

TR-IT-0030

Labelling of Spontaneous Speech Data Using DAT-Link for AD and Xwaves+ for Segmentation

シンガー ハラルド
Harald Singer

ゼリグマン マーク
Mark Seligman

1993.12

In this report, we will describe how to download speech data, which was recorded on DAT, to UNIX files on a computer using DAT-Link. The files are then prepared for labelling in xwaves. This work was motivated by a request of Mark Seligman, who wanted to investigate the pause lengths in spontaneous conversations. This report is accessible for itl group members via `/homes/singer/tex/TR-IT-0030/tr.tex`.

目次

1	Introduction	1
2	Example Session	2
3	AD and DownSampling	4
4	Conversion of Files	5
5	Labeling	6
6	Conclusion and Future Improvements	7
	参考文献	9
	付録 A Setup Files	10
	A.1 .cshrc	10
	A.2 .twmrc or .vtwmc	10
	A.3 xwaves script	10
	A.4 full path names	11
	付録 B DA program for Xwaves+	11

1 Introduction

In ATR's new 7 year project, "spontaneous speech" is a major keyword. ATR is planning to collect a large database of spontaneous speech. Most of the data collection will be done by outside companies. However, the degree of spontaneity will be limited by not allowing responsive expressions (あいづち), maximum turn length of 10 seconds, etc.

Researchers, who would like to work with less constrained spontaneous speech must thus collect and process their data personally. In this report, we will explain our approach to download conversation data from DAT tape and label this data. The data was collected at ATR in a multimedia setup [2] [3]. Speaker A is on the left track, speaker B on the right track. The data had been recorded using a head-mounted Sennheiser HMD-410-6 microphone with not much cross-talk.

For labelling purposes, it is especially important to have the 2 speakers in a conversation synchronized and displayed at the same time, to find out about discourse cues etc. ¹

The overall process can be roughly described as

1. download DAT tapes to stereo files via DAT-Link hardware
2. split up stereo files into 2 channels and add SPHERE headers
3. segment automatically into pause units (optional) ²
4. correct pause unit labels by hand with xwaves

All processing was performed on a Sun SS10, where DAT-Link[4] and Xwaves+[1] software and hardware had been installed. Necessary changes to setup files are explained in Appendix A. The following chapters give an example session, DA (digital to analog) shell script, header conversion shell script and labelling shell script.

¹In previous "downloading work" at ATR, each track was downloaded separately and thus synchronization is difficult.

²As this is an open research topic it will not be described in this report.

2 Example Session

Assuming that you have a stereo DAT tape, i.e. left channel speaker A, right channel speaker B, we are going to DA this tape, convert the headers, autosegment the data and finally hand-correct the labels. We actually did this on a Sun SS10 with DAT-Link installed (hostname as62).

1. setup

- (a) verify hardware connections between DAT, DAT-Link and host
- (b) narecord must be in your path (`$ which narecord`) and NetAudio server must be running (`$ ps aux | grep netaudiod`)
- (c) make sure your DAT tape is write protected
- (d) hopefully xwaves has been properly installed

2. download and downsample to 12 kHz

- (a) put DAT tape into DAT recorder
- (b) enter without `CR` (carriage return)!

```
$ narecord -f ESPS -s 12000 -ostereo test.12Kstereo
```

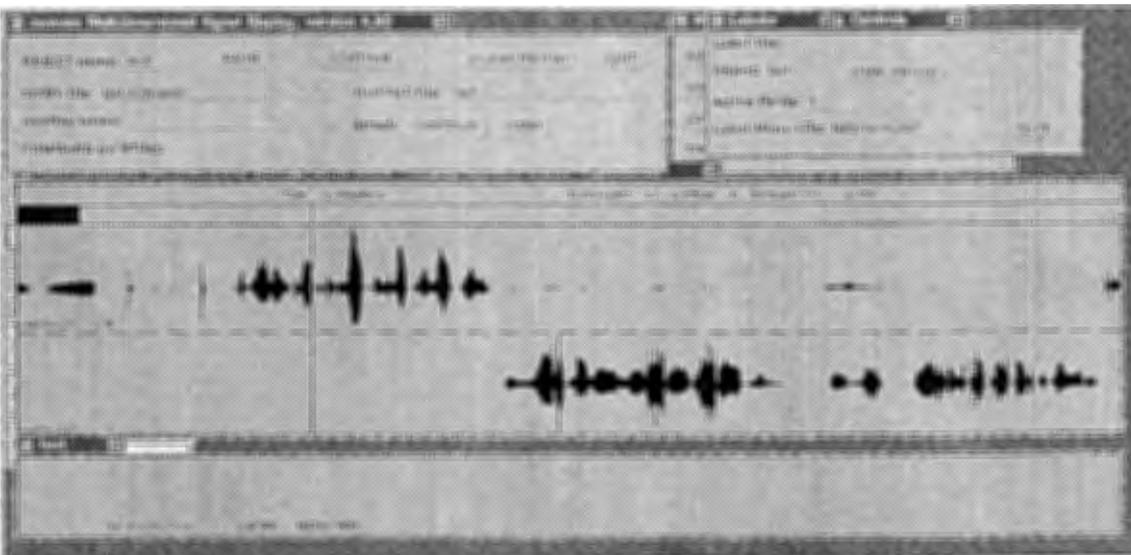
- (c) hit play button of DAT and then `CR` on keyboard
- (d) hit `Ctrl-C` on keyboard if desired conversation has finished; leave a few extra seconds because of buffering

