identifying the T-wave we did not attempt to evaluate the corresponding intervals.

**Table 1.** Absolute differences (ms) between two independent observers with respect to the timing of events, summarized for all data sets as well as for those determined using the 61-channel system (61 ch) and the 37 channel system (37 ch).

		<i>n</i>	Missed events	Identification rate (%)	Absolute differences between observers				
					Minimum	First quartile	Median	Third Quartile	Maximum
P <sub>onset</sub>	All	252	1	99.6	0	0	1	3	17
	37 ch	87	0	100.0	0	0	2	3	14
	61 ch	165	1	99.4	0	1	1	3	17
P <sub>end</sub>	All	252	1	99.6	0	1	2	3	19
	37 ch	87	0	100.0	0	1	2	3	19
	61 ch	165	1	99.4	0	1	1	3	16
QRS <sub>onset</sub>	All	253	0	100.0	0	0	1	2	9
	37 ch	87	0	100.0	0	0	1	2	9
	61 ch	166	0	100.0	0	0	1	2	7
QRS <sub>end</sub>	All	253	0	100.0	0	0	1	3	10
	37 ch	87	0	100.0	0	1	1	3	10
	61 ch	166	0	100.0	0	0	1	2	10
T <sub>onset</sub>	All	231	22	91.3	0	3	7	16	82
	37 ch	80	7	92.0	0	3	7	20	62
	61 ch	151	15	91.0	0	3	6	15	82
T <sub>end</sub>	All	228	25	90.1	0	2	8	20	120
	37 ch	78	9	89.7	0	2	9	32	120
	61 ch	150	16	90.4	0	3	7	19	90

#### 2.4. Statistics

Values are presented as means  $\pm$  standard deviation (SD) or, for skewed data, as measures of location based on quantiles (minimum, 1st quartile, median, 3rd quartile, maximum). Interobserver differences between systems were compared using the Mann-Whitney U test. Simple linear regression was used to quantify the dependency of interobserver differences on gestational age. In the examination of the CTI using 1, 7 or 61 channels, the difference between the estimates of the combinations of two-channel configurations (1 versus 7, 1 versus 61, 7 versus 61 channels) was compared to zero using Student's *t*-test for dependent samples. The same test was used for examining the differences in CTI as determined using magnetometer and gradiometer configurations. In order to examine the effect of 37- and 61-channel configurations in the determination of the CTI, multivariate linear regression with respect to gestational age and system (binary variable: 37-channel system = 0, 61-channel system = 1) was performed. Statistical significance was assumed when  $p < 0.05$ .

### 3. Results

#### 3.1. Interobserver reproducibility

The observers were able to find almost all events in the signal traces. Whereas the depolarization times posed virtually no difficulties, low T-wave signal amplitude reduced the rate of identification of onsets and ends to about 90% (table 1). Agreement between the observers was highest in QRS<sub>onset</sub> followed by QRS<sub>end</sub>, P<sub>onset</sub> and P<sub>end</sub> with median values at or close to the temporal resolution of the signal (table 1). Overall the timing of the QRS complex could be more consistently determined than that of the P-wave. Difficulties in the

**Table 2.** Dependency of interobserver differences on gestational age: correlation coefficients ( $r$ ), linear regression parameters and the corresponding estimated change in timing values over the 27 weeks of observation ( $p$ -values for the slope  $<0.05$  in bold type).

	$r$	Slope	y-intercept (ms)	$p$	Change over period of observation (ms)
P <sub>onset</sub>	-0.06	-0.03	3.1	0.368	-0.7
P <sub>end</sub>	-0.15	-0.08	5.3	<b>0.019</b>	-2.2
QRS <sub>onset</sub>	-0.14	-0.03	2.1	<b>0.031</b>	-0.8
QRS <sub>end</sub>	-0.29	-0.09	4.3	<b>0.000</b>	-2.3
T <sub>onset</sub>	-0.11	-0.23	18.9	0.102	-6.3
T <sub>end</sub>	-0.13	-0.44	29.5	<b>0.047</b>	-12.0

**Table 3.** Absolute differences (ms) between two independent observers with respect to the CTI as calculated on the basis of their respective identification of waveform boundaries.

