Examples for Using Speech Signal Processing Toolkit
Ver. 3.9
SPTK working group
December 25, 2015

Contents

1 Basics

1.1 Help message ............................................. 3
1.2 Data type conversion between "little endian" and "big endian." ........................................................................... 3
1.3 Dump a binary data file ........................................ 3
1.4 Data type conversion from "short int" to "float" .................................................. 3
1.5 Plotting speech waveform on X-window ........................................ 3
1.6 Save the figure in an Encapsulated PostScript file .................................................. 4
1.7 Play a sound file ............................................. 3
1.8 Cut a portion out of a file ........................................ 4

2 Pitch Extraction from Speech Waveform

2.1 A pitch extractor ............................................. 5
2.2 Plotting the extracted pitch contour .................................................. 6

3 Speech Analysis/Synthesis Based on Mel-Cepstral Representation

3.1 Mel-cepstral analysis of speech .................................................. 6
3.2 Plotting spectral estimates from mel-cepstrum .................................................. 6
3.3 Plotting the spectral estimate with the FFT spectrum .................................................. 7
3.4 Speech synthesis from mel-cepstrum .................................................. 8

4 Speech Analysis/Synthesis based on LPC

4.1 LPC analysis of speech ........................................ 9
4.2 Plotting spectral estimates from LPC coefficients .................................................. 10
4.3 Plotting the spectral estimate with the FFT spectrum .................................................. 10
4.4 Speech synthesis from LPC coefficients .................................................. 11
4.5 Obtain PARCOR coefficients from LPC coefficients .................................................. 12
4.6 Speech synthesis from PARCOR coefficients .................................................. 12
4.7 Obtain LSP coefficients from LPC coefficients .................................................. 13
4.8 Speech synthesis from LSP coefficients .................................................. 13

5 Speech Analysis/Synthesis Based on Mel-Generalized Cepstral Representation

5.1 Mel-generalized cepstral analysis of speech .................................................. 14
5.2 Plotting spectral estimates from mel-generalized cepstrum .................................................. 15
5.3 Plotting the spectral estimate with the FFT spectrum .................................................. 16
5.4 Speech synthesis from mel-generalized cepstrum .................................................. 17
1  Basics

1.1  Help message

impulse -h

1.2  Data type conversion between “little endian” and “big endian.”

Files:  
  data.short: speech data included in this example (short integer, 16 kHz sampling, little endian)
  data.short-b: speech data (short integer, 16 kHz sampling, big endian)

swab +s < data.short > data.short-b

1.3  Dump a binary data file

Files:  
  data.short: speech data included in this example (short integer, 16 kHz sampling)

dmp +s data.short | less

1.4  Data type conversion from “short int” to “float”

Files:  
  data.short: speech data included in this example (short integer, 16 kHz sampling)
  data.float: speech data (float, 16 kHz sampling)

x2x +sf < data.short > data.float

1.5  Plotting speech waveform on X-window

Files:  
  data.short: speech data included in this example (short integer, 16 kHz sampling)

gwave +s data.short | xgr

\footnote{By clicking links in this PDF file, your PC may play some speech files, which were converted from “float” format into “wav” format (16 kHz sampling, 16-bit integer).}

\footnote{If you compiled SPTK with “--enable-double” option, please use “+sd” option instead of “+sf” and “+d” option instead of “+f”.

3
1.6 Save the figure in an Encapsulated PostScript file

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)
figure.eps: Encapsulated PostScript file

```
gwave +s data.short | psgr > figure.eps
```

1.7 Play a sound file

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)

Note: This works only on Linux, Solaris, and FreeBSD.

da +s -s 16 -a 100 data.short

1.8 Cut a portion out of a file

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)
2 Pitch Extraction from Speech Waveform

2.1 A pitch extractor

Files: `data.short`: speech data included in this example (short integer, 16 kHz sampling)

Conditions: frame period: 80 points (5 ms)
- minimum fundamental frequency for search: 80 Hz
- maximum fundamental frequency for search: 165 Hz

Note: Options should be adjusted for each speech data.

```
x2x +sf data.short | pitch -a 1 -s 16 -p 80 -L 80 -H 165 > data.pitch
```
2.2 Plotting the extracted pitch contour

Files: data.pitch: pitch data extracted from speech data "data.short" (float)

Conditions: Minimum value of vertical axis: 0.0
Width: 15 cm
Height: 4 cm

fdrw -y 0 250 -W 1.5 -H 0.4 < data.pitch | xgr

3 Speech Analysis/Synthesis Based on Mel-Cepstral Representation

3.1 Mel-cepstral analysis of speech

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)
data.mcep: mel-cepstrum (float)

Conditions: frame length: 400 points (25 ms)
frame period: 80 points (5 ms)
window: Blackman window
analysis order: 20
frequency warping parameter: $\alpha = 0.42$
FFT size: 512 points

x2x -sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 | mcep -l 512 -m 20 -a 0.42 > data.mcep

3.2 Plotting spectral estimates from mel-cepstrum

Files: data.mcep: mel-cepstrum (float)