3. view and listen in stereo (only possible with DAT-Link)

- (a) enter

```
$ setenv AUDIOPLAY as62:0.0
$ xwaves test.12Kstereo
```

- (b) put empty DAT tape into DAT recorder and push pause/record+play.
- (c) click and drag to listen in stereo



☒ 1: Xwaves

4. convert headers

(a) enter

```
$ demux test.12Kstereo test.tmp1 test.tmp2
$ bhd test.tmp1 | addhdr -l -r 12000 > test.left
$ h_edit -Ssample_byte_format=01 test.left
$ bhd test.tmp2 | addhdr -l -r 12000 > test.right
$ h_edit -Ssample_byte_format=01 test.right
```

After splitting up the ESPS stereo file into 2 files (demux), the ESPS header is removed (bhd), the SPHERE header is added (addhdr) including byte-swapping and the byte ordering information is corrected (h_edit).

5. automatic speech/non-speech segmentation and conversion to ESPS label format

(a) enter

```
$ Segment -i test.left -O test.left.LB
$ gawk -f atr2esps.awk type=atr test.left.LB > test.left.lab
```

6. view and correct

(a) enter

```
$ xwaves test.left test.right
```

(b) attach if xlabel not yet attached

(c) enter

```
test.left.lab
```

in window with title Labeler in row with title Label File

3 AD and DownSampling

With DAT-Link, DA and down sampling can be done in one step. Furthermore, 2 channels can be recorded at the same time. This drastically reduces the necessary amount of work.

The following script shows how to download a total of 16 conversations. The resulting stereo files will be in ESPS format.

```
#!/bin/csh -f
# from DAT to ESPS stereo files for all conversations
# example:
#   recordall.csh
# stop with Ctrl-C at end of each conversation

onintr CATCH

set MAXSEC = 1200      # assumed to be 20 minutes maximum
set SAMP = 16000

set FILETYPE = ( TL MM )
set CONV = ( 01 02 03 04 05 06 07 08 )

foreach F ( $FILETYPE )
  foreach C ( $CONV )
    set STEREO = "$F_"$C".stereo"
    echo "please enter tape for $STEREO"
    echo "CR when done and play button pressed"
    set ANSWER = $<

    narecord -f ESPS -s $SAMP -ostereo -t $MAXSEC $STEREO

  end
end

CATCH:
  ls -l $STEREO
end

exit 0
### EOF
```

The downloaded, downsampled files are shown here:

```
-rwxrwx---  1 singer  research 11252205 Nov 26 16:38 MM_01.stereo
-rwxrwx---  1 singer  research  6458349 Nov 26 16:41 MM_02.stereo
...
-rwxrwx---  1 singer  research  8423917 Nov 26 15:40 TL_01.stereo
-rwxrwx---  1 singer  research  3935213 Nov 26 15:42 TL_02.stereo
...
-rwxrwx---  1 singer  research 30727145 Nov 26 18:06 TL_08.stereo
```

4 Conversion of Files

The ESPS format is somehow hard to deal with, as it is not fully documented. We thus convert each ESPS stereo file into 2 SPHERE files, which can also be read by xwaves.