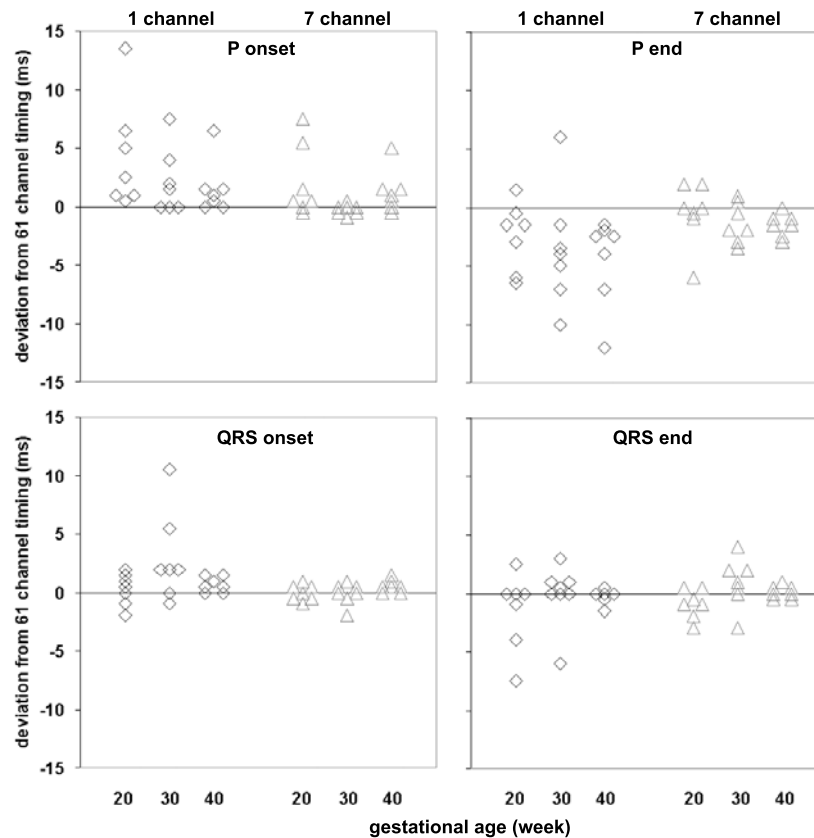
	$n$	Minimum	First quartile	Median	Third Quartile	Maximum
P-wave	252	0	1	3	7	24
PQ interval	252	0	1	2	4	18
PR interval	252	0	1	2	3	17
QRS complex	253	0	1	2	4	14
ST segment	231	0	4	8	16	82
T-wave	228	0	5	13	26	114
QT interval	228	0	3	8	22	120

recognition of T-wave onset and end led to a higher variability between the observers with median values around 7–8 ms and differences greater than 20 ms in roughly 25% of the identifiable events. The biomagnetometer system used played no role in reproducibility: there were no statistically significant differences in interobserver variability between the 37- and 61-channel configurations. There was a weak correlation of interobserver reproducibility to gestational age: the differences tended to be smaller towards the end of pregnancy (table 2). The improvement in agreement between the observers over the period of observation (27 weeks), estimated on the basis of regression analysis, was about 1–2 ms for the P-wave and QRS complex and above 6 ms for the T-wave, whereby the waveform onsets showed more constancy than the ends.

The interobserver differences in the identification of events led to corresponding variability in the estimation of the CTI. Table 3 shows that the deviations were small for the P-wave, the PQ and PR intervals as well as the QRS complex, the median differences representing approximately 5% or less of these CTI values. The intervals involving repolarization demonstrated less agreement, the median difference for the T-wave corresponding to approximately 10% of the T-wave duration. The differences in the QT interval may be attributed to the difficulty in consistently identifying T end.

