The Research Council of Norway (NFR)

NORSAR

NORSAR Scientific Report No. 1-95/96

Semiannual Technical Summary

1 April - 30 September 1995

Kjeller, November 1995

7.4 Automatic onset time estimation based on autoregressive processing

Introduction

In order to support the developments at the GSETT-3 IDC, we have during this reporting period been experimenting with the use of an autoregressive method for automatic onset time estimation, denoted AR-AIC. This method has for several years been operational in the processing of data from the Japanese national seismic network, and the software has been provided to us by scientists from the Japanese NDC.

In this paper we have adapted the Japanese method for application to GSETT-3 data, with emphasis on developing an automated procedure that includes new features such as multiple narrow-band filters, the concept of "usable bandwidth" and a quality measure of the estimated onset time.

IDC onset time estimation

We have investigated the automatic phase picking at the GSETT-3 IDC, and found that the automatic picks are consistently late compared to the onset times determined by the analysts. Fig. 7.4.1 shows some characteristic examples where the automatic onsets, denoted S, are all late. In order to quantify the bias of the automatic phase picking procedure at the IDC, we have in Fig. 7.4.2 plotted the time difference between manual and automatic onsets for all P-phases with SNR > 50 for the time period January-September 1995. We see that the automatic onsets are usually late for the entire time interval, and this behavior becomes even more pronounced during the last 3 months of the period.

A new signal processing package which is scheduled to be installed at the GSETT-3 IDC will hopefully take care of the deficiencies of the current procedure used for onset time estimation. It should be noted that the current onset estimation procedure has been adapted from the algorithm used for automatic arrival time picking at NORSAR (Mykkeltveit & Bungum, 1984), and also that our experience is that the implementation at NORSAR does not provide such delayed onsets.

Autoregressive method

We will first give a brief description of the Japanese autoregressive method for onset time estimation, and for details we refer to Kamigaichi (1994), GSE/JAPAN/40 (1992), Yokota et al (1981) and Maeda (1985).

Generally speaking, autoregressive (AR) models are employed to represent the seismic waves, and Akaike's Information Criterion (AIC) is used to determine the AR order and to estimate the arrival times of the seismic signals.

Fig. 7.4.3 illustrates the basic concepts of the method:

- An initial onset is given, either from the time of the declared STA/LTA-based signal detection or from another onset time estimator. The original data is shown in the lower panel of Fig. 7.4.3.
- AR coefficients are computed from data in two windows, one located in the noise preceding the initial onset (F-window) and another located within the signal (S-window).
- The data are filtered with two prediction error filters, derived from the AR coefficients of the F- and S-windows, respectively (see 2nd and 3rd panel of Fig. 7.4.3).
- Finally, the Akaike Information Criterion (AIC) (see upper panel) is applied as a criterion to estimate the optimal division point of the time series. This division point will be the minimum of the AIC-curve, and is taken to be the onset of the seismic signal.

The F-window was in this study defined to start 7 s ahead of the initial onset, whereas the S-window started 1 s after the initial onset. Both windows had a length of 4 s. As seen from Fig. 7.4.3, the AIC was computed for a 12 s interval, starting 7 s ahead of the initial onset. This parameterization can, of course, be adjusted to accommodate different types of applications of the method.

We will in the following discuss the AIC onset time estimation utilizing the AR-coefficients of both F- and S-windows. There is also an option for utilizing the AR-coefficients of the F-window only. This option will in the text be referred to as AIC_F or AR-AIC_F.

Performance for high SNR teleseismic signals

As a first evaluation of AR-AIC, we analyzed teleseismic GSETT-3 data with high SNR, primarily P-phases from the Chinese nuclear test on 17 August 1995. First, we picked the phase onsets manually on the raw unfiltered waveforms using the NORSAR analysis tool, EP, with high resolution graphics (Fyen, 1989). Secondly, we ran the AR-AIC method on the same data set, using the automatic onsets from the IDC processing as the initial start time. The results are shown is shown in Fig. 7.4.4, and we see that there is an excellent correspondence between the manual and AR-AIC onset estimates for these high SNR teleseismic signals. The mean time difference is less than 2 milliseconds and the standard deviation is 0.04 s. As a result of this close correspondence, we will in the following use these AR-AIC onsets as the reference. The reason for this change of reference is purely due to convenience, as we in this way avoided retyping the manual onsets.

In Fig. 7.4.5, we show the time difference between the AR-AIC onsets and the automatic (SigPro) time picks at the IDC. As expected from the results given in Fig. 7.4.2, the automatic onsets are consistently late, with a mean time difference of 0.45 s.

Similarly, we compared the analyst reviewed IDC picks with the AR-AIC onset (and indirectly also the manual picks using the EP program). The results are shown in Fig. 7.4.6, and we see that for this data set the manual picks at the IDC are often early, with a mean of about 0.2 s. We also see from the figure that there is a sub-set of the data which is in quite close agreement, whereas another sub-set is about 0.3 s early. We do not know the reason

for this time difference, but there are two factors that can be of importance. One is the limited time the IDC analysts are able to spend on refining the time picks due to the daily workload. Another possible source of error is the compensation for the group delay of the bandpass filters used prior to the phase picking. This is a topic that should be revisited, as the current procedure for time adjustment due to the group delay of the bandpass filters clearly has deficiencies.

Implementation of AR-AIC for processing of GSETT-3 data

From applying the AR-AIC method to signals with various frequency contents, signal-to-noise ratios and complexities, we found that some preprocessing was necessary to ensure stable performance of AR-AIC. In particular, an assessment of the usable bandwidth of the signal, followed by bandpass filtering and decimation was necessary when processing low SNR signals, especially at low frequencies. Once the onset time was estimated, we found it helpful to calculate an accompanying quality measure. The idea behind this quality measure was to have a tool that could be used to automatically distinguish between "good" and "bad" onsets, and possibly also to get an associated uncertainty. The flowchart for automatic operation of AR-AIC is given in Fig. 7.4.7. We will in the following describe in more detail the procedures for the assessment of the usable signal bandwidth and the quality of the AR-AIC onset.

Usable bandwidth

The estimation of the usable bandwidth of the signal was done by filtering the signal with a set of relatively narrow bandpass filters, and then for each of these filters we computed the maximum SNR (STA/LTA) within a time interval around the initial onset. The usable bandwidth was then estimated from a comparison between the maximum SNR's of the different filter bands. Specifically,

- We estimated the maximum SNR within the time interval (s-2,s+3) sec. for a set of narrow bandpass filers, where s is the initial onset time. The 3rd order Butterworth filters used in this study were (in Hz): 0.5-1.5, 0.8-1.8, 1.0-2.0, 1.5-3.0, 2.0-4.0, 3.0-5.0, 4.0-6.0, 6.0-8.0, 8.0-10.0, 10.0-16.0, 14.0-20.0. The high end of the filter bands were limited by the Nyquist frequency.
- We then found the filter band providing the highest SNR, called SNR_{max}. If the neighboring filters had SNRs within a factor 5 of SNR_{max} and at the same time had SNR > 4.5, then the usable bandwidth was extended to include these neighboring filter bands.

An example illustrating the algorithm is given in Table 7.4.1. It should be noticed that no rigorous testing has been conducted to come up with the parameters of this algorithm, but they are derived from experiments with limited data sets and from our experience with processing of seismological data.

After having estimated the usable bandwidth of the signal, we filtered the data with a 2nd order Butterworth filter for this bandwidth, and then decimated the data in accordance with the high cutoff frequency of the bandpass filter. The necessity of doing filtering and decimation for processing of low SNR signals is illustrated in Fig. 7.4.8. This signal does only have a usable SNR in the filter band 1.0-2.0 Hz, as shown in the lower panel. The result from applying AR-AIC to the unfiltered data is shown in the upper panel, where S is the initial onset time and A is the

AR-AIC onset. The second panel shows the result after filtering, but without decimation, and, finally, the third panel shows the result after both filtering and decimation. Obviously, the AR-AIC onset after filtering and decimation gives the best result.

