THE CURRENT STATUS OF INFRASOUND DATA PROCESSING AT THE INTERNATIONAL DATA CENTRE

Nicolas Brachet and John Coyne

Provisional Technical Secretariat of the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

ABSTRACT

One component of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) Preparatory Commission is the International Data Centre (IDC), which receives and processes data from a globally distributed network of seismic, hydroacoustic, infrasound, and radionuclide stations. Currently the Operational system at the IDC makes only limited use of infrasound data during routine processing. However, active development and testing is being done in the development environment, as described in this paper. Infrasound data at the IDC are processed at the station level using the Progressive Multi-Channel Correlation (PMCC) algorithm, which calculates the correlation between sensors at an infrasound array station. If the signal is sufficiently correlated and consistent over an extended period of time and frequency range, a detection is created. Groups of detections are then categorized, and detections are assigned the phase name I (infrasound), IPx (seismic P), ISx (seismic S), or N (noise). Currently, around 90% of detections are identified as noise. Non-noise detections are then used in network processing, along with detections from seismic and hydroacoustic technologies. The result is an automatically generated bulletin that includes phases detected at seismic, hydroacoustic, and infrasound stations. This automatic bulletin including all waveform technologies is planned to be routinely reviewed by analysts during the interactive review process. Specialized software is being developed that allows analysts to visualize the PMCC results that were automatically calculated and stored during station processing. These results are presented in plots of azimuth and speed versus time and frequency. This form of display is very powerful, especially for low signal-to-noise ratio (SNR) signals, which may be very difficult to visualize with only waveform data. Procedures for the routine interactive review of infrasound data are currently being explored and developed at the IDC.
INTRODUCTION

The CTBTO is tasked with monitoring compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT), which bans nuclear weapon explosions underground, in the oceans, and in the atmosphere. The verification regime includes a globally distributed network of seismic, hydroacoustic, infrasound, and radionuclide stations that collect and send data to the IDC in Vienna, Austria, shortly after the data are recorded at each station. Upon receipt at the IDC, the time series data from each seismic, hydroacoustic, and infrasound station are automatically processed (Figure 1) at the station level (station processing). The results of station processing serve as input to network level (network processing). Network processing results in automatic event locations, which are reported in bulletins known as Standard Event Lists (SELs). Three SELs are successively made at the IDC: SEL1 includes seismic and hydroacoustic data, and is produced two hours after real time; SEL2 includes seismic, hydroacoustic, and infrasound data and is available six hours after real time; SEL3 also includes seismic, hydroacoustic, and infrasound data and is available twelve hours after real time. Seismic data from auxiliary seismic stations are requested after each SEL and are used to refine event locations in subsequent bulletins. The bulletin production deadlines are staggered to accommodate late-arriving data and the signal propagation times for all technologies.

The SEL3 bulletin is reviewed by human analysts, during which the automatic results are corrected and any late-arriving data not available for SEL3 processing are considered. The result of the interactive review process is the Reviewed Event Bulletin (REB), which is typically available in less than 10 days after real time. The creation of the REB triggers a post-location processing pipeline, which includes processes such as surface wave magnitude estimation and event characterization. Additional bulletins are formed as a result of these processes. After Entry Into Force (EIF) of the CTBT, the delay for producing the REB is planned to be two days.

Figure 1. Schematic overview of data processing at the IDC.

The IDC is actively developing software and procedures that will be used to routinely process infrasound data at the IDC. Currently, in the IDC Operational System, infrasound data are routinely processed at the station level but are not considered during network processing and consequently do not contribute to any SEL bulletin. Analysts may manually associate infrasound data during interactive review, but this is currently done only on an exceptional basis. Consequently, infrasound stations rarely appear in the REB at the current time. The remainder of this paper discusses the current status of infrasound processing in the IDC development environment.

RESEARCH ACCOMPLISHED

IDC Development Environment

Development at the IDC is done in a mixed Solaris and Linux environment, where the full suite of IDC application software is routinely operated 24 x 7. All development and maintenance activities are initially tested and refined in this environment before promotion into the Operational System. The full set of IMS data available in Operations is present in the development environment, as well as stations that are still being tested and have not yet been promoted into Operations. As of June 2006, data from 35 infrasound stations are processed in the IDC development environment (Figure 2). This accounts for 58% of the planned infrasound network of 60 stations, as defined in Annex 1 to the Protocol of the CTBT. Thirty-four of these stations are in the IDC Operational System.
Infrasound Station Processing

Note: Throughout this text, when specific values are given, it is important to remember that many of the values are interrelated and must be considered in the proper context. Changing a value may have unforeseen consequences elsewhere. In addition, active development is being done in many of these areas, and values are subject to change and refinement. The incentive for including these values is to provide a technical insight into the infrasound processing currently being done at the IDC.