### 3.2. Dependency on the number of available channels

For the evaluation of the 21 data sets in which the waveform boundaries were determined on the basis of 1, 7 and 61 channels, the interobserver differences were not dependent on the number of channels available for evaluation. On the presumption that the 61-channel configuration permitted the most accurate estimation of the onsets and ends of the waveform



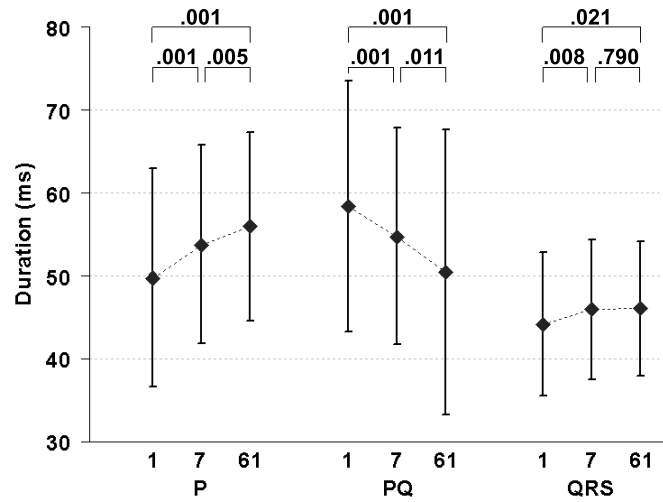
**Figure 2.** Timing of P-wave and QRS boundaries using 1- or 7-channel configurations relative to the timing determined using the 61-channel configuration ( $\diamond$ :  $\text{timing}_{61 \text{ channels}} - \text{timing}_{1 \text{ channel}}$ ;  $\triangle$ :  $\text{timing}_{61 \text{ channels}} - \text{timing}_{7 \text{ channels}}$ ). Values (in ms) are shown for the results obtained around the 20th, 30th and 40th week of gestation.

boundaries, the estimation of the timing of events using the 1- or 7-channel configuration was compared to that using 61 channels. The results indicated that using fewer channels led to a delayed setting of signal onsets and a premature setting of signal ends (figure 2) and this was particularly clear for the P-wave. Furthermore, greater shifts in timing were more likely to occur earlier in pregnancy.  $\text{QRS}_{\text{onset}}$  was least affected by either the number of channels or by week of gestation.

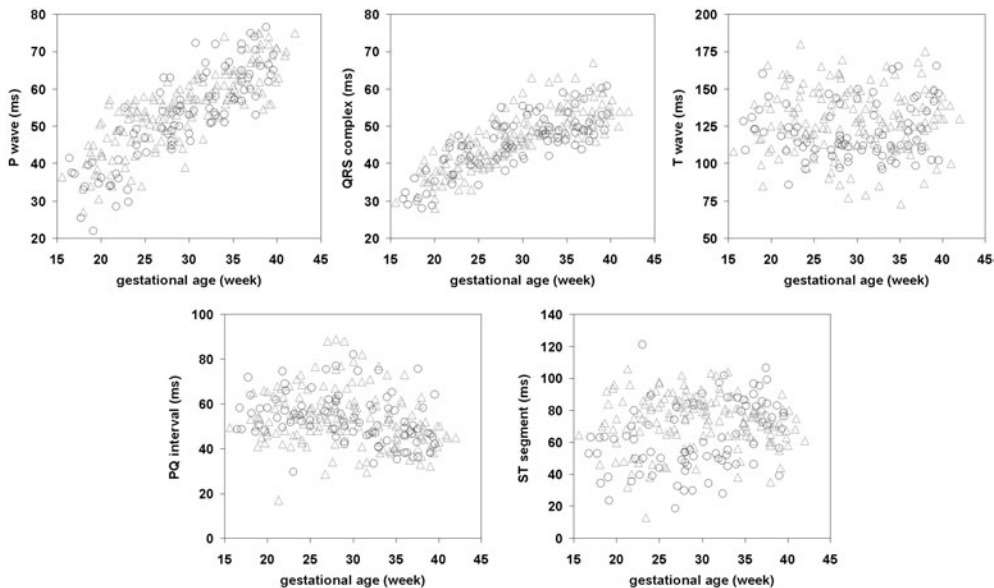
The dependency of the perceived timing of events on the number of channels analysed led to an increase in the duration of the P-wave and the QRS complex and a decrease in the duration of the PQ intervals when more channels were examined. Figure 3 shows that these differences are particularly clear for the P-wave and the PQ interval and that the differences between the 7- and 61-channel configurations were not as large as those to the 1-channel configuration.

Visually comparing the values of the CTI determined in the 87 data sets recorded with the 37-channel biomagnetometer to the 166 data sets recorded with the 61-channel biomagnetometer while taking gestational age into account revealed a clear difference only for the ST segment (figure 4): lower values occurred more often for the 37-channel system.