We have also made some preliminary tests on how the application of this 2nd order causal Butterworth filter for the usable bandwidth influenced the arrival time estimates. The high SNR P-phases that were previously analyzed as shown in Fig. 7.4.4, were bandpass filtered and decimated prior to AR-AIC processing. The time differences between AR-AIC computed on unfiltered data and data filtered in the usable bandwidth are shown in Fig. 7.4.9. For this data set we can see that there is no need to introduce any corrections for the filter. But before drawing any definite conclusions on the filter effects on the onset time estimates, we need to investigate more thoroughly the effect of varying SNR, bandwidth, filter order and signal frequency content.

Quality of the onset time estimates

The uncertainty of manually determined phase onsets is obviously dependent on the SNR of the signal. In addition, manual phase picks are often accompanied with a flag indicating the instantaneous or emergent nature of the arrival.

We have during our work with the AR-AIC method found that it would be very valuable to attach to the automatically determined onsets some additional parameters that can subsequently be used to derive associated picking uncertainties. In addition, we would like to know the degree of success of the estimation procedure, e.g. in terms of a flag indicating whether the algorithm truly succeeded or possibly failed.

The human observation of a seismic phase is attributed to an amplitude increase and/or a change in the frequency content of the data. If the trace is properly filtered, an amplitude increase should be observable. In this study, we have therefore decided to derive additional signal parameters from the time domain data, filtered in the band that provides the highest SNR. To analyze the amplitude increase we found it convenient to create the envelope of the data from the bandpass filtered trace and its Hilbert transformed counterpart. The Hilbert envelope was gently smoothed with a lowpass filter. This procedure is illustrated in Fig. 7.4.10.

We defined the following set of measurements to be made on the envelope:

- $NOISE_{max}$ was taken to be the maximum of the envelope within a 3 second interval preceding the automatically estimated onset.
- $AMP_{0.5}$, $AMP_{1.0}$, $AMP_{2.0}$, $AMP_{3.0}$ and $AMP_{5.0}$ were the maximum of the envelope within 0.5, 1.0, 2.0, 3.0 and 5.0 seconds after the onset, respectively. The corresponding (quality) signal-to-noise ratios $QSNR_{0.5,...,5.0}$ were defined to be $AMP_{0.5,...,5.0}/NOISE_{max}$.
- $T_{QSNR1.5}$ was the time from the onset to the point where QSNR exceeded 1.5. $QSNR_{fp}$ was the signal to noise ratio of the first local peak of the Hilbert envelope in an interval from $T_{QSNR1.5}$ to 5 seconds after the onset. T_{fp} were the time from the onset to the first local peak, and T_{max} were the time from the onset to the point where the maximum QSNR was found (within 5 seconds of the onset).

When searching for the best frequency band for bandpass filtering, we searched among the same filters as those used for determining the usable bandwidth, but we did now use $QSNR_{3.0}$ as the criterion for determining the best filter.

In order to get an idea on how to use the envelope measurements to quantify the quality of the automatic AR-AIC onsets, we analyzed a limited data set of 122 phases associated to events in the IDC Reviewed Event Bulletin (REB). The onsets of all phases were manually picked by using the EP program to get a reference for comparing the automatic onsets. Fig. 7.4.11 shows the difference between the AR-AIC, hereafter also denoted $AR\text{-}AIC_{F+S}$, onsets and the manual picks as a function of $QSNR_{2.0}$. The data points labelled **F** represent phases that we were unable to pick manually in a confident way, primarily due to low SNR. We see from the figure that for $QSNR_{2.0}$ lower than 5, the scatter increases significantly, as the algorithm had a tendency to make an early trigger. An interesting observation during our testing was that the $AR\text{-}AIC_F$ method, utilizing only the autoregressive coefficients of the preceding noise window, often gave the correct onset in the cases where $AR\text{-}AIC_{F+S}$ made the wrong decision. For this data set, we found that by using the time difference between the two types of AR-AIC onsets together with the quality measurements $QSNR_{0.5}$, $T_{QSNR1.5}$ and SNR_{max}, we were able to obtain a rule for identifying the cases where we should use the $AR\text{-}AIC_F$ onset instead of $AR\text{-}AIC_{F+S}$. The results are given in Fig. 7.4.12, and we clearly see that the scatter at low SNR is significantly reduced (except for the low quality **F** onsets).

In automatic operation of AR-AIC it is important to identify the cases where the method failed as well as the cases where the phase onsets are very uncertain. First of all, the phases that we were unable to pick manually, labelled **F**, should be identified as a low quality onset. From utilizing the quality measurements $T_{QSNR1.5}$, $QSNR_{5.0}$ and the time difference between the initial onset and the AR-AIC onset, we were able to categorize as low quality 20 out of 22 **F** onsets, while retaining 90 out of 100 acceptable onsets. The results are shown in Figs. 7.4.13 and 7.4.14. As expected, we see from Fig. 7.4.13 that the time difference between the manual and the automatic onsets decreases with increasing $QSNR_{2.0}$. As an illustration, we separated the data into two populations based on a $QSNR_{2.0}$ of 6, and found that the standard deviation was 0.15 s for the high SNR population and 0.5 s for the low SNR population.

We have with this example shown that it is possible to use the envelope quality measurements to indicate how well the automatic AR-AIC onsets match the manual picks, as well as a tool to identify low quality onsets. In addition, the envelope quality measurements were used to decide between the use of $AR\text{-}AIC_{F+S}$ and $AR\text{-}AIC_F$. A next step will be to analyze a larger data set that also contains detections that are unassociated to seismic events. In this way we can get a better picture of the operational performance of AR-AIC and the associated quality measurements.

Conclusions

We have in this study shown that by including processes like determination of usable bandwidth, filtering, decimation and quality assessments, the AR-AIC method for onset time estimation can be adapted to work on a wide range of seismic signals. In particular, we have found it convenient to be able to distinguish between *reliable* and *unreliable*

onsets. In this way, we can avoid using erroneous arrival time data in the subsequent event location procedures, and thus being able to improve the location precision of the automatic processing system.

It is also our goal to be able to give more weight to the most reliable phase onsets. In the location procedure at the IDC this is done by associating the arrival times with a given uncertainty, currently being only a function of phase type. In order to investigate how the uncertainty of the AR-AIC onsets depends on the envelope quality measurements described above, it is necessary to analyze events for which ground truth information is available, e.g. in terms of accurate locations provided by local networks. During the next reporting period we plan to conduct such a study for a set of events located in the Japan area, with high quality locations provided by the Japanese National Seismic Network.

T. Kværna

References

Fyen, J. (1989): Event Processor program package. Semiannual Technical Summary, 1 October 1988 - 31 March 1989, NORSAR Sci. Rep. No 2-88/89, Kjeller, Norway.

GSE/JAPAN/40 (1992): A Fully Automated Method for Determining the Arrival Times of Seismic Waves and its Application to an on-line Processing System. Paper tabled in the 34th GSE session in Geneva, July 1992.

Kamigaichi, O. (1994): Automated identification of arrival time, etc. using AR-model. Paper presented at the GSE workshop in Tokyo, Japan, 14-16 March 1994.

Maeda, N. (1985): A method for reading and checking phase time in auto-processing system of seismic wave data (in japanese with English abstract), *J. Seismol. Soc. Jpn.*, 38, 365-379.

Mykkeltveit, S. and H. Bungum (1984): Processing of seismic events using data from small-aperture arrays. *Bull. Seism. Soc. Am.* 74, 2313-2333.

Yokota, T., S. Zhou, M. Mizoue and I. Nakamura (1981): An automatic measurement of arrival time of seismic waves and its application to an on-line processing system (in Japanese with English abstract), *Bull. Earthquake Res. Inst. Univ. Tokyo*, 55, 449-484.

