

ActSds and OdfSds: Programs for converting INTERACT and The Observer data files into SDIS timed-event sequential data files

ROGER BAKEMAN

Georgia State University, Atlanta, Georgia

AND

VICENÇ QUERA

Universidad de Barcelona, Barcelona, Spain

In this article, we describe programs for converting Mangold International's INTERACT and Noldus Information Technology's The Observer data files to sequential data interchange standard (SDIS) timed-event sequential data files. Users who convert their INTERACT or The Observer data files can then take advantage of various flexible and powerful data modification and computational procedures available in the Generalized Sequential Querier, a program that assumes SDIS-formatted files.

Mangold International's INTERACT (www.mangold.de; Mangold, 2006) and Noldus Information Technology's The Observer (www.noldus.com; Noldus, 1991; Noldus, Trienes, Hendriksen, Jansen, & Jansen, 2000; Noldus Information Technology, 2003) are sophisticated software packages for the collection and analysis of observational data. Typically, coders work with multimedia digital files (other possibilities include video tapes or live coding); the results of their coding with INTERACT are stored in tab-delimited text files with the extension ACT, and the results with The Observer are stored in text files with the extension ODF (observer data file). The sequential data interchange standard (SDIS) is a standard format for sequential data (Bakeman & Quera, 1992); files using this format have the extension SDS. The programs described in the present article—ActSds and OdfSds—convert ACT or ODF into SDS files. In particular, OdfSds is a more general version of an earlier ODF to SDS conversion program (Bakeman & Quera, 2000) that required Version 3 Observer data files, whereas OdfSds should work with files produced by various versions of The Observer.

Conversion is useful for investigators who use either INTERACT or The Observer to code behavior but who then wish to perform at least some of their analyses with the Generalized Sequential Querier (GSEQ)—a program that analyzes SDIS-formatted data (Bakeman & Quera, 1995). The Observer and INTERACT were designed initially to facilitate data collection. In contrast, GSEQ has no data collection capability but was designed to facilitate more sophisticated data analyses in conjunction with standard statistical packages. Thus, the conversion programs

described in the present article are useful for investigators who wish to take advantage of the extensive data modification and analytic capabilities of GSEQ. For example, users can define new codes that occur only at those times when two or more earlier behaviors were all coded (a logical AND), or at times when any one of a set of behaviors was coded (a logical OR); other logical operations are also available. Furthermore, users can define new codes that are tied to the onset, offset, or occurrence times of existing codes. For example, a new code might include the 10 sec before or the 5 sec after the onset of an alarm cry, or a stretch of time that includes seconds coded for baby crying and 5 sec after. GSEQ allows for lag sequential analysis to any of a number of positive or negative lags and also includes a procedure that converts time-based to event-based data. Another useful feature in GSEQ—especially when observers are first being trained—lets the user define codes within a specified time window as an agreement and then computes the κ agreement statistic for two observers using this flexible window.

INTERACT is commercially available from Mangold International; The Observer is commercially available from Noldus Information Technology, and GSEQ for Windows (GSW), ActSds, and OdfSds can be downloaded from the authors' Web pages (see the Details and Availability section).

Events, Time Formats, and SDIS Data Types

Basic to observational research is the event. Coders are asked to detect and identify (i.e., code) events as they occur in the stream of behavior. Programs like INTERACT and The Observer are useful because they permit users to view

R. Bakeman, bakeman@gsu.edu

and review video records and identify events of interest; the programs then automatically record when those events occurred. Both INTERACT and The Observer store information about coded events in data files; both also have ancillary files that contain information about the codes used and other information. ActSds and OdfSds process only the data files; hence, these programs should work with data files produced by different versions of INTERACT and The Observer—even when the format of the ancillary files changes—as long as the basic data file format remains constant.

INTERACT records times as hh:mm:ss:ff, where hh is hours, mm is minutes, ss is seconds, and ff is frames; ActSds lets the user specify whether there are 30 frames per second (the US standard; NTSC), or 25 frames per second (the European standard; PAL). The Observer records time in decimal format (e.g., sss.dd), so the frames per second is not an issue. Both ActSds and OdfSds convert all times to a decimal format and let the user specify whether to round to the nearest 100th or 10th or nearest second. Note, however, that rounding to the 100th cannot make video more precise than the 25th or 30th of a second used for recording. By default, ActSds and OdfSds round to the nearest second, recognizing that human observers are involved, that coded behaviors usually last a few seconds at least, and that—as Bakeman and Quera (1995) argued—it does not make sense to use more precision than is reasonable or required.