**Figure 3.** Duration of the P-wave, PQ interval and QRS complex (mean  $\pm$  SD in ms), estimated on the basis of 1, 7 and 61 channels in 21 data sets; *p*-values were determined on the basis of a paired *t*-test.



**Figure 4.** Duration of the cardiac time intervals with respect to gestational age, determined in FMCG recorded with a 37-channel (O, *n* = 87) and a 61-channel ( $\Delta$ , *n* = 166) biomagnetometer.

We examined the differences in CTI duration using multivariate linear regression analysis in which the effect of gestational age and the biomagnetometer used were considered. This showed that the duration of the P- and T-wave, the ST segment and the QT interval were estimated shorter for the system with fewer channels (table 4). Furthermore, with respect to the correlation to gestational age, the analysis also confirmed that the P-wave and the QRS complex are dependent on age. The coefficients of determination suggested only a

**Table 4.** Dependency of the CTI on biomagnetometer system (37 channel = 0, 61 channel = 1) and gestational age: coefficient of determination ( $r^2$ ), regression coefficients and corresponding  $p$  values estimating the effect of gestational age and system on the CTI ( $p$ -values <0.05 in bold type).

	$r^2$	Coefficient (age)	$p$	Coefficient (system)	$p$
P-wave	0.68	1.37	<b>0.000</b>	2.18	<b>0.008</b>
PQ interval	0.07	-0.48	<b>0.000</b>	-0.29	0.846
PR interval	0.18	0.89	<b>0.000</b>	1.90	0.264
QRS complex	0.63	0.93	<b>0.000</b>	0.90	0.142
ST segment	0.08	0.34	0.063	10.49	<b>0.000</b>
T-wave	0.02	0.11	0.594	5.84	<b>0.037</b>
QT interval	0.16	1.38	<b>0.000</b>	16.85	<b>0.000</b>

**Table 5.** Values of CTI (ms) as determined using a low (LTS) and a high (HTS) temperature SQUID system in two different fetuses, one in the 37th and one in the 38th week of gestation. The deviation from the values obtained with 61 channels are given in parentheses.

	37th week			38th week		
	LTS <sub>61</sub>	LTS <sub>1</sub>	HTS	LTS <sub>61</sub>	LTS <sub>1</sub>	HTS
P-wave	61	50 (18%)	44 (28%)	67	34 (49%)	38 (43%)
PQ interval	55	64 (16%)	67 (22%)	40	46 (15%)	46 (15%)
QRS complex	54	51 (6%)	44 (19%)	49	47 (4%)	52 (6%)

LTS<sub>61</sub>, values based on 61 channels; LTS<sub>1</sub>, values based on the channel closest to the position of the single HTS channel.

minimal effect of age on the PQ, PR and QT intervals, negative in the case of the PQ interval (table 4).

### 3.3. Effect of sensor type

*Gradiometer and magnetometer.* The estimated timing of the P-wave onset and end differed significantly between the gradiometer and magnetometer configurations: the difference ( $t_{\text{magnetometer}} - t_{\text{gradiometer}}$ ) was  $5.2 \pm 4.6$  ms for  $P_{\text{onset}}$  ( $p = 0.023$ ) and  $8.8 \pm 6.2$  ms for  $P_{\text{end}}$  ( $p = 0.010$ ), indicating that these events were identified later in the magnetometer configuration. The differences for the QRS complex were minimal and not statistically significant:  $0.6 \pm 1.6$  for  $QRS_{\text{onset}}$  and  $-0.2 \pm 0.8$  for  $QRS_{\text{end}}$ . This led to virtually no difference in the calculated durations of the QRS complex (gradiometer versus magnetometer:  $52.6 \pm 3.7$  ms versus  $51.8 \pm 4.4$ ,  $p = 0.280$ ), a slightly shorter P-wave using gradiometers ( $67.8 \pm 52.2$  ms versus  $71.4 \pm 9.8$ ,  $p = 0.224$ ) and a significantly longer PQ interval ( $42.9 \pm 7.6$  ms versus  $34.7 \pm 4.1$ ,  $p = 0.020$ ).