Table 7.4.1. Example illustrating the use of multiple narrow-band filters to arrive at a "usable bandwidth", as described in the text. In this case, the usable bandwidth is 1.5-5.0 Hz

Band	SNR	Comment
1.0-2.0 Hz	4.4	Below 4.5 and below a factor 5
1.5-3.0 Hz	5.0	OK
2.0-4.0 Hz	24.3	Maximum
3.0-5.0 Hz	6.1	OK
4.0-6.0 Hz	4.6	Below a factor 5

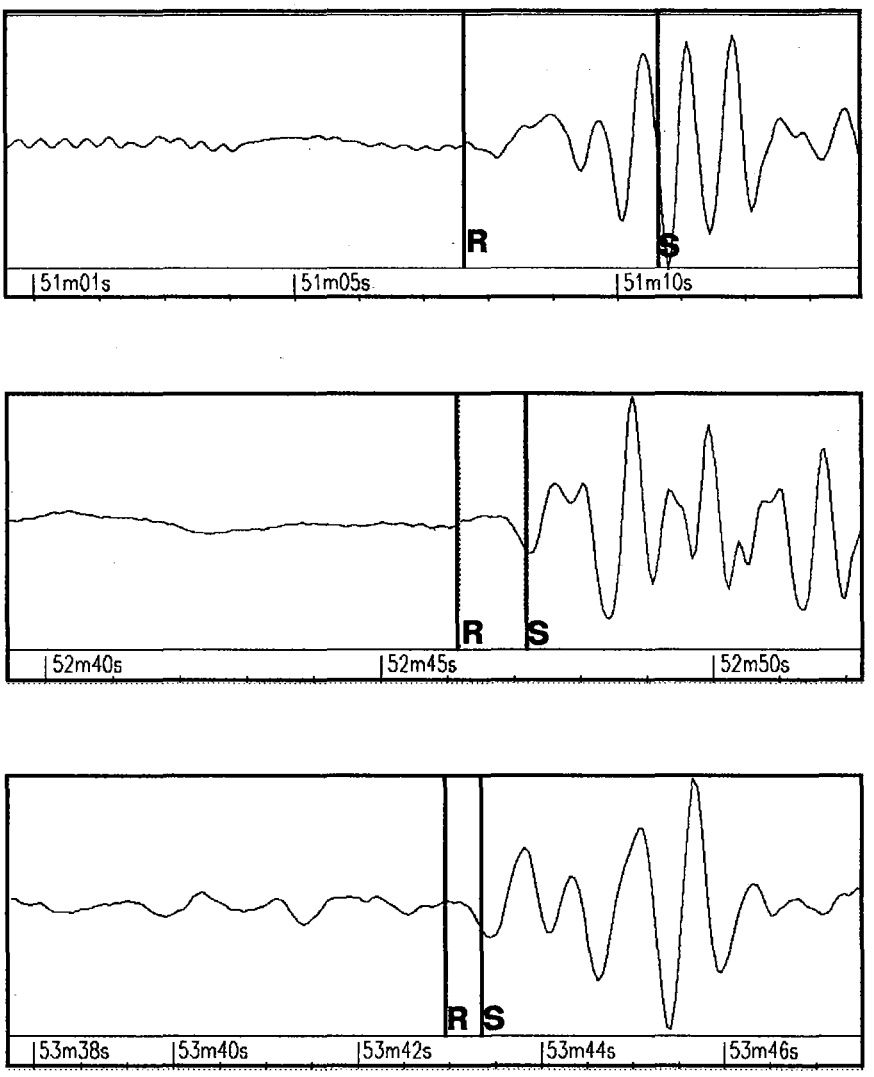

*Fig. 7.4.1. Characteristic examples of automatic (**S**) and manual (**R**) onset time estimation at the IDC.*

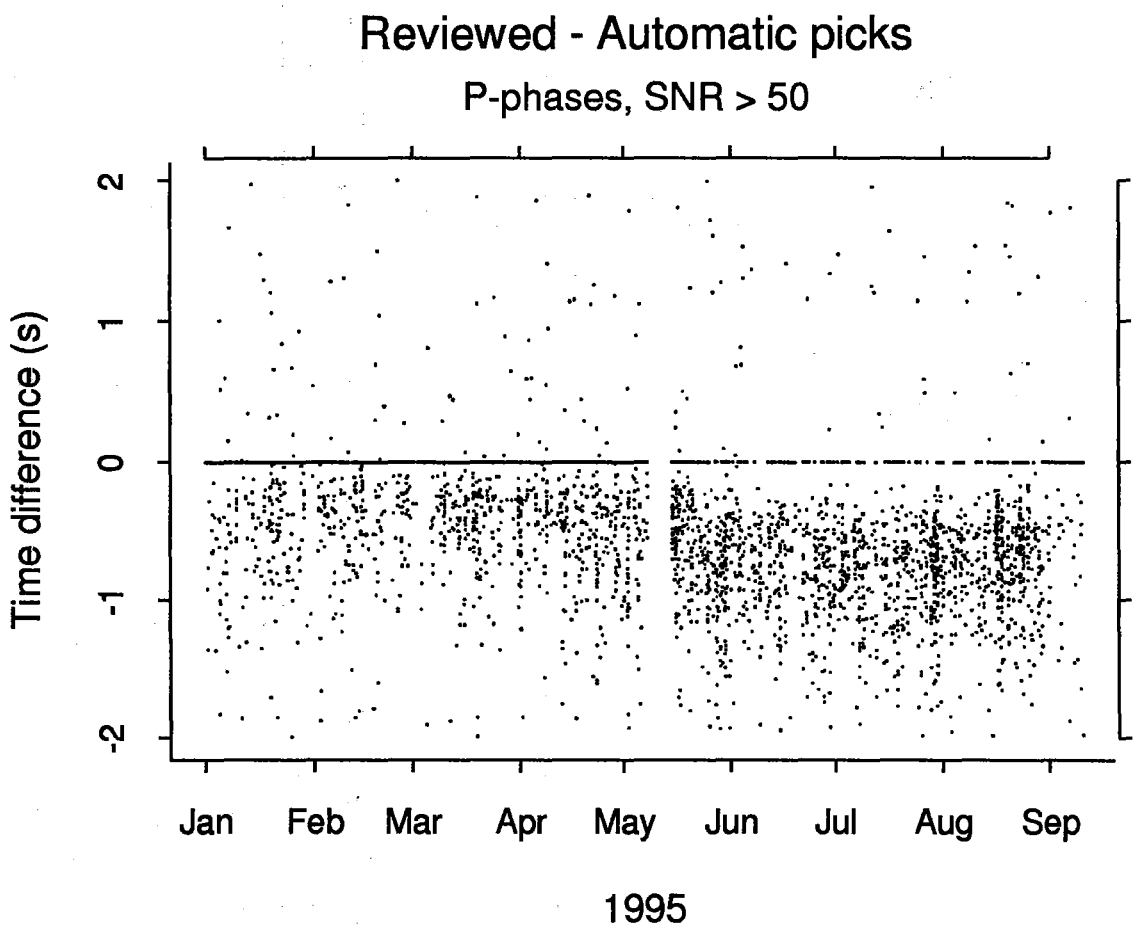

Fig. 7.4.2. Time difference between manually reviewed and automatic picks at the IDC for P-phases with SNR > 50 for the time period January-September 1995.