The following script shows how to convert the above mentioned 16 conversations into SPHERE format. We choose to byte-swap the files and also downsample them from 16 kHz to 12 kHz as most our signal processing software is written for 12 kHz and so that they can be easily played by a DASBOX on a DECstation.³

```
#!/bin/csh -f
# splits ESPS stereo file into left and right channel

# temporary files
set LEFT = left.esps
set RIGHT = right.esps

set FILETYPE = ( TL MM )
set CONV = ( 01 02 03 04 05 06 07 08 )

foreach F ( $FILETYPE )
  foreach C ( $CONV )

    set STEREO = "$F_"$C".stereo"

    echo "splitting $STEREO into left and right channels"
    demux $STEREO $LEFT $RIGHT

    echo "$F"_A_"$C".12K: cut ESPS header, byte swap, add SPHERE Header
    bhd $LEFT | srconv is=16 os=12 | addhdr -l -r 12000 > "$F"_A_"$C".12K
    h_edit -Ssample_byte_format=01 "$F"_A_"$C".12K

    echo "$F"_B_"$C".12K: cut ESPS header, byte swap, add SPHERE Header
    bhd $RIGHT | srconv is=16 os=12 | addhdr -l -r 12000 > "$F"_B_"$C".12K
    h_edit -Ssample_byte_format=01 "$F"_B_"$C".12K

  end
end

rm $LEFT $RIGHT
exit 0
### EOF

-r-xrwx---  1 singer  research  5626880 Nov 26 19:24 MM_A_01.12K
-r-xrwx---  1 singer  research  5626880 Nov 26 19:25 MM_B_01.12K
...
-r-xrwx---  1 singer  research  4212736 Nov 26 18:39 TL_A_01.12K
-r-xrwx---  1 singer  research  4212736 Nov 26 18:40 TL_B_01.12K
...
```

³The original downloading into stereo files should be done at 16 kHz and these files should be kept somewhere as the future standard at ATR will probably be 16 kHz. Previously downsampled and labelled data can thus directly be used without any additional labelling work. If, on the other hand, speech data is downloaded again from the DAT tape there will always be a time alignment problem.

5 Labeling

Labelling is the most labor intensive part of the job. We are using Xwaves+ by ENTROPICS[1]. As Xwaves+ labels events and not durations we define the events shown in Tab. 1.

表 1: Event labels

event type	symbol
start of turn	s
end of turn	e
start of segment inside turn	d
end of segment inside turn	f
start of responsive expression	r
end of responsive expression	t

A typical example for a responsive expression (あいずち) would be “hai”(はい). An example of a labelled conversation is given in Fig. 2.