*LTS and HTS.* Compared to the values based on 61 channels, the values of the P-wave and QRS complex determined using only one channel, whether LTS or HTS, were lower, and those of the PQ interval were higher (table 5). This is in agreement with the results on the effect of the number of channels, described above. For the fetus in the 37th week, the single-channel LTS results were closer to the 61-channel results (6–18% deviation) than those of the HTS results (19–28% deviation). For the fetus in the 38th week, the difference in accuracy for the single channel LTS and HTS results did not differ much (4–49% versus 6–43% deviation, respectively).

#### 4. Discussion

The reliable determination of relevant parameters is a requirement for the clinical application of fetal magnetocardiography. It has been shown that fetal CTI may be of value in the identification of pathologies such as intrauterine growth retardation (van Leeuwen *et al* 2001) or cardiac hypertrophy (Pardi *et al* 1986). Work comparing the CTI of healthy fetuses between different laboratories has shown that systematic differences in their determination may be presumed (Stinstra *et al* 2002). Here we have examined some of the factors which may underlie these discrepancies.

One important result is that differences in event timing between observers who have agreed on how to determine signal offsets and ends are minimal. This agreement incorporates the definition of the signal onset and end as well as the viewing modalities such as the time and amplitude windows and channel display. Defining events as the first or last identifiable deflection from the zero-line in any of the channels available maximizes the accessible information and avoids incorrect estimation of the CTI (see also figure 1(a)). This definition has been explicitly accepted by the majority of groups working in this field (Grimm *et al* 2003). As perceived signal deflections are dependent on the time window and amplitude gain (Kautzner 2002), using predefined windows avoids systematic under and overestimation. Also, we have not used the so-called butterfly plot (overlapping display of all channels) as we found that noise and varying dc offsets tended to mask low amplitude changes otherwise visible when viewing the channels individually.

We found that reproducibility between observers was not influenced by the biomagnetometer used or number of channels available for examination. We did note a small but consistent effect of gestational age which may be attributed to the role of the signal-to-noise ratio (SNR). The increase in fetal signal amplitude during pregnancy (van Leeuwen *et al* 1999b, Li and Wakai 2002) should enable a more unambiguous identification of the deviation from the zero line. Indeed, our results show that the end of the waveforms were generally more ambiguous than their respective onsets but that the precision in their determination also tended to increase more as pregnancy progressed (see tables 1 and 2). The effect of SNR was also most certainly responsible for the difficulty in identifying the T-wave as ventricular repolarization mostly resulted in signals with a slow, low amplitude drift (figure 1(a)). Accordingly, the agreement between observers was lowest for intervals involving the T-wave. In contrast, the degree of agreement in the majority of the events defining the P-wave and QRS complex was high: one may expect a divergence of only a few milliseconds in most cases. However, some signal traces did occur in which the discrepancy was great enough to warrant the use of two independent observations in order to avoid gross misappraisal in such cases. A higher degree of certainty would certainly be obtainable using more than two observers. However, practical constraints will probably preclude this option in most situations and, as we have shown, if the observers have agreed on the method, differences will be minimal.

An alternative to manual procedures is the automatic identification of waveform boundaries on the basis of appropriate software. In the assessment of the adult cardiac signal, various approaches have been suggested and implemented with modest to good success. Their application to fetal signals has, to our knowledge, not taken place. On the other hand, numerous methods for detecting fetal ventricular activity exist and these may be used as a starting point in the detection of onsets and termination of the waveforms. Although one must expect more difficulties in the correct identification in the fetal signal because of the lower signal-to-noise ratio, automatic procedures offer the prospect of enhanced reproducibility. The results presented here might serve as a reference for future work in this domain.

The area covered during data acquisition plays a role in the accurate determination of the CTI. Using more channels of the same data set led to longer P-waves and QRS complexes, indicating that inadequate coverage will lead to an underestimation of these times. This effect was greater in the P-wave as well as earlier in pregnancy, confirming the role of the SNR. For the P-wave, even the use of the 37-channel system still led to shorter values when compared to the larger 61-channel system: with the concomitant increase in the sensor area of coverage from 284 cm<sup>2</sup> to 800 cm<sup>2</sup>, we found an increase of on average 2 ms. We interpret the apparent shortening of the PQ interval with gestational age as being not so much a physiological change but rather another consequence of SNR. Higher relative noise levels will lead to delayed identification of a signal upstroke, more so if the upstroke is not rapid. Similarly higher relative noise levels will be associated with a premature detection of the end of the signal, the true end being masked by the noise. Improvement in the detection of P<sub>end</sub> due to the higher P-wave amplitudes later in pregnancy should lead to a later identification of this event. This would mean that the PQ interval is overestimated at earlier gestational ages but would not affect the PR interval.