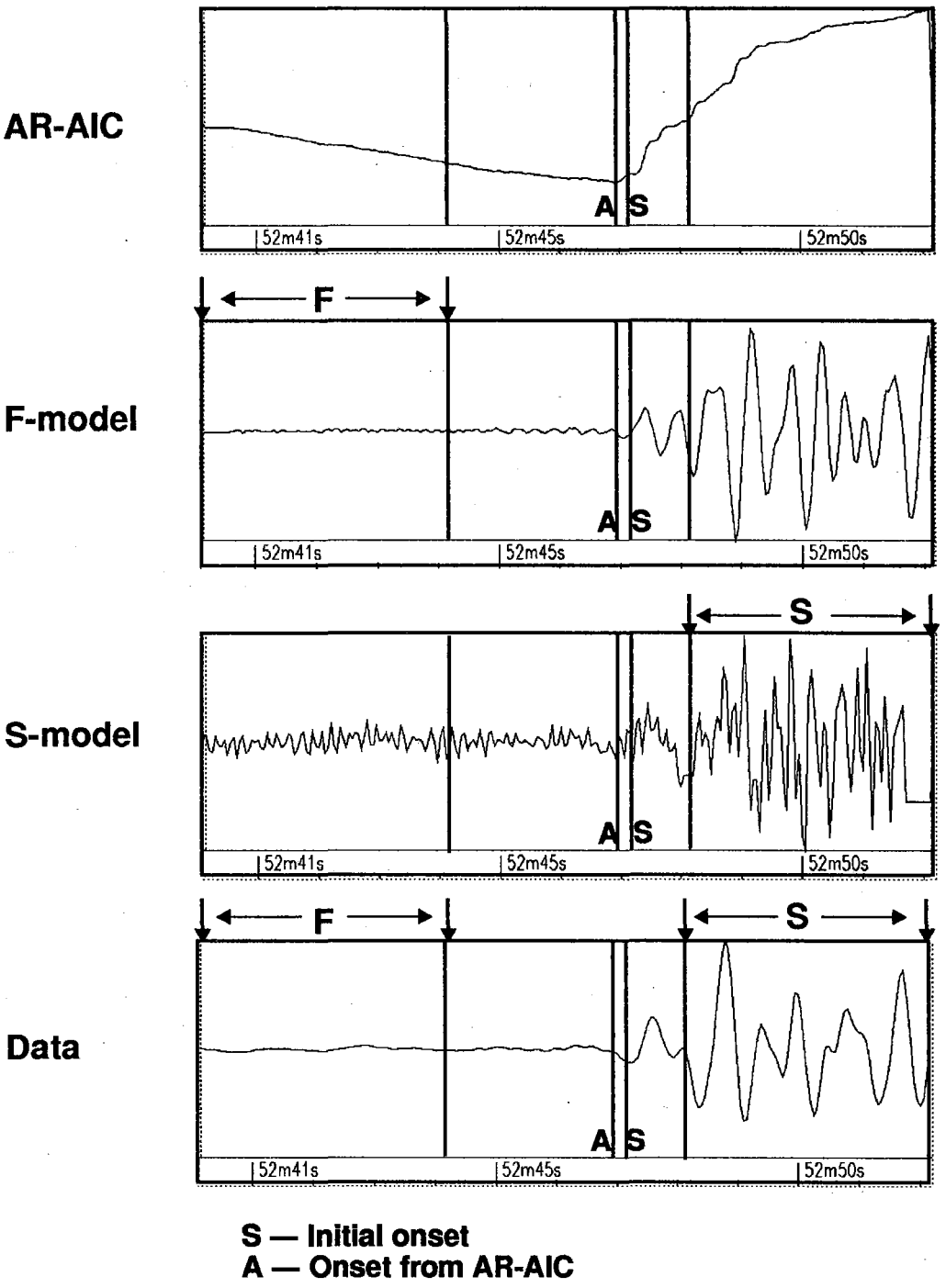

Fig. 7.4.3. Illustration of the basic concepts of onset time estimation using the AR-AIC method. The lower panel shows the data with a seismic signal. The third panel from the top shows the data filtered by a prediction error filter derived from the AR-coefficients of the 4 sec S-window positioned within the signal. The second panel from the top shows the data filtered by a prediction error filter derived from the AR-coefficients of the 4 sec F-window positioned in the noise preceding the signal. The upper panel shows the AIC used to estimate the optimal division of the time series. The minimum is taken to be onset of the seismic signal.

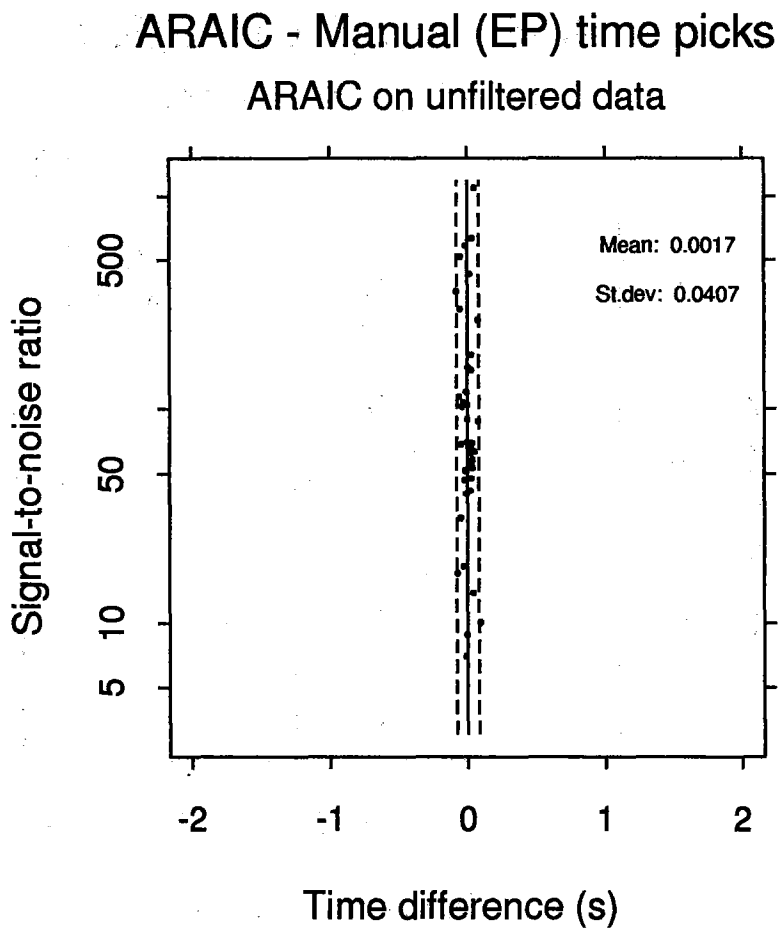

Fig. 7.4.4. Time difference between AR-AIC onsets estimated on unfiltered data and manually picked onsets (EP) for a set of high SNR teleseismic signals. The dashed lines indicate a distance of two standard deviations from the mean.

