



Site Effects Assessment Using Ambient Excitations

SESAME

**European Commission – Research General Directorate
Project No. EVG1-CT-2000-00026 SESAME**

**Report on the
Interface software**

University of Potsdam, Germany

WP06

**Array measurements:
Derivation of dispersion curves**

Deliverable D15.06

June 2003

List of Contents

Summary.....	2
Waveform interface software tool 'retrieve' - Description and Usage	2
A simple Example.....	8
Acknowledgements	6
References	6

List of Contributors

Matthias Ohrnberger, IGUP, Potsdam, Germany

Project coordinator: Pierre-Yves Bard

Task B Leader, WP06 Leader: Frank Scherbaum

Summary

The software tool '**retrieve**' is a simple command line tool, allowing to extract single or three-component waveform data from an existing GIANT database for a given time window and station. The software tool intends to facilitate the interfacing of waveform data between different programs used within the SESAME project, i.e. to export the data acquired within Task B (array measurements of ambient noise) in the SESAME Ascii Format (SAF) for further use with the H/V processing software developed in Task A. In order to account for the instrument characteristics of the recorded data, the existing calibration information is exported automatically in a 'poles and zeroes' representation (PAZ) in the calibration file format defined in GSE1.0 (CAL1 section, PAZ format).

'retrieve' is entirely written in C and makes use of the RAIMA data base manager library for accessing the database structures of a GIANT database. The software is distributed as binary executables for Linux Intel platforms (Kernel 2.x) as well as for SPARC SUN Solaris systems (2.5.x and higher). Due to the licensing conditions of the developer's license of the RAIMA database manager, the source code can not be provided.

Waveform interface software tool retrieve -Description and Usage

The software tool 'retrieve' shall facilitate the access to waveform data acquired during field campaigns within SESAME Task B. Its main feature is the conversion from data formats used with the GIANT database (in our case GSE2.0) to the SESAME Ascii format which is used within the workgroups of SESAME Task A (H/V processing). Similar to the software tools HVDB and cap (see deliverable D05.06 - quality control software), retrieve is used from the command line. The user has to specify the GIANT database from which he/she wants to export waveform data by setting several environment variables (compare also GIANT manual, Appendix 1 of D05.06) **before** calling 'retrieve'.

Example (for csh):

```
user@host:~> setenv IATSN_BASE lefkas
user@host:~> setenv IATSN_CAL /home/user/lefkas/cal/
user@host:~> setenv DBDPATH /home/user/lefkas/base
user@host:~> setenv DBFPATH /home/user/lefkas/base
user@host:~> setenv IATSN_DATA /home/user/lefkas/
```

Example (for sh or bash):

```
user@host:~> export IATSN_BASE=lefkas
```

```

user@host:~> export IATSN_CAL=/home/user/lefkas/cal/
user@host:~> export DBDPATH=/home/user/lefkas/base
user@host:~> export DBFPATH=/home/user/lefkas/base
user@host:~> export IATSN_DATA=/home/user/lefkas/

```

The environment variable IATSN_BASE is the name of the database to be selected. IATSN_CAL points to the directory of the calibration files connected to the database. The '/' at the end of the path is required. IATSN_DATA points to the directory from which relative path and waveform filename entries have been stored in the GIANT database (usually the root directory of a database location). As before, the '/' at the end of the given path is mandatory. Environment variables DBDPATH and DBFPATH point to the directory where the database files are located.

Note 1: For unix or unix-like systems upper and lower case letters are distinguished. The environment variables must be specified therefore in capital letters as shown above in the example.

Note 2: For all of the databases provided in the deliverable D05.06, we have used a common directory structure. Although it is not mandatory to use this directory structure for organizing database files, calibration information, and waveform files, we recommend to keep it 'as is'. For the following description, we will use implicitly the pre-configured directory structure for convenience.

After the user has selected the GIANT database from which he/she wants to export waveform data, the command line tool 'retrieve' is called with the following syntax:

```

user@host:~> retrieve -f YYYY[MM[DD[hh[mm[ss.sss]]]]] -l
YYYY[MM[DD[hh[mm[ss.sss]]]]] -g GP01+GP02 -o path-to-output [-c component] [-s] [-b]
[-z] [-i integer] [-e]

```

Necessary command line parameters are the -f, -l, -g, -o and -t options, followed by an appropriate argument. The -c, -b, -z, -i, -s, and -e flags are optional. The optional command line flags -c and -i require an argument, whereas -s, -e, -b and -z must not have an argument. Calling 'retrieve' without any argument or in case that an error occurred during the parsing of the command line, the following help message is displayed:

```

=====
USAGE: retrieve -optflag optarg -optflag optarg ...