Comparing the two-multichannel systems also led to discrepancies in the estimation of the ST segment and the T-wave and consequently the QT interval. The data show that for the 37-channel system a considerable number of ST segment estimations were substantially shorter than those based on the 61-channel system. It may be presumed that this was due to the low frequency and low amplitude nature of the T-wave: having 61 channels to evaluate permitted a more consistent identification of signal shifts. Overall, our impression is that the results for the ST segment, the T-wave and the QT interval should not be overestimated as the determination of the T-wave involves a high probability of error.

These results confirm that the number of channels and the area of coverage play a nontrivial role in the estimation of fetal CTI. It seems that when more information on the waveforms is available, the signal onsets are placed earlier and the signal ends later and that this is true even for larger, multichannel systems. In order to reduce interindividual variability in the estimation of the CTI, more measuring sites should be preferred over fewer. Data acquisition should place emphasis on appropriate placement of the biomagnetometer in order to include sites with optimal peak-to-peak amplitudes of the P-wave and QRS complex. It should be noted that the channel with maximum QRS amplitude is not necessarily the channel which allows the most reliable determination of onset and/or end of the P-wave (van Leeuwen *et al* 2001). This helps explain the greater dependency of the P-wave duration on the number of available channels compared to that of the QRS complex in our results. When a single- or few-channel system is being used, several sequential acquisitions over the maternal abdomen will be necessary in order to ensure the reliability of the estimation of the CTI.

It is known that the use of gradiometer sensor configurations has the advantage of reducing the effect of strong, more distant signal artefacts but this is coupled to the disadvantage of possible changes in the morphology of the signal of interest (Burghoff *et al* 1997). Our results comparing a magnetometer with respect to a gradiometer configuration show that the duration of the QRS complex is not affected but that the PQ interval duration will most likely be underestimated with a possible overestimation of the P-wave. Factors that may be invoked to explain these results include: noisier magnetometer data, change in waveform shape and change in waveform amplitude. If noise were the primary factor then, on the basis of the foregoing discussion, we would expect the timing of P<sub>onset</sub> to be delayed and P<sub>end</sub> to be earlier. As this is true only for P<sub>onset</sub>, we may assume that the changes in P-wave morphology play a major role. Indeed, the estimated rms noise was similar (magnetometer data:  $9.6 \pm 3.6$  fT; gradiometer data:  $6.3 \pm 3.9$  fT). Visual inspection of the gradiometer and magnetometer averaged data sets revealed that for both P-wave and QRS complex waveform shape and

amplitude were altered in the majority of channels (see also figure 1(b)). Due to the overall high amplitude of the QRS complex, the changes in waveform did not seem to affect the estimation of event timing. On the other hand, the relatively lower P-wave amplitude often led to a slurring of the waveform which led to a different identification of event timing. The PR interval will be changed to the extent that  $P_{\text{onset}}$  is affected.

With respect to the differences between LTS and HTS sensors, the tendency to more accurate estimation using a single LTS sensor compared to the single-channel HTS sensor most likely reflects the influence of noise. The fact that this was not as clear for the data set of the fetus in the 38th week (Table 5) may be explained by the difficulty in positioning the HTS sensor at precisely the location of one of the LTS channel position: using QRS morphology as a criterion, it was difficult finding comparable signals for this acquisition.

In summary, we investigated several factors which may affect the accuracy of determination of CTI based on fetal magnetocardiography. Using a defined procedure for the setting of signal events, interobserver reliability was good to excellent. Increasing the number registration sites led to more accurate identification of these events. The differences in wave morphology between magnetometer and gradiometer configuration led to deviations in timing whereas the differences between LTS and HTS sensors seemed to be primarily due to noise. Overall, SNR played an important role in the accurate determination of CTI and the effects of gestational age on reproducibility can largely be reduced to the changes in signal amplitude associated with fetal maturation. Reliability and reproducibility of fetal CTI will be enhanced by strategies which lead to an increase of SNR, which guarantee sufficient maternal abdomen coverage, which implement consistent evaluation procedures and which avoid differences in sensor configuration.