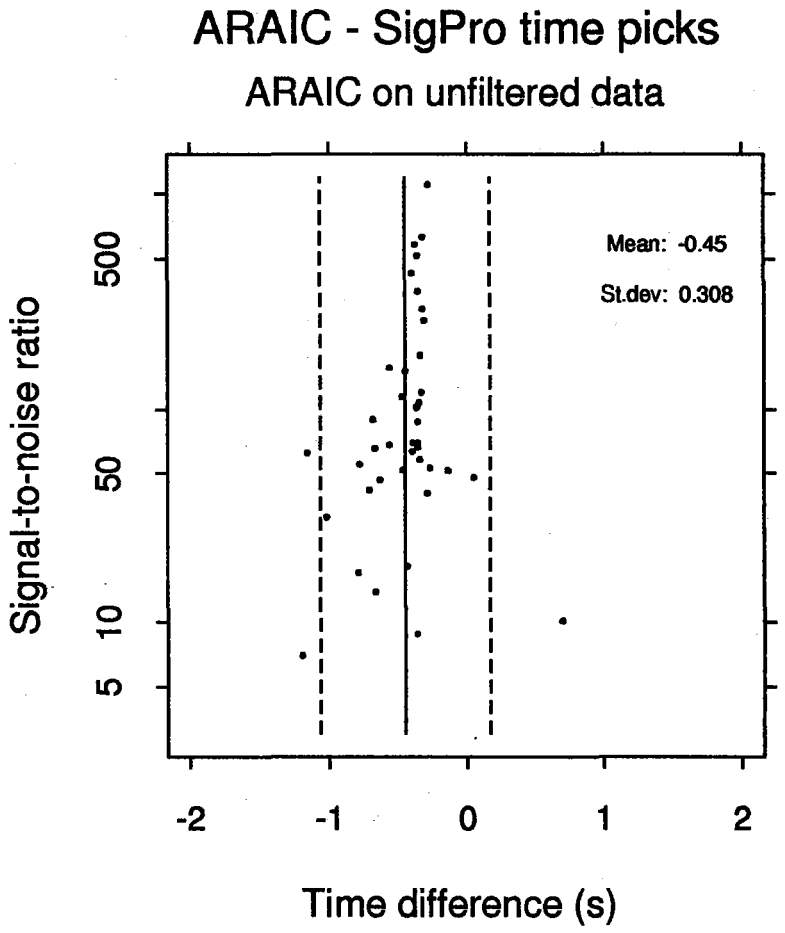

Fig. 7.4.5. Time difference between AR-AIC onsets estimated on unfiltered data and the automatic onsets provided by the signal processing at the IDC (SigPro). The data set is the same as in Fig. 7.4.4. Notice that the SigPro onsets are consistently late. The dashed lines indicate a distance of two standard deviations from the mean.

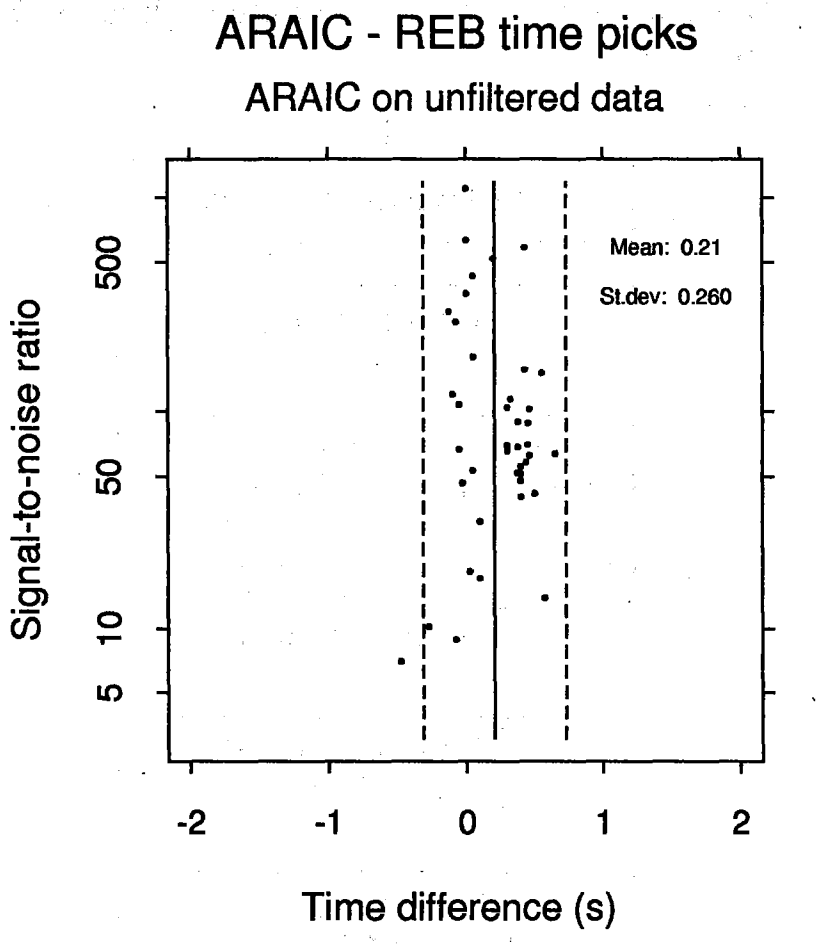

Fig. 7.4.6.Time difference between AR-AIC onsets estimated on unfiltered data and the analyst-reviewed picks at the IDC. The data set is the same as in Fig. 7.4.4. Notice that the analyst-reviewed picks at the IDC are often early. The dashed lines indicate a distance of two standard deviations from the mean.

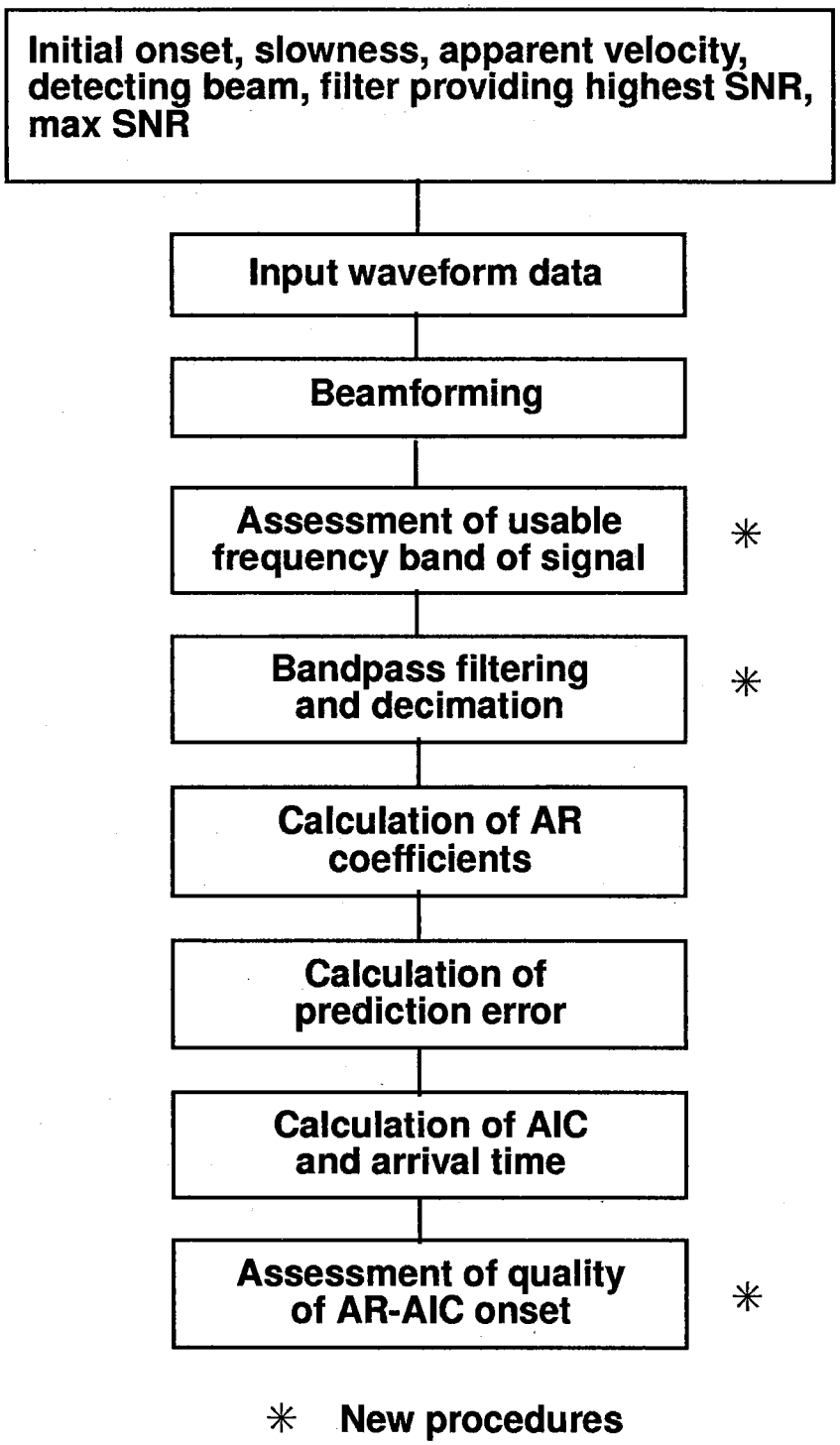

Fig. 7.4.7. Flowchart showing the different steps involved in the automatic operation of AR-AIC