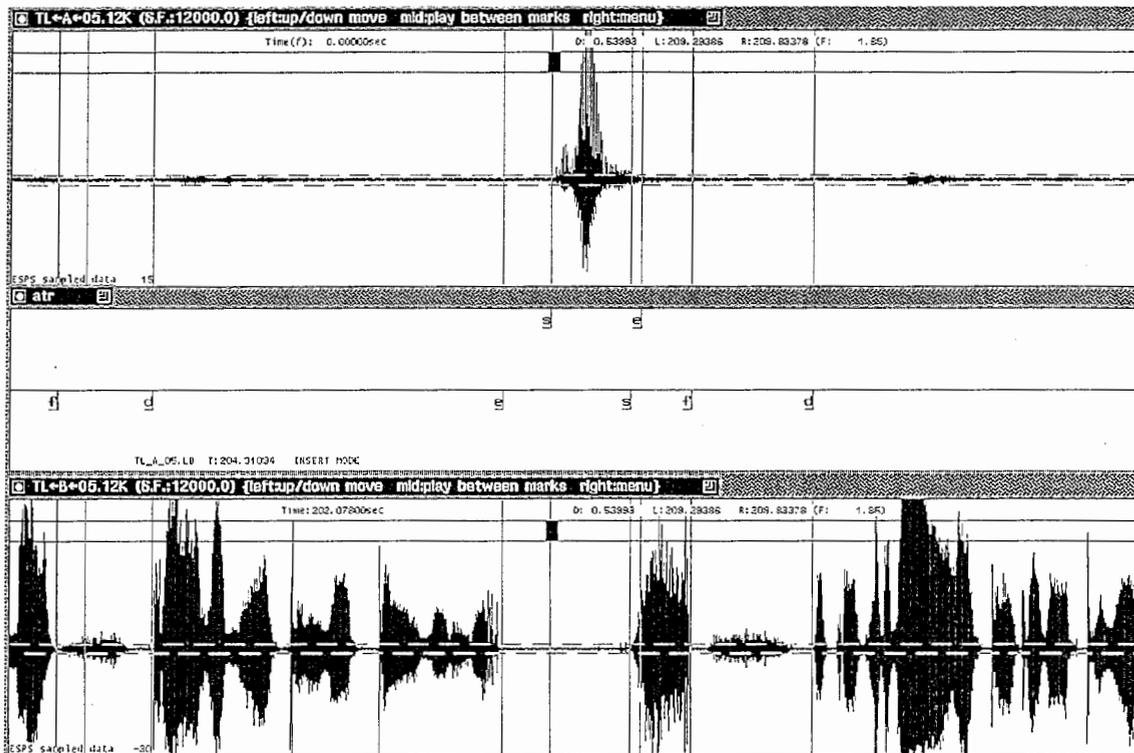


図 2: Xwaves window for labelling

The labeller scrolls through the conversation by LM, MM and RM (left, middle or right mouse button) on either wave file. In Insert mode in the xlabel window, event types can be entered via keyboard. With LM on the xlabel window the corresponding wave between left and right labels is played back (see xwaves manual for further details[1]).

The following shell script loads the desired conversation for labelling.

```
#!/bin/csh -f
# example:
#   label.csh -m TL -c 01 -t A

# temporary wave command file
```

```

set WAVECOM = ".wavecom$$"

###          set argline defaults          ###
set MODE = "TL"
set CONV = "01"
set TOP = "A"

#-----
###          check command line arguments  ###
set com = $0
while($#argv)
switch($argv[1])
    case -c*:
        set CONV = $argv[2]; shift; breaksw
    case -m*:
        set MODE = $argv[2]; shift; breaksw
    case -t*:
        set TOP = $argv[2]; shift; breaksw
    default:
        goto USAGE
endsw
shift
end

if ( $TOP == "A" ) then
    set BOTTOM = "B"
else
    set BOTTOM = "A"
endif

set FILET = "$MODE"_"$TOP"_"$CONV"
set FILEB = "$MODE"_"$BOTTOM"_"$CONV"

cat > $WAVECOM << EOF
waves set ref_size 5 ref_step 5
set play_prog /homes/singer/bin/waves_play
set plot_min -1400 plot_max 1400
make name atr file $FILEB.12K loc_x 1 loc_y 620
send make name atr file $FILEB.LB
make name atr file $FILET.12K loc_x 1 loc_y 160
send make name atr file $FILET.LB
EOF

xwaves $WAVECOM
rm $WAVECOM
exit 0

USAGE:
    echo "usage: $com:t"
    echo " [-c conversation number]          VAL: $CONV"
    echo " [-m MODE]                             VAL: $MODE"
    echo " [-t TOPSPEAKER]                         VAL: $TOP"
    exit 1
### EOF

```

6 Conclusion and Future Improvements

This report described how to efficiently create speech files on a workstation. A task which up to now took about a week can now be accomplished in some hours. Basically, researchers can now easily collect and

label databases which are necessary for their research.

Recently, ATR ITL has become a beta test site for ENTROPICS Aligner software. This software requires only orthographic word labels, which can be entered for example by using `xchart`, which is included in the `Xwaves` distribution. HMM Viterbi alignment is then used to generate the best matching phoneme labels. Using this software, the task of labelling the necessarily huge amounts of speech for probabilistic models can be speeded up by orders of magnitude and with minimal costs.

参考文献

- [1] Entropic Research Laboratory. *waves+*, 1993.
- [2] R. Furukawa, F. Yato, and K. Loken-Kim. Analysis of telephone and multimedia dialogues. Technical Report TR-IT-0020, ATR, 1993. (in Japanese).
- [3] K. Loken-Kim, F. Yato, K. Kurihara, L. Fais, and R. Furukawa. EMMI-ATR environment for multi-modal interaction. Technical Report TR-IT-0018, ATR, 1993.
- [4] Townshend Computer Tools. *DAT-Link/DAT-Link+ User's Manual*, 1992.

付録 A Setup Files

A.1 .cshrc

If you use csh or tcsh, you should add the following to your resource file. If you use bash, please add something similar.

```
# for ESPS
setenv ESPS_BASE /usr/local/esps
setenv ELM_HOST as57
set path = ( $path $ESPS_BASE/bin )
setenv MANPATH $MANPATH:"$ESPS_BASE"/man"
setenv BBOX_QUIT_BUTTON 1

# for DAT-Link
set path = ( $path /usr/local/datlink/bin /usr/pi/bin )
```

A.2 .twmrc or .tvtwmrc

If you use twm or tvtwm window manager, the following lines have to be added to the resource files. Doesn't work on the HP's!

```
UsePPosition "non-zero"
RandomPlacement
```