## References

- Burghoff M, Steinhoff U, Haberkorn W and Koch H 1997 Comparability of measurement results obtained with multi-SQUID-systems of different sensor configuration *IEEE Trans. Appl. Superconduct.* **7** 3465–8
- Dunajski Z and Peters M 1995 Development of the fetal magnetocardiograms from the 13th week of gestation onward *Biomagnetism: Fundamental Research and Clinical Applications* ed C Baumgartner, L Deecke, G Stroink and S J Williamson (Amsterdam: Elsevier Science) pp 704–8
- Grimm B, Haueisen J, Huotilainen M, Lange S, van Leeuwen P, Menendez T, Peters M J, Schleussner E and Schneider U 2003 Recommended standards for fetal magnetocardiography (fMCG) *Pacing. Clin. Electrophysiol.* **26** 2121–6
- Horigome H, Takahashi M I, Asaka M, Shigemitsu S, Kandori A and Tsukada K 2000 Magnetocardiographic determination of the developmental changes in PQ, QRS and QT intervals in the foetus *Acta Paediatr.* **89** 64–7
- Kähler C, Schleussner E, Grimm B, Schneider A, Schneider U, Nowak H and Seewald H J 2002 Fetal magnetocardiography: development of the fetal cardiac time intervals *Prenat. Diagn.* **22** 408–14
- Kautzner J 2002 QT interval measurements *Card. Electrophysiol. Rev.* **6** 273–7
- Li Z and Wakai R T 2002 Amplitude of the P and QRS components of the fetal MCG in normal and fetal arrhythmia subjects *Proc. 13th Int. Conf. Biomag.* ed H Nowak, J Haueisen, F Giessler and R Huonker (Berlin, Offenbach: VDE) pp 636–8
- Menendez T, Achenbach S, Moshage W, Flüg M, Beinder E, Kollert A, Bittel A and Bachmann K 1998 Pränatale registrierung fetaler Herzaktionen mit magnetokardiographie *Z. Kardiol.* **87** 111–8
- Pardi G, Ferrazzi E, Cetin I, Rampello S, Baselli G, Cerutti S and Civardi S 1986 The clinical relevance of the abdominal fetal electrocardiogram *J. Perinat. Med.* **14** 371–7
- Quinn A, Weir A, Shahani U, Bain R, Maas P and Donaldson G 1994 Antenatal fetal magnetocardiography: a new method for fetal surveillance? *Br. J. Obstet. Gynaecol.* **10** 866–70
- Stintra J, Golbach E, van Leeuwen P, Lange S, Menendez T, Moshage W, Schleussner E, Kähler C, Horigome H, Shigemitsu S and Peters M 2002 Multicentre study on the fetal cardiac time intervals using magnetocardiography *Br. J. Obstet. Gynaecol.* **109** 1235–43
- van Leeuwen P 1997 Applications of fetal magnetocardiography *Biomed. Tech. (Suppl. 1)* **42** 265–8

- van Leeuwen P, Hailer B, Bader W, Geissler J, Trowitzsch E and Grönemeyer DH 1999a Magnetocardiography in the diagnosis of fetal arrhythmia *Br. J. Obstet. Gynaecol.* **106** 1200–8
- van Leeuwen P, Lange S, Hackmann J, Klein A, Hatzmann W and Grönemeyer D 2001 Assessment of intra-uterine growth retardation by fetal magnetocardiography *Proc. 12th Int. Conf. on Biomagnetism* ed J Nenonen, R J Ilmoniemi and T Katila (Espoo: Helsinki University of Technology) pp 603–6
- van Leeuwen P, Lange S, Schüßler M and Lajoie-Junge L 1999b Determination of changes in fetal cardiac time intervals and T/QRS ratio during pregnancy using magnetocardiography *Proc. 10th Int. Conf. on Biomagnetism* ed C J Aine, Y Okada, G Stroink, S J Swithenby and C C Wood (Berlin: Springer) pp 581–4
- van Leeuwen P, Zhang Y, Krause H J, Wolters N and Grönemeyer D 2003 Registration of fetal cardiac activity using LTS and HTS SQUID biomagnetometers *Biomed. Tech. (Suppl. 1)* **48** 372–3