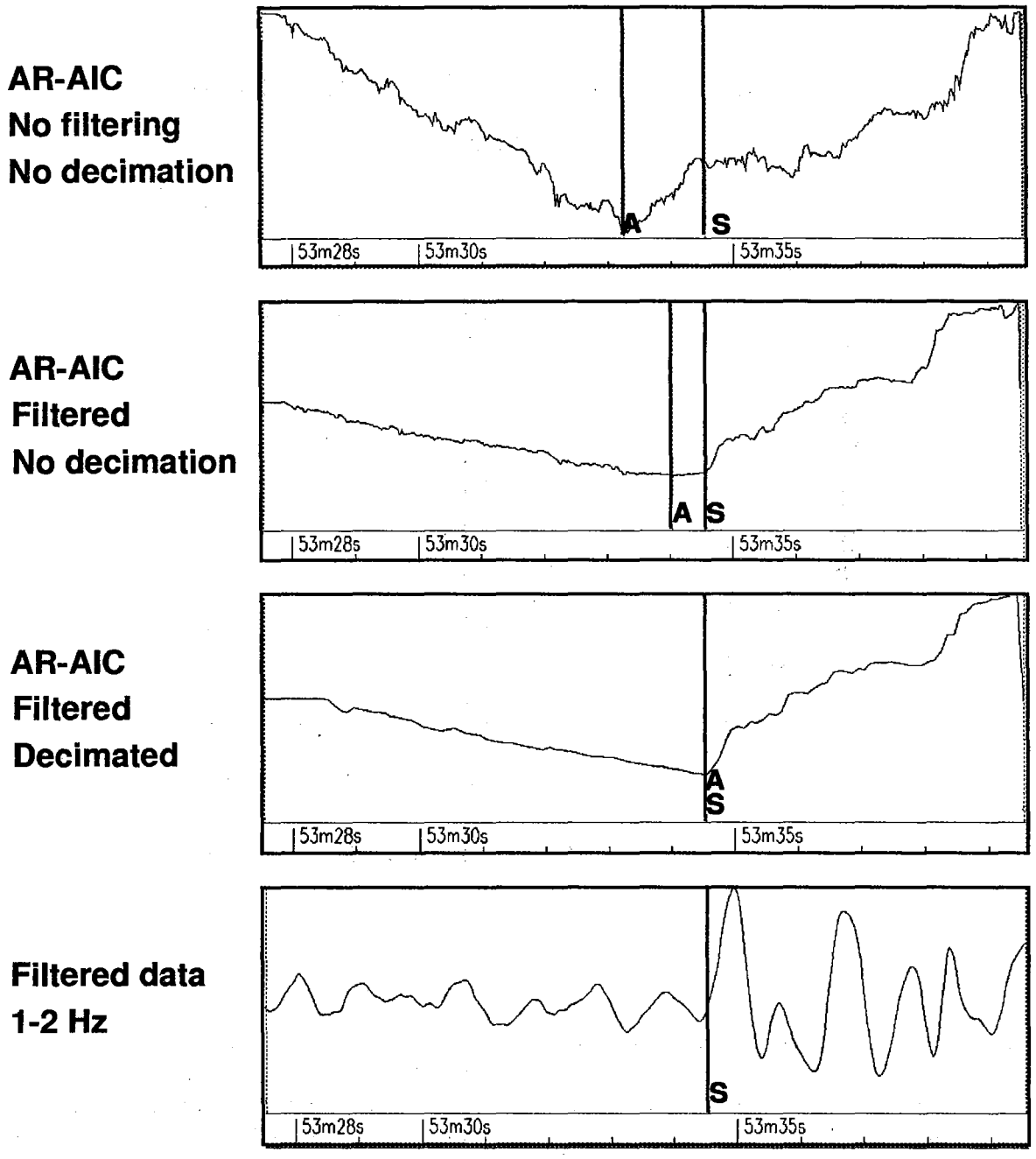

*Fig. 7.4.8. Illustration of the necessity of doing filtering and decimation prior to onset time estimation by the AR-AIC method. **S** is the initial onset, and **A** represents the AR-AIC onset. The lower trace shows the data bandpass filtered in the usable bandwidth of 1-2 Hz. The top panel shows the AIC-curve after processing the raw data. The second panel shows the AIC-curve after processing the filtered data and the third panel shows the AIC-curve after processing the filtered and decimated data. Notice that both filtering and decimation were necessary to get the correct onset.*

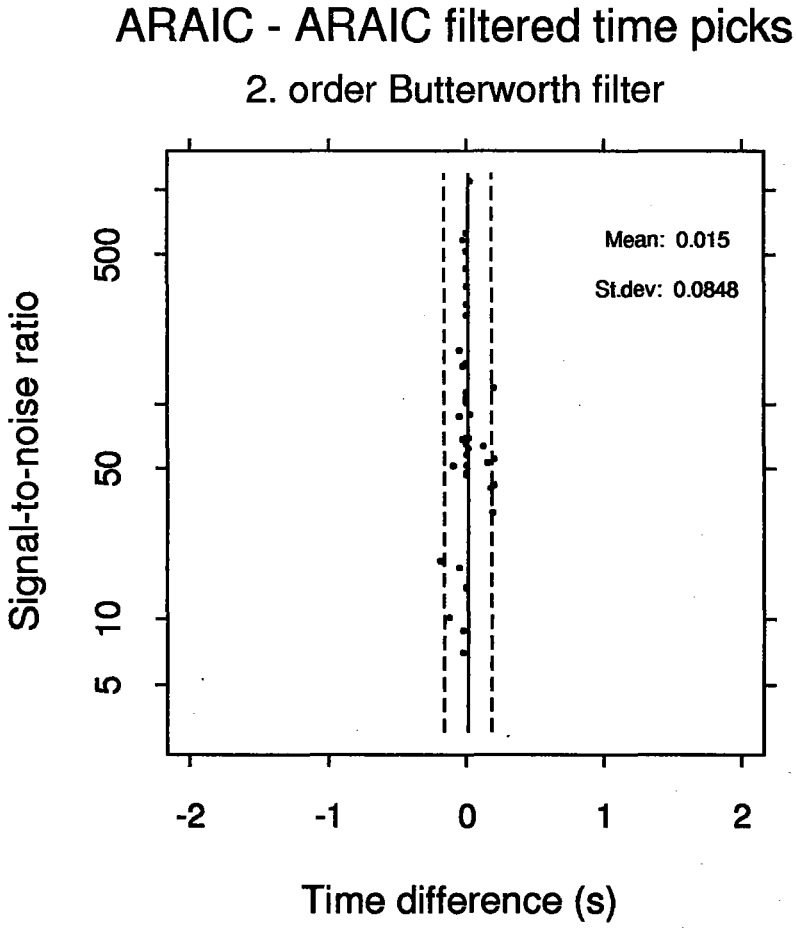

Fig. 7.4.9. Time difference between AR-AIC onsets estimated on high-SNR unfiltered data and the AR-AIC onsets estimated on data filtered in the usable frequency band. Notice the very small systematic bias. Although filtering introduces some scatter in the estimates, it is important to be aware that filtering is essential for processing low-SNR signals. The data set is the same as in Fig. 7.4.4. The dashed lines indicate a distance of two standard deviations from the mean.