SDIS permits several time formats. ActSds and OdfSds convert INTERACT and The Observer data files into SDIS timed-event sequential data (TSD) files. Data in TSD files consist of codes along with their associated onset and (explicit or implicit) offset times. There are other SDIS data types (Bakeman & Quera, 1995), but essentially these are notational conveniences; all types are compiled by GSEQ into files with a single binary format, which GSEQ then uses for analysis. When codes are organized into mutually exclusive and exhaustive (ME&E) sets of codes, coders do not need to enter offset times explicitly, because the onset of another code in the set implies the offset for the previous code in that set. Both INTERACT and The Observer let users define sets of ME&E codes in an ancillary file. INTERACT then enters offset times for codes in such sets explicitly in the data file. Thus, ActSds produces entries of the SDIS form “Code,OnsetTime-OffsetTime.” In contrast, The Observer does not enter offset times explicitly in the data file. Thus, OdfSds produces entries of the SDIS form “Code, +OnsetTime” (termed a *context code*); offset times are determined by the GSEQ compiler from information concerning sets of ME&E codes given in the SDIS declaration, as will be described later.

Subjects and Sessions

Observational researchers—even as they lavish time on careful coding of behavior—have in mind eventual data analysis, which will be shaped by the study’s design. Thus, it is important at the outset to specify the basic sampling units (individual participants, parent–child dyads, families, or other groups—cases in SPSS, or subjects in older literature) and the research factors that constitute the design. Traditionally, such factors are described as between subjects (e.g., gender with two levels, male and female) or within

subjects (e.g., subjects observed more than once—e.g., at 12, 18, and 24 months of age, or in different contexts). When repeated measures exist, *analytic units*, each identified with a repeated measure, are nested within sampling units. Hereafter, when we use the term *unit*, we mean an analytic unit, either one per sampling unit for between-subjects or more than one for within-subjects designs.

GSEQ lets the user compute various scores from an observational session. These scores are then organized by the research factors for subsequent analysis. In other words, we are defining *session* as synonymous with *analytic unit* (Bakeman & Quera, 1995, defined a session as a sequence of coded events for which continuity can be assumed—i.e., an uninterrupted stretch of time that is coded—but in the present article, we recognize that a session might be segmented by either planned or unplanned breaks; there will be more on this later). Statistics and indices derived from the coded data for an observational session constitute scores; scores from several sessions (analytic units) are then organized by the between- and within-subjects factors of the design and analyzed using conventional statistical techniques. (Howe, Dagne, & Brown [2005] used the term *episode* instead of session; we prefer to reserve episode for stretches of behavior within sessions—e.g., a growling episode made up of individual growls.)

Each ODF file includes data for a single unit, whereas each ACT file can contain data for one or more units (termed *sets*, or, in an earlier version of INTERACT, *takes*). In contrast, an SDS file contains data for all units included in a particular study, which is why ActSds and OdfSds let the user select several files for conversion to a single SDS file. In SDS files, the unit label (i.e., name or identifier)—if present—is enclosed in angle brackets and precedes information about that unit’s events. ActSds uses the set label if it is present in the ACT file, and OdfSds defaults the label to “<Unit#1>,” “<Unit#2>,” and so on. However, before conversion, both programs let the user edit these unit (i.e., session) labels as desired.

SDIS permits a session to consist of more than one segment, separating each with a semicolon; thus, the general form of a session in an SDIS file is

<identifier> sequential codes and
times [; sequential codes and times]/

where square brackets enclose optional material, and the slash indicates the end of data for that unit (see Figure 1). Any breaks in continuity might be planned (e.g., one recording per day for a week for one child) or unplanned (e.g., an unexpected event intrudes during data collection). INTERACT does not have notation for an interrupted recording within a set; hence, each unit produced by ActSds consists of one uninterrupted session. Typically, an ODF file comprises data for one session, indicated by a line that contains {start}, subsequent lines with times and codes, and a final line that contains the offset time followed by {end}. However, a second {start} may follow a previous {end} (in some versions of The Observer), or there may be a {susp} followed by a {resu} (for suspend and resume). If an ODF file contains more than one session segment, they are separated in the SDS file with semicolons.