Signal Detection

Signals recorded at infrasound stations are detected at the IDC using the PMCC algorithm (Cansi, 1995). PMCC is an array processing technique that detects coherent energy crossing the array. This algorithm initially calculates the correlation of signals between triplets of sensors at an infrasound station. If the signals are correlated and consistent, waveforms from other sensors at the station are progressively considered and are retained if the signals are sufficiently correlated and consistent. As additional sensors are added, the obtained results are refined and improved. Correlation is performed for overlapping time windows and multiple frequency bands. Currently, due to limited processing capacity, processing is done using time windows 50 s long with a 10 s increment between adjacent windows and 10 frequency bands between 0.1 and 4.0 Hz. There are plans to introduce new hardware and the Linux operating system, where processing will be done using variable window lengths and 15 frequency bands between 0.02 and 4.0 Hz.

If waveforms are sufficiently correlated and consistent for a specific time window and frequency band, the results are stored in a so-called pixel, which provides an insight into how the results are subsequently visualized. The pixel height represents the processing bandwidth, while the pixel width represents the time increment between adjacent windows. Each pixel also includes detection attributes such as azimuth, speed (trace velocity), consistency, root mean square (rms) amplitude, and Fisher statistic values (F-stat).

The next stage of the detection processing consists of aggregating pixels with similar attributes into larger groups that are referred to as families. Attributes that factor into the family grouping process include time, frequency, azimuth, and speed. If there is a sufficient number of pixels in a family, the family will give rise to a detection. When a detection is formed, a number of attributes are calculated for the family, including time duration of the family, frequency bandwidth, centre frequency, speed, azimuth, and F-stat.
Figure 3 shows a typical display of PMCC processing results. The top two panels show azimuth and speed, respectively. Individual PMCC pixels can be seen, as well as groups of similar pixels that form a family. The lower panel shows waveform data and includes the detections that were identified as I phases. The phases are positioned at the start of each family.

Figure 3. Visualization of PMCC results, illustrating azimuth and speed pixels, and waveforms (unfiltered) with detected infrasound arrivals.

Detection Categorization

After events are detected, the next step of station processing is detection categorization, followed by phase identification. It is instructive to consider the mission of the CTBTO to understand and appreciate the phase identification process applied to infrasound data at the IDC. The overall mission is to monitor compliance with the CTBT, and this is accomplished by routinely producing, daily, a global bulletin of events that were observed by a global network of seismic, hydroacoustic, and infrasound stations. The reported events are usually impulsive and are observed on multiple stations. If the bulletin included single-station events, the uncertainty for those events would be extremely large, and the number of events in the bulletin would increase substantially.

Given these requirements of a global bulletin of events recorded by multiple stations, there are a number of signals recorded at infrasound stations that are not of interest for CTBT monitoring. Such signals include ocean activities (offshore microbaroms or surf on the coast), thunderstorms, small mine blasts in the near vicinity of the station, and aircraft takeoffs and landings. The phase identification process at the IDC is designed to identify such signals as noise (N phase), so these phases are not considered during network processing. If this were not the strategy, there would be a good chance that such signals would be misassociated with other phases (e.g., with seismic phases) during network processing, and the automatic bulletins would be overwhelmed with these false events, which would need to be reviewed and discarded during the interactive review process.
Detection categorization is accomplished in the following two steps.

1. **Apply categorization on individual detections**

Any detection that is high frequency or that does not have an infrasonic speed is identified as noise. This is accomplished with the following tests:

- $\text{min}_\text{freq} > 1.75 \text{ Hz}$
- $\text{speed} < 290 \text{ m/s}$
- $\text{speed} > 500 \text{ m/s}$

2. **Apply categorization on clusters of detections**

The first step is to build clusters of detections with similar characteristics. Two detections (or families) are clustered together if the two detections are close in time, azimuth, and frequency. This is accomplished with the following tests:

- $\abs{\text{time}_2 - (\text{time}_1 + \text{durations}_1)} \leq 5400 \text{ s}$
- $\abs{\text{az}_2 - \text{az}_1} \leq \max(1., 2 \times \max(\text{std dev az}_1, \text{std dev az}_2))$
- $\abs{\text{cfreq}_2 - \text{cfreq}_1} \leq \max(2 \times \text{std dev freq}_1, 0.5 \times \text{bandwidth}_1)$
- $\abs{\text{cfreq}_2 - \text{cfreq}_1} \leq \max(2 \times \text{std dev freq}_2, 0.5 \times \text{bandwidth}_2)$
- $\min(\text{common\_bandwidth} / \text{bandwidth}_1, \text{common\_bandwidth} / \text{bandwidth}_2) \geq 30\%$

A cluster is identified as noise for long-duration clusters, short-duration clusters, and repetitive clusters. This is accomplished if any of the following criteria are met:

**Criterion 1 (long-duration clusters)**

- Detection is not the first in a cluster,
- Cluster duration is $\geq 3600 \text{ s}$, and
- $(\text{detection\_fstat} - \text{min\_cluster\_fstat}) \leq (3 \times \text{std\_dev\_cluster\_fstat})$

**Criterion 2 (short-duration clusters)**

- Cluster duration $< 60\text{ s}$

**Criterion 3 (repetitive clusters)**

- Detection is not the first in a cluster,
- Cluster is composed of more than 5 detections, and
- $(\text{detection\_fstat} - \text{min\_cluster\_fstat}) \leq (3 \times \text{std\_dev\_cluster\_fstat})$

Note: Non-infra detections (detections with speeds $\geq 2900 \text{ m/s}$) are incompatible with the categorization process, and consequently are not considered at this stage, and are directly put into the phase identification process.
Phase Identification

During the phase identification process, phase names are assigned to detections. Detections that have been identified as noise during the categorization process are named N. Detections with speeds over 5700 m/s are interpreted as P-type signals and are named IPx. Detections with speeds between 2900 m/s and 5700 m/s are named ISx. All other detections are interpreted as infrasound arrivals and are named I.

Figure 4 shows two views of the IDC detection list for the station I04AU for the time period 22 May 2006 through 28 May 2006. Each panel of data contains the detections for one day. The display shows the azimuth for the detections, and the azimuth scale is at the bottom of each side. The plots on the left-hand side show all PMCC detections for this time period, including many long-duration signals caused by surf noise and microbaroms. The plots on the right-hand side show the non-noise detections after the detection characterization and phase identification process. The numbers between the plots show the number of detections for each day before and after this process.

Figure 4. Example of IDC detection list before and after detection characterization and phase identification. Detections which were identified as noise are removed from the display on the right-hand side.

Network Processing

During network processing, non-noise phases are used as candidates for forming events that appear in the automatic bulletins. The location algorithm used at the IDC is an iterative non-linear least-squares inversion originally developed by Jordan and Sverdrup (1981) which was later modified by Bratt and Bache (1988) to include azimuth and slowness observations. An event with only infrasound observations will appear in an automatic IDC bulletin if the event is seen by at least two infrasound stations. At the current time, infrasound phases can be associated with events that are up to 60 degrees from a station. During network processing, the time, azimuth, and slowness attributes from arrivals are considered for different hypothetical events. If a candidate event is found, it is saved and will appear in the automatic SEL2 and SEL3 bulletins. Network processing in the development environment currently builds mixed technology events using seismic, hydroacoustic and infrasound phases.
The automatic association algorithm uses the same infrasound travel time model that was originally introduced by the Prototype International Data Center (PIDC) in Arlington, USA. This velocity model basically uses a constant slowness of 350s/deg (i.e., celerity of 318m/s) regardless of the range between source and receiver. Some work is currently underway to refine this propagation model using a dataset of reference events collected by the IDC during the last 5 years.

Large efforts have been produced to decrease the number of false associations with infrasound arrivals. The introduction of the detection categorization strategy (previously described) has contributed to significantly lowering the number of automatic events with infrasound down to a manageable level. The experimental SEL3 bulletin currently produces about 60 events per day, including arrivals from one or several of the 34 infrasound stations. Currently, infrasounds contribute to about 35% to the total events formed with seismic and hydroacoustic technologies. These results are encouraging, but the proportion of false infrasound associations is still too high and unacceptable for an analyst’s interactive review in Operations. Work is currently underway to explore additional techniques to limit infrasound associations based on a distance/frequency relationship. This process may be further refined by considering other attributes such as duration and amplitude.