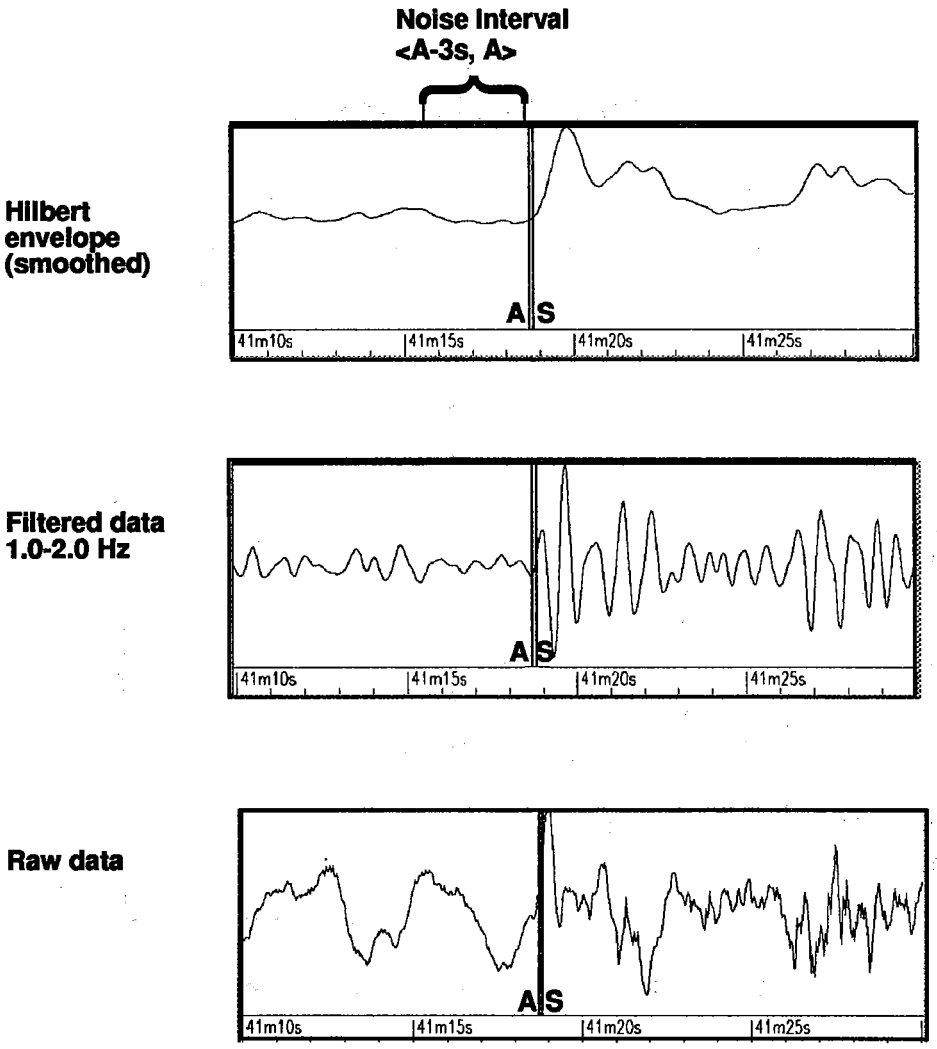

Fig. 7.4.10. Figure showing the raw data (lower panel), the data filtered in the best frequency band (middle panel) and the smoothed envelope (top panel) computed from the filtered time series and its Hilbert transformed counterpart. The 3 sec noise interval is indicated on the top panel.

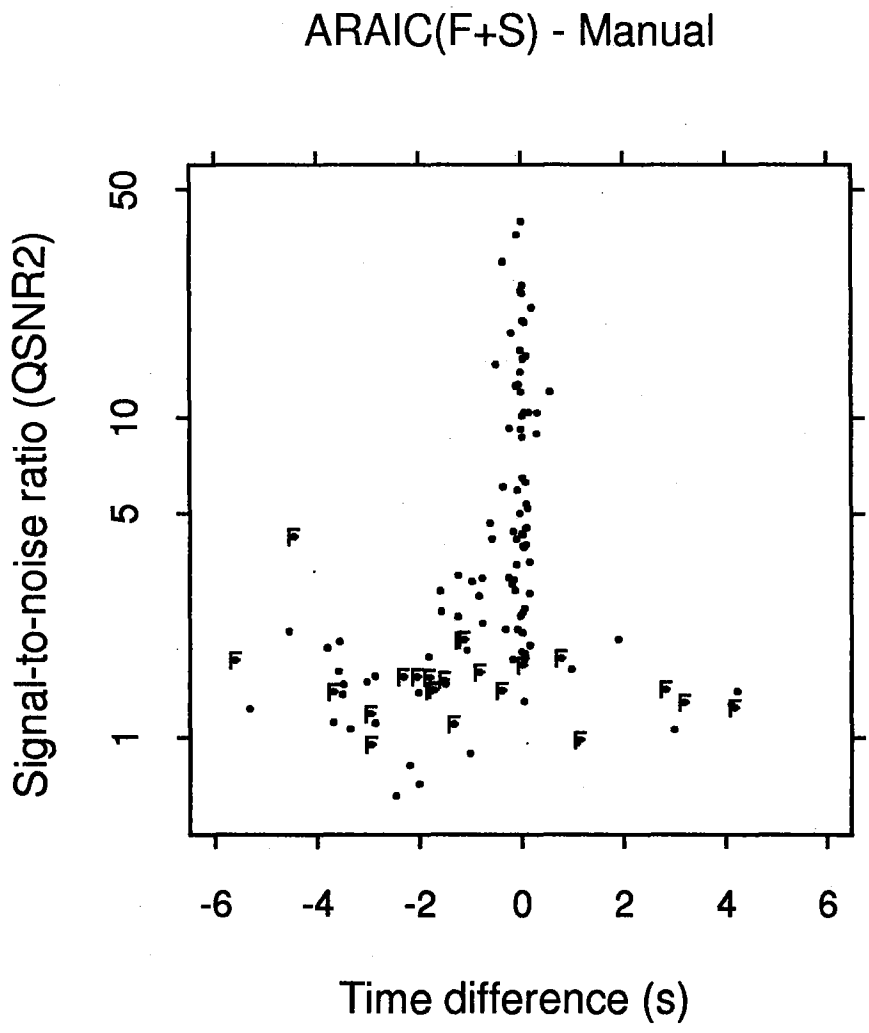

Fig. 7.4.11. Time difference between the AR-AIC$_{F+S}$ onsets and manually picked onsets shown as a function of QSNR$_{2.0}$. The data points labelled F represent phases that we were unable to pick in a confident way, primarily due to low SNR.