INTERACT ACT file:	
<pre>*****
SYSTEM: Mom-Child System2

COLUMN: Mother
COLUMN: Child
 Type Entry Exit Memo Mother Child S 00:00:00:00 00:00:00:00 Child Thomas T 21:05:39:00 21:06:18:00 Play session E 21:05:39:00 21:05:44:00 E 21:05:40:00 21:05:45:00 smiling looking at toy E 21:05:49:00 21:05:55:00 looking at himself E 21:06:01:00 21:06:03:00 looking at mother E 21:05:54:00 21:06:04:00 talking E 21:06:04:00 21:06:05:00 eye contact E 21:06:13:00 21:06:17:00 looking at himself E 21:06:15:00 21:06:18:00 smiling</pre>	
Converted SDIS file:	
<pre>% SDS file saved by ActSds Version 1.2 30/11/2007 17:59:38 Timed <seconds> Child_eye_contact Child_looking_at_himself Child_looking_at_mother Child_looking_at_toy Mother_smiling Mother_talking ; <Play_session> ,1265:39 Mother_smiling,1265:40-1265:45 Mother_talking,1265:54-1266:04 Mother_smiling,1266:15-1266:18 & Child_looking_at_toy,1265:39-1265:44 Child_looking_at_himself,1265:49-1265:55 Child_looking_at_mother,1266:01-1266:03 Child_eye_contact,1266:04-1266:05 Child_looking_at_himself,1266:13-1266:17 ,1266:18 /</pre>	
The Observer ODF file:	Converted SDIS file:
<pre>NEWTASK2.CNF Justin1 06-27-1997 00:00:00.00 {indvar} {start} 0.0 Occupied 13.4 Gaze 16.4 Occupied 18.8 Gaze 25.4 Social 30.1 Occupied 38.5 {end} {notes}</pre>	<pre>% SDS file saved by OdfSds Version 1.0 30/11/2007 17:53:25 Timed <seconds> (Gaze Occupied Social) ; <Unit#1> Occupied,+0 Gaze,+13 Occupied,+16 Gaze,+19 Social,+25 Occupied,+30 ,39 /</pre>

Figure 1. Examples of INTERACT and The Observer data files converted into SDIS files by programs ActSds and OdfSds, respectively.

A technical note: SDIS allows for explicit session start and stop times; if absent, they default to the onset time of the first code and the offset time of the last code, respectively. If an ACT file has explicit start and stop times listed in the Set line, then ActSds uses these; OdfSds uses the time on the {end} line as an explicit session stop time.

Research Factors

In its current version, INTERACT allows for sets with their associated events to be nested within groups. Thus, group labels could represent the levels of a research factor or variable; for example, the variable could be gender with two groups: male and female. The Observer lets users name levels of independent variables within each ODF file, one per line, on lines between a line that contains {indvar} and the line that contains {start}. SDIS permits users to define up to 7 unique variables that correspond to factors that reflect

the design of the study, along with up to 12 unique levels for each variable. For example, one factor might be gender with two levels (male, female), and a second factor might be treatment with two conditions (experimental, control), thereby defining four conditions, or cells, for this 2×2 design (i.e., two between-subjects variables, each with two levels).

Assigning units to conditions (i.e., levels of each factor) in SDIS is useful primarily when condition information (i.e., the values or levels for each variable) is exported to standard statistical programs (e.g., SPSS or SAS) along with summary statistics for each analytic unit so that, for example, standard ANOVAs may be performed. Additionally, all summary tables produced by GSEQ may optionally be pooled over levels of a factor, which is often useful during initial exploratory analyses of one's data.

In SDIS, factors—if used—are indicated by terms for levels enclosed in parentheses just before the slash that

terminates each session; for example, given the two variables mentioned earlier, a unit might end with

```
<identifier> sequential codes
and times (male, control)/
```

ActSds and OdfSds let the user name as many as seven factors (none, or one, or two are more typical) and specify their names. Then, factor levels can be assigned to each unit. In ActSds, factor level labels can also be formed from file names or group or set labels. In OdfSds, if users want the lines following {indvar} to be used as level names, then they name the number of factors, and for each, they indicate which line after {indvar} should be used for the factor level label. Factor names default to Factor#1 and so on, but the user can edit both factor names and their level names as desired. If the user specifies factors and their levels, then the appropriate syntax for the SDS file is generated.