**Interactive Review**

The primary tool used by analysts at the IDC to routinely review waveform data is the Analyst Review Station (ARS). ARS allows an analyst to review each SEL3 event and to make corrections and additions as needed. The majority of the ARS interface is occupied with waveform data. During seismic data review, FK results are visualized using another tool. This set up of ARS and the FK tool has worked well for reviewing seismic data.

After data are processed using PMCC, it is critical to visualize the PMCC pixels in time and frequency along with the waveforms, in order to correctly comprehend and interpret the results. This is especially true for low SNR arrivals, which may be very difficult to visualize with only waveform data. In order to accommodate this need, a software tool has been developed that allows the PMCC pixels that were calculated during station processing to be visualized during the interactive review process.

This tool has a number of innovative features, including the following:

- Plots of azimuth and speed pixel information as a function of frequency and time, as well as the original time series data, phases identified during station processing, and meteorological observations at the station.
- Ability to toggle the pixel display between all pixels, only pixels that are members of a family, or only pixels that are members of a non-noise family.
- Polar diagram plot that is integrated with the main window and displays pixel information in polar coordinates of azimuth and speed. Figure 5 shows the waveforms recorded at I07AU when a fast-moving object passed over the array. The polar diagram shows the azimuth and change of speed for this fast-moving object.
- Integration with ARS.

Analysts typically use ARS to review all waveform data from an event. In ARS, a user may select a phase recorded at an infrasound station and press a button to send a message to the PMCC tool, which displays the PMCC results surrounding the arrival. Once those results are displayed, a user may retime, rename, or delete the phase, where each action is synchronized with ARS.

The tool can also be used independently of ARS to directly read and display PMCC results of station for a specific time interval. In all cases the tool reads data from the relational database management system, including the pixel and arrival information that was calculated and stored during station processing.

Work is currently underway at the IDC to develop procedures for the routine interactive review of infrasound data. Currently, in the Operational System, interactive review is done by an analyst who selects a 4-hour time block and analyzes all events and waveforms in that time period. Events that are toward the beginning of the time block sometimes require data from the previous time block to be reviewed.
The process of analyzing infrasound data may require a different procedure for a number of reasons, including the following:

- Different types of display needed to visualize and comprehend the results of station processing.
- Much slower wave propagation times so that associated infrasound arrivals are often outside the time block being reviewed.
- A significant number of additional stations, which could have a severe impact on ARS performance.
- The infrasound channels to display for each station in ARS is still an open issue. For seismic arrays, a beam is typically displayed in ARS; however, beams are not currently made during infrasound station processing.

Given these reasons, different approaches for reviewing infrasound data are being considered and are in various stages of implementation and testing. In one approach, infrasound data are treated the same as other waveform data, and selected channels are displayed for a time block in ARS. Data outside the time block are displayed as needed. In the second approach, infrasound waveforms are not displayed in ARS. When an event is reviewed that contains infrasound arrivals, the arrival is selected in the location dialog window of ARS and is sent to the PMCC review tool which displays the data. The results can be modified in the tool and sent back to ARS. The third approach is to do a “first pass” review of infrasound data using only the PMCC review tool in order to validate or discard the infrasound arrivals contributing to SEL3 events. After this step, arrivals and events that have a reasonable probability of being genuine are reviewed in the context ARS.

Figure 5. Example of PMCC pixel display and polar diagram, showing a fast-moving aircraft traveling from the southeast to the northwest across the array. The units in the polar display are azimuth (deg) and speed (m/s).
CONCLUSION(S) AND RECOMMENDATIONS

This paper describes the current status of infrasound processing in the development environment at the IDC. The PMCC algorithm is being used to routinely detect signals observed at infrasound stations. Detections with similar characteristics are categorized into clusters, which facilitates the assignment of phase names to similar detections. Infrasound signals are then assigned the phase names N, I, IPx, or ISx. Around 90% of infrasound detections are currently identified as noise. Non-noise arrivals are considered during network processing, which considers arrivals from all waveform technologies. Currently, about 35% of the SEL3 events in the development environment include infrasound arrivals. A specialized tool has been developed to interactively review the results of infrasound station processing, and procedures are currently being developed for the routine interactive review of infrasound data.

DISCLAIMER

The views expressed herein are those of the authors and do not necessarily reflect the views of the CTBTO Preparatory Commission.

ACKNOWLEDGEMENTS

We would like to thank the members of the infrasound specialist group at the IDC, who have provided valuable feedback concerning the interactive review software. The members of this group are Abdou Ndiath, Misrak Fisseha, Feyza Ocal, Kirill Sitnikov, and Gadi Turyomurugyendo.

REFERENCES