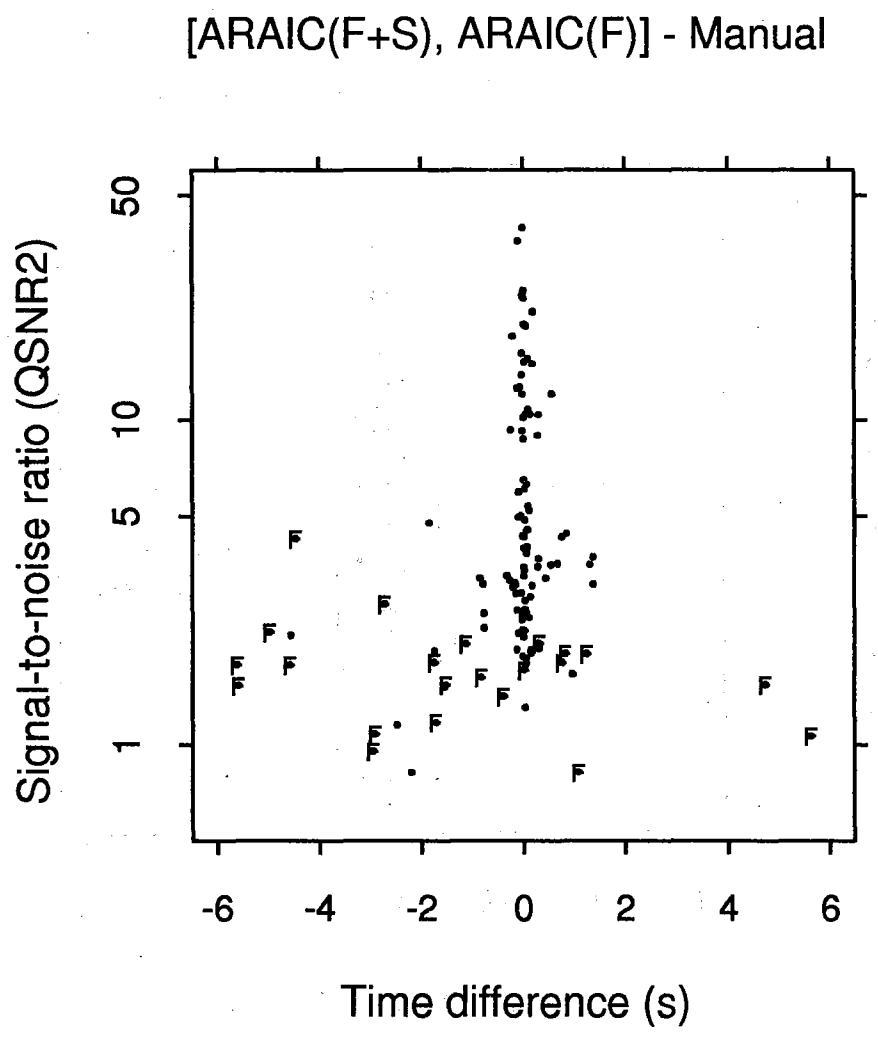

Fig. 7.4.12. Same as Fig. 7.4.11, but based on certain criteria of the quality measurements, the AR-AIC$_F$ onsets were used instead of the AR-AIC$_{F+S}$ onsets. See text for details.

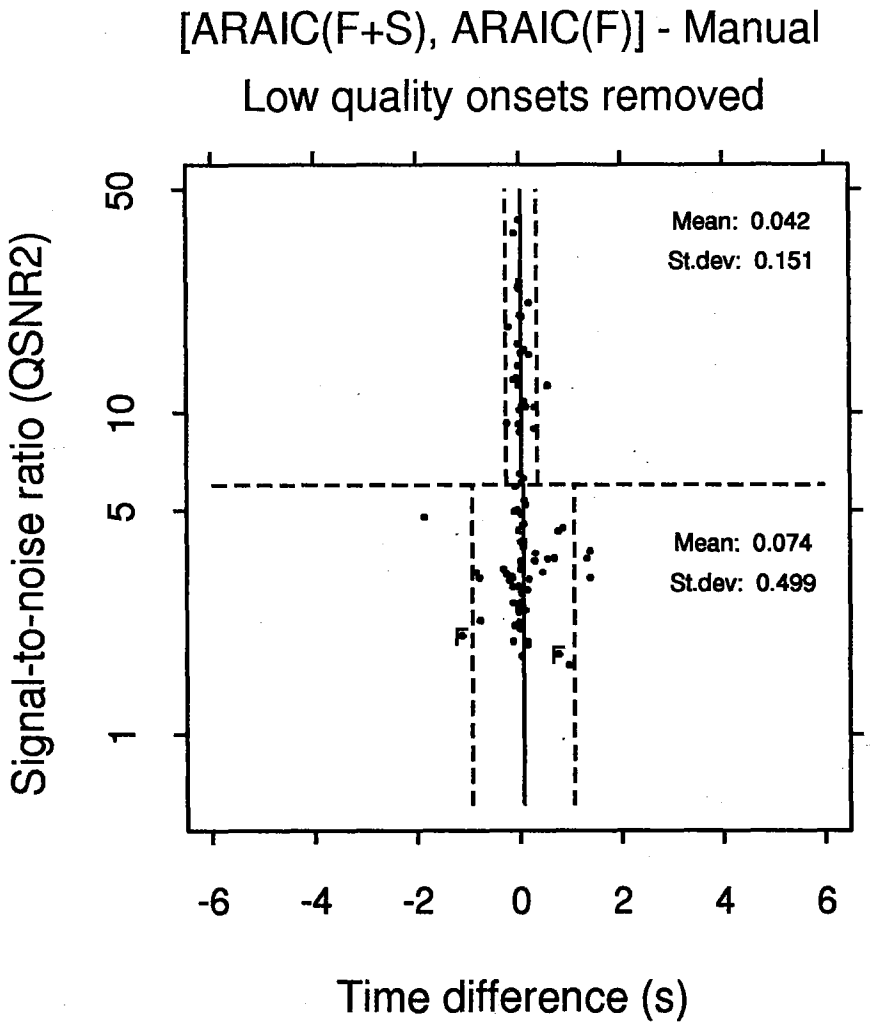

Fig. 7.4.13. Same as Fig. 7.4.12, but with low quality onsets removed. Notice the difference in the scatter between the high and low $QSNR_{2.0}$ populations. The dashed lines indicate a distance of two standard deviations from the mean.

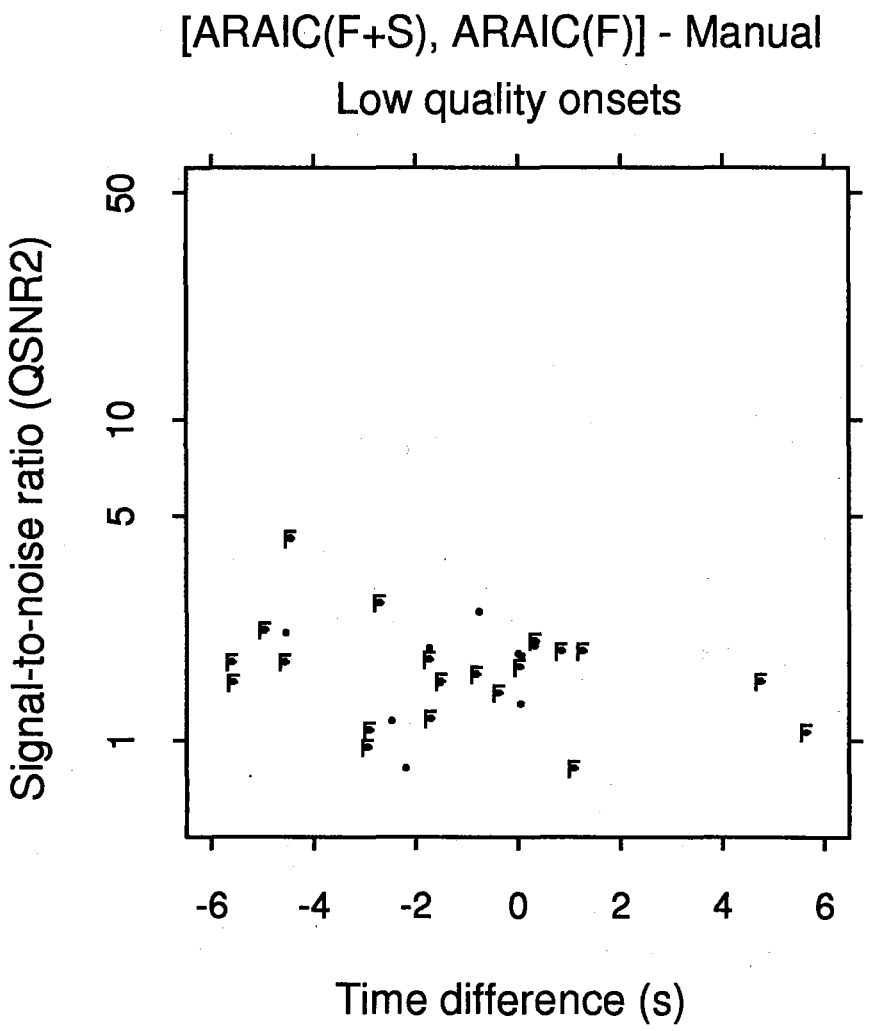

Fig. 7.4.14. This figure shows the data identified as low quality onsets by utilizing the envelope quality measurements