INTERACT Categories and The Observer's Actors, Behaviors, and Modifiers

INTERACT organizes codes into categories (in the ACT datasheet, each category is a separate column), with each event on a separate line. Data lines in an ODF file contain information about the events coded. These lines have the general form

```
time [Act,] Beh [,Mod] [,Mod] [/*comment]
```

where *time* is a decimal number indicating seconds, *Act* indicates an actor code, *Beh* a behavior code, *Mod* a modifier code (e.g., the recipient of the behavior, its intensity or direction, etc.; Noldus, 1991), and square brackets indicate optional material. SDIS has a simpler structure. Its codes are unitary, not multipartite; moreover, GSEQ's flexible data modification commands let the user derive useful new codes from the initial ones. Thus, ActSds and OdfSds create codes for the SDS file by concatenating category and initial codes (INTERACT), or actor, behavior, and modifier codes (The Observer). Ideally, for readability of output, SDIS codes should be relatively brief—seven characters or fewer. Hence, both ActSds and OdfSds let the user edit the proposed SDIS code names before conversion. Further details are provided in technical write-ups that can be downloaded along with the programs.

SDIS Type, Code, and Condition Declarations

SDIS files begin with a type declaration and, optionally, a list of permissible codes and conditions, if any. ActSds and OdfSds generate declaration line(s) for the SDS file like this:

```
Timed (codes) [*conditions];
```

Timed indicates a TSD file, and the square brackets indicate optional material that is present only if factors (i.e., conditions) are defined. In its output, GSEQ orders codes as they are listed in this declaration, so users may wish to declare the codes in a way that makes sense to them. Furthermore, per SDIS syntax, codes that constitute a ME&E

set should be enclosed in parentheses. By default, ActSds does not define any ME&E sets, and any sets defined by OdfSds may not be correct (OdfSds assumes, e.g., that all behavior codes, or all codes derived from each actor [if any] constitute a single ME&E set). In either case, the user should edit the SDS file before compiling so that the declaration reflects the assignment of codes to ME&E sets. Two examples of data conversion are shown in Figure 1.

ActSds and OdfSds Programs: Details and Availability

The ActSds and OdfSds programs consist of three or four tabbed pages—each of which represents a step in data conversion—and three menus (*file*, *edit*, and *about*). Each tab contains a “when-done” button that lets the user move to the next tab. Again, further details are provided in technical write-ups that can be downloaded along with the programs. A zip file containing the ActSds program and its technical write-up, as well as the OdfSds program and its technical write-up, may be downloaded at no charge from www.gsu.edu/~psyrab/sg.htm or from www.ub.es/comporta/sg.htm (select “English” and “Download” or “Español” and “Descargar”). The programs are written in Pascal using Borland's Delphi Professional Version 7.0 and assume a Windows 95 or later environment.

AUTHOR NOTE

Correspondence concerning this article should be addressed to R. Bakeman, Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, GA 30302 (e-mail: bakeman@gsu.edu).

REFERENCES

- BAKEMAN, R., & QUERA, V. (1992). SDIS: A sequential data interchange standard. *Behavior Research Methods, Instruments, & Computers*, *24*, 554-559.
- BAKEMAN, R., & QUERA, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. Cambridge: Cambridge University Press.
- BAKEMAN, R., & QUERA, V. (2000). OTS: A program for converting Noldus Observer data files to SDIS files. *Behavior Research Methods, Instruments, & Computers*, *32*, 207-212.
- HOWE, G. W., DAGNE, G., & BROWN, C. H. (2005). Multilevel methods for modeling observed sequences of family interaction. *Journal of Family Psychology*, *19*, 72-85.
- MANGOLD, P. (2006, September). *Getting better results in less time: When using audio/video recordings in research applications make sense*. Paper presented at the 3rd Congress of the European Society on Family Relations, Darmstadt, Germany.
- NOLDUS, L. P. J. J. (1991). The Observer: A software system for collection and analysis of observational data. *Behavior Research Methods, Instruments & Computers*, *23*, 415-429.
- NOLDUS, L. P. J. J., TRIENES, R. J. H., HENDRIKSEN, A. H. M., JANSEN, H., & JANSEN, R. G. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, & Computers*, *32*, 197-206.
- NOLDUS INFORMATION TECHNOLOGY (2003). *The Observer: Professional system for collection, analysis, presentation and management of observational data* [Reference manual, Version 5.0]. Wageningen, The Netherlands: Author.

(Manuscript received December 11, 2007;
revision accepted for publication February 23, 2008.)