A.3 /bin/xwaves

If several users want to use Xwaves+ on the same CPU, the connection to xlabel gets confounded. The solution is to have your private xwaves script, which should be in your path and before the standard xwaves command.

```
#!/bin/csh -f
# This script provides a relatively simple solution to the problems
# encountered when more than one user wishes to run xwaves on a given
# host at the same time. To prepare for the use of this method, each
# xwaves user needs to do the following:
#
# If you don't already have a personal bin directory called $HOME/bin,
# then create one.
#
# Place a copy of this script in a file called $HOME/bin/xwaves and then
# do:
#
# chmod a+rx $HOME/bin/xwaves
#
# touch $HOME/bin/send_xwaves
# chmod a+rw $HOME/bin/send_xwaves
#
# Modify your shell initialization resource file (e.g. $HOME/.cshrc) so
# that $HOME/bin is before $ESPS_BASE/bin on your UNIX PATH.
#
# source $HOME/.cshrc
#
# rehash (if you are running the C-shell)
#
# THE ABOVE OPERATIONS ONLY NEED TO BE PERFORMED ONCE. Thereafter, you
# should be able to use xwaves (and send_xwaves) immediately upon login.
# set echo
```

```

rm -f $HOME/bin/send_xwaves

set prognum = 100$$

echo $ESPS_BASE/bin/send_xwaves -P $prognum '$*' > $HOME/bin/send_xwaves
chmod a+rxw $HOME/bin/send_xwaves

$ESPS_BASE/bin/xwaves -P $prognum $*

exit 0
#### EOF

```

A.4 full path names

All programs and scripts used in this report can be currently accessed on as62 via

```

/usr/local/datlink/bin/narecord
/usr/local/esps/bin/bhd
/usr/local/esps/bin/demux
/usr/pi/bin/addhdr
/usr/pi/bin/srconv
/usr/pi/bin/h_edit
/home/atrh16/singer/bin/SunOS/Segment
/home/atrh16/singer/bin/atr2esps.awk
/home/atrh16/singer/bin/xwaves

```

付録 B DA program for Xwaves+

We want to use the AD/DA converters connected to our workstations. For the DECstation, that would be the DASBOX, for the SUN's the built-in AD/DA converter. The following script automatically finds out which DA to start. Admittedly, this is very machine-dependent. This script is accessed in Xwaves+ either via \$HOME/.wave_pro or with set play_prog waves_play in a Xwaves+ command file.

```

#!/bin/csh -f
# H.Singer 15FEB92
# just an example how to use DA on a remote host
# uses environment variable AUDIOPLAY, which should be used like DISPLAY
#   setenv AUDIOPLAY atrq15:0.0

# get the current directory, there is no elegant way!!
set DIR = `pwd | sed -e 's/\//tmp_mnt//`
echo $DIR

if (! $?AUDIOPLAY) then
  echo "@WARNING: AUDIOPLAY variable not set"
  switch ($USER)
  case singer:
    setenv AUDIOPLAY atrq15:0.0; breaksw
  case seligman:
    setenv AUDIOPLAY as24:0.0; breaksw
  case kitagawa:
    setenv AUDIOPLAY as28:0.0; breaksw
  case fais:
    setenv AUDIOPLAY as12:0.0; breaksw
  default:
    echo "no default audio provided for $USER"; exit 1
  endsw

```

```
endif

# extract sampling frequency; has to be adjusted to kHz
@ SF = 'hditem -i record_freq $2' / 1000

# offset and number of samples
set START = 'echo $1 | sed -e 's/-r//' | sed -e 's/:+.*$//''
set COUNT = 'echo $1 | sed -e 's/-r.*:+'/'

echo $START $COUNT $2 $SF

switch ( $AUDIOPLAY )
  case atrq15:0*:
    # DASBOX
    set DAOUT = /usr/dasbox/bin/daout
    if ( 'hditem -s -i sample_byte_format $2' ) then
      # NIST header
      @ START += 1024
    endif
    rsh atrq15 "setenv DASBOX 140; cd $DIR; $DAOUT -O $START -T $COUNT -f $SF $2"
    breaksw
  case as52:0*:
    # DAT-Link
    naplay -f ESPS $*
    breaksw
  case as*:
    # for department 4
    set AHOST = 'echo $AUDIOPLAY | gawk -F: '{print $1}''
    echo $AHOST
    # Sparc2 built-in
    set ENCODING = lin16LSB
    /home/as53/awb/bin/sun4/seligmansubwave $START $COUNT $2 |\
      srconv gain=4.0 it=Short ot=ulaw is=12 os=8 |\
      rsh $AHOST "(cat >/dev/audio)"
    breaksw
  default:
    echo "DA is not supported for <$AUDIOPLAY>"
    exit 1
endsw

exit 0
### EOF
```