```

option flags:

- f: <arg: window start - time string: 19970720>
- l: <arg: window end - time string: 1997072001>
- time strings are accepted as:
 - YYYY (just year)
 - YYYYMM (year and month)
 - YYYYMMDD (year, month and day)
 - YYYYMMDDHH (...)
 - YYYYMMDDHHMM (...)
 - YYYYMMDDHHMMSS (...)
 - YYYYMMDDHHMMSS.SSS (...)
- g: <arg: station (or station list)>
- Example: P012 or P012+WLMN+NN99
- o: <arg: output file path>
- c: <arg: component>
- 1: selects only Z-components

2: selects only N-components
3: selects only E-components
99: selects all 3 components
-t: <arg: output file type>
A: PITSa ASCII
G: GSE2.0
S: SESAME ASCII format

Processing parameters:

-s: <noarg> seidl's restitution
-b: <noarg> butterworth bandpass
-z: <noarg> zero phase bbp
-i: <arg: decimation factor>
-e: <noarg> extra time correction

=====

- The obligatory command line flags **-f** and **-l** specify the time window of waveform data which shall be exported from the database by giving the start time (-f option - first time) and the end time (-l option - last time). The time format consists of a variable length time string. The following strings are accepted for the -f and -l flags:

YYYY => full year specification. This time argument is interpreted as 00 h 00' and 00" of the 1st of January for the given year.

YYYYMM => year and month specification. The time argument is interpreted as the first day of the given month and year, at 00 h 00' and 00".

YYYYMMDD => year, month and day specification. The time is interpreted as 00 h 00' 00" on the given day, month and year.

YYYYMMDDHH => year, month, day and hours specification. The time is interpreted as starting (ending) at the full hour.

YYYYMMDDHHMM => year, month, day, hours and minutes specification. The time is interpreted as starting (ending) at the full minute.

YYYYMMDDHHMMSS => year, month, day, hours, minutes and seconds specification. The time is interpreted as starting (ending) at the full second.

YYYYMMDDHHMMSS.SSS => full time specification. The time is interpreted as starting (ending) time of the sample closet to the specified times.

- By making use of the mandatory **-g** option, the user specifies the station (or station list) for which the waveform data shall be extracted. The expected string argument is either a single station name or a list of station names concatenated with '+' signs.
- The **-o** option is used to select the (absolute or relative) output directory, into which the exported waveform data will be written.
- The user can select different output file formats with the **-t** option. Three output waveform formats are currently supported: Pitsa Ascii Format, GSE2.0 format (compressed CM6) and the SESAME Ascii Format (SAF). The argument to the -t flag is a single character, either 'A', 'G', or 'S', specifying either Pitsa Ascii, GSE2.0 or SAF to be written.

Note 3: Depending on the selected output format, 'retrieve' behaves somehow differently. a) only for the SAF output file type, a single data file is written for three-component data. In case of Pitsa Ascii or GSE2.0 format, each user-specified channel will be written into its own data file. The naming convention used for the output data files in case of Pitsa Ascii or GSE2.0 format is: "YYYYMMDDHHMMSS_C.STAT", where the

first string indicates the start time of the record, C stands for the component (given as integer: 0 == Z, 1 == N, 2 == E) and STAT is the station name as given in the database. For SAF files, the filename convention specified in the Meeting Minutes of WP03 (Nice, April 2002) is used: "NNNNNN_YYYYMMDD_HHMM.saf". The first six characters are filled with the station name. In case the station name is shorter than 6 characters, we fill the string with '#' until it reaches the desired length. The date and time format indicates the start time of the record and the extension is set to ".saf". No component information must be given, as SAF files always contain all three components stored in three columns.

b) in case of writing SAF waveform files, the calibration information connected to each station and channel is automatically extracted from the database and is written to the specified output directory. The file name of this calibration file is additionally recorded in the header of the corresponding ".saf" file in order to facilitate the linkage between the files. The naming convention used follows the specifications given in the Meeting minutes of the WP03 workshop, held in Nice in April 2002. File names are created like: "NNNNNN_YYYYMMDD_HHMM_C.cal". The base of the file name is easily identified with the corresponding waveform file, then the component information is appended (as before specified as integer) and the extension is set to ".cal".

- The -c flag can be used to specify the components which shall be extracted from the given database. The argument is an integer value: 1 (selects Z-component), 2 (selects first horizontal component - usually N), 3 (selects second horizontal component) and 99 (selects all 3 components). The -c flag is optional, if not specified, the default is set to 99 (all 3 components will be written).

Note 4: The SAF requires a 3-column output. Thus in order to not conflict with other reading routines, we always write three columns to the output file, whether all three components have been selected or not. In case of selecting a single component, the other two data columns are filled with 0 values.

- The -i option can be used in case, the user wants to export waveform data with a reduced sampling rate. The algorithm used is a simple integer decimation after applying a digital anti-alias filter with a cut-off frequency set to 40% of the new sampling rate ($0.4 \cdot f_c$). The argument passed to the -i flag is a positive integer value.
- The flags -b and -z are used to bandpass the data before exporting the waveforms. Both flags take no argument. -b specifies, that a Butterworth bandpass filter shall be applied to the data. The -z flag specifies whether the filtering is performed only in the forward direction (if -z is not set) or both in forward and backward direction (if -z is set = zerophase filtering). Specifying the -z flag without using the -b option has no effect. The filter parameters, lower and upper corner frequencies and the number of conjugate complex pole pairs ("number of sections") for approximating the filter response, will be asked from the user interactively - a manual input is required.
- The -e option is primarily intended for the use with further array processing applications. Using the -e flag will ask for each station-channel combination for a know time correction for this trace. This option has been valuable for correction purposes when timing problems have occurred during recording.
- Using the -s option allows to simulate an instrument response using a method described by Seidl (1980). The original instrument response and the desired new instrument response are combined to design a filter operation which is then applied to the seismic waveform data. The parameters of the desired instrument response (corner frequency f_c and percentage of critical damping h) will be given interactively by the user after program

start.

Note 5: Using the SAF output option, the data unit will be changed from μV (original unit of waveform data) to nm/s. Thus the conversion factor for calculating amplitudes from the integer values used in the format will be given in nm/s/C instead of $\mu V/C$.

Calling 'retrieve' will start the following action. First, the command line is parsed and checked for missing input or lack of consistency. Then a connection to the database is established and a query for waveform data within the specified time window and for the given stations is performed. Consistency checks are performed to assure, that no change in sampling rate or data gaps occur when waveform files have to be concatenated. In case that further input is required (options -s, -e, and -b) the user is asked for the parameters interactively. For all traces, the overall offset is removed. Any further processing (if -s, -b, or -i have been given) will then be performed before writing the output files in the desired format. Using the SAF output option, additionally calibration files are written. If the -s option has not been used, the original calibration files are copied. In case of specifying the -s flag, the calibration information is re-calculated according to the choice of parameters which have been given from the user.

A simple Example

For the purpose of illustration, here is a simple example of a test run from 'retrieve' applied to the SESAME data set 'lefkas'. Here, we extract 30 min waveform data of two stations in SAF format:

```
user@host: > retrieve -f 20020803100000.000 -l 20020803103000.000 -g GP07+GP06 -
c 99 -o ./ -t S
```

```
This is the correct calib file!
Searching station GP06 - channel 1 *Done
Searching station GP06 - channel 2 *Done
Searching station GP06 - channel 3 *Done
Searching station GP07 - channel 1 *Done
Searching station GP07 - channel 2 *Done
Searching station GP07 - channel 3 *Done
Searching station GP06 - channel 1 *Done!
Searching station GP06 - channel 2 *Done!
Searching station GP06 - channel 3 *Done!
Searching station GP07 - channel 1 *Done!
Searching station GP07 - channel 2 *Done!
Searching station GP07 - channel 3 *Done!
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_0.P06
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_0.P06
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_1.P06
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_1.P06
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_2.P06
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_2.P06
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_0.P07
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_0.P07
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_1.P07
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_1.P07
opening file /home/clust/SESAME/GREECE/lefkas/20020803/100000_2.P07
Filename: /home/clust/SESAME/GREECE/lefkas/20020803/100000_2.P07
We removed offset -0.733690 for station 1
```

```

We removed offset 3.692686 for station 2
We removed offset -3.568281 for station 3
We removed offset -4.730209 for station 4
We removed offset -1.173067 for station 5
We removed offset -2.074188 for station 6
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
cfreq: 0.199550 and damping 0.705402 read from calfile, gain 0.400000
This is a GSE1 calibration file
This is a GSE1 calibration file
Executing command: cp ./cal/0006_006.n ./GP06##_20020803_1000_1.cal
This is a GSE1 calibration file
Executing command: cp ./cal/0006_006.e ./GP06##_20020803_1000_2.cal
This is a GSE1 calibration file
This is a GSE1 calibration file
Executing command: cp ./cal/0007_007.n ./GP07##_20020803_1000_1.cal
This is a GSE1 calibration file
Executing command: cp ./cal/0007_007.e ./GP07##_20020803_1000_2.cal

```

The messages written to standard output in the terminal are lengthy on purpose. This allows the user to easier recognize processing errors and will hopefully also allow to send precise bug-reports to the author (please report any bugs of this software tool to mao@geo.uni-potsdam.de).

In the example the output path was set to the current directory (.). These are the files that have been created by running 'retrieve':

```
user@host: > ls -lrt
```

```
...
```

```

-rw-r--r-- 1 mao  research  156 Jul 6 14:55 GP06##_20020803_1000_2.cal
-rw-r--r-- 1 mao  research  156 Jul 6 14:55 GP06##_20020803_1000_1.cal
-rw-r--r-- 1 mao  research  156 Jul 6 14:55 GP06##_20020803_1000_0.cal
-rw-r--r-- 1 mao  research  156 Jul 6 14:56 GP07##_20020803_1000_2.cal
-rw-r--r-- 1 mao  research  156 Jul 6 14:56 GP07##_20020803_1000_1.cal
-rw-r--r-- 1 mao  research  156 Jul 6 14:56 GP07##_20020803_1000_0.cal
-rw-r--r-- 1 mao  research 6233852 Jul 6 14:56 GP07##_20020803_1000.saf
-rw-r--r-- 1 mao  research 6384532 Jul 6 14:56 GP06##_20020803_1000.saf

```

For each station a waveform data files has been written, as well as one calibration file for each component of each station. This is what the header of one of the waveform data looks like:

```
user@host: > head -20 GP07##_20020803_1000.saf
```

```

SESAME ASCII data format (saf) v. 1
# REMARK: file has been extracted from a GIANT database with utility 'retrieve'
# Command line: /ldat1/mao/sesame_test/get_src/retrieve -f 20020803100000.000 -l 20020803103000.000 -g
GP07+GP06 -c 99 -t S
# Preprocessing parameters used:
# Sampling frequency in Hertz
SAMP_FREQ = 124.999994
# number of samples (long)
NDAT = 224999
# start time of record
START_TIME = 2002 08 03 10 00 0.000
# response file name
RESPFILE = GP07##_20020803_1000_0.cal

```

```
SENS_EIGFREQ = 0.199550
SENS_DAMPING = 0.705402
# seismometer generator constant [uV/nm/s]
TRANS_FACTOR = 0.400000
# RESPFILE H1 = GP07##_20020803_1000_1.cal
# SENS_EIGFREQ H1 = 0.199550
# SENS_DAMPING H1 = 0.705402
# RESPFILE H2 = GP07##_20020803_1000_2.cal
# SENS_EIGFREQ H2 = 0.199550
# SENS_DAMPING H2 = 0.705402
# conversion factor [V/counts]
CONV_FACTOR = 2.42849e-11
#####
-38365316 -449396 -1081995
-47839652 -11525318 6815334
-29341554 -13397247 9860803
-8385779 -4230119 5494333
3685502 3257597 4097555
```

Note: only the calibration information of the vertical component has been inserted here without a '#' (indicating a comment). We have chosen to do so, in order to not 'confuse' reading routines for SAF waveform files.

Acknowledgements

This project (Project No. EVG1-CT-2000-00026 SESAME) is supported by the European Commission – Research General Directorate.

References

Seidl, D., The Simulation Problem for Broad-Band Seismograms, *Journal of Geophysics*, Vol. 48, pp. 84-93, 1980.