Conditions: analysis order: 20
frequency warping parameter: $\alpha = 0.42$
FFT size: 512 points
plotted frames: from 10-th to 135-th
sampling frequency: 16 kHz

bcut +f -n 20 -s 10 -e 135 < data.mcep | mgc2sp -m 20 -a 0.42 -g 0 -l 512 | grlogsp -l 512 -x 8 | xgr
3.3 Plotting the spectral estimate with the FFT spectrum

Files: data.mcep: mel-cepstrum (float)

Conditions: analysis order: 20
  frequency warping parameter: $\alpha = 0.42$
  FFT size: 512 points
plotted frame: 65-th
sampling frequency: 16 kHz

( x2x +sf < data.short | frame -l 400 -p 80 | \ bcut +f -l 400 -s 65 -e 65 \ window -l 400 -L 512 | spec -l 512 | \ glogsp -l 512 -x 8 -p 2 ; \ bcut +f -n 20 -s 65 -e 65 < data.mcep | \ mgc2sp -m 20 -a 0.42 -g 0 -l 512 | glogsp -l 512 -x 8 ) | xgr

3.4 Speech synthesis from mel-cepstrum

Files:
data.pitch: pitch data extracted from speech data "data.short" (float)
data.mcep: mel-cepstrum (float)
data.mcep.syn: synthesized speech (float)

Conditions:
frame period: 80 points (5 ms)
analysis order: 20
frequency warping parameter: $\alpha = 0.42$

excite -p 80 data.pitch | \ mlsadf -m 20 -a 0.42 -p 80 data.mcep > data.mcep.syn
gwave data.mcep.syn | xgr
4 Speech Analysis/Synthesis based on LPC

4.1 LPC analysis of speech

Files:  
data.short: speech data included in this example (short integer, 16 kHz sampling)  
data.lpc: LPC coefficients (float)

Conditions:  
frame length: 400 points (25 ms)  
frame period: 80 points (5 ms)  
window: Blackman window  
analysis order: 20

```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 |  
lpc -l 400 -m 20 > data.lpc
```
4.2 Plotting spectral estimates from LPC coefficients

**Files:** data.lpc: LPC coefficients (float)

**Conditions:** analysis order: 20

```
bcut +f -n 20 -s 10 -e 135 < data.lpc |\
spec -l 512 -n 20 | grlogsp -l 512 -x 8 | xgr
```

or

```
bcut +f -n 20 -s 10 -e 135 < data.lpc |\
mgc2sp -m 20 -a 0 -g -1 -n -u -l 512 |\
grlogsp -l 512 -x 8 | xgr
```

---

4.3 Plotting the spectral estimate with the FFT spectrum

**Files:** data.lpc: LPC coefficients (float)
Conditions: analysis order: 20
plotted frame: 65-th
sampling frequency: 16 kHz

\(\text{x2x +sf < data.short | frame -l 400 -p 80 | }\)
\(\text{bcut +f -l 400 -s 65 -e 65 | }\)
\(\text{window -l 400 -l 512 | spec -l 512 | }\)
\(\text{glogsp -l 512 -x 8 -p 2 ;}\)
\(\text{bcut +f -n 20 -s 65 -e 65 < data.lpc > data.tmp ;}\)
\(\text{spec -l 512 -n 20 -p data.tmp | glogsp -l 512 -x 8 ;}\)
\(\text{rm data.tmp ) | xgr}\)

4.4 Speech synthesis from LPC coefficients

Files: data.pitch: pitch data extracted from speech data "data.short" (float)
data.lpc: LPC coefficients (float)
data.lpc.syn: synthesized speech (float)

Conditions: frame period: 80 points (5 ms)
analysis order: 20

excite -p 80 data.pitch | poledf -m 20 -p 80 data.lpc > data.lpc.syn
gwave +f data.lpc.syn | xgr


da +f -s 16 data.lpc.syn

4.5 Obtain PARCOR coefficients from LPC coefficients

Files: data.lpc: LPC coefficients (float)
data.par: PARCOR coefficients (float)

Conditions: analysis order: 20

lp2par -m 20 < data.lpc > data.par

4.6 Speech synthesis from PARCOR coefficients

Files: data.pitch: pitch data extracted from speech data “l” (float)
data.par: PARCOR coefficients (float)
data.syn: synthesized speech (float)

Conditions: frame period: 80 points (5 ms)
analysis order: 20
4.7 Obtain LSP coefficients from LPC coefficients

**Files:**
- data.lpc: LPC coefficients (float)
- data.lsp: LSP coefficients (float)

**Conditions:**
- analysis order: 20
- split number of unit circle: 256

```
lpc2lsp -m 20 -n 256 < data.lpc > data.lsp
```

4.8 Speech synthesis from LSP coefficients

**Files:**
- data.pitch: pitch data extracted from speech data """" (float)
- data.lsp: LSP coefficients (float)
- data.lsp.syn: synthesize speech (float)
Conditions: frame period: 80 points (5 ms)
  analysis order: 20

excite -p 80 data.pitch | lspdf -m 20 -p 80 data.lsp > data.lsp.syn

gwave +f data.lsp.syn | xgr

da +f -s 16 data.lsp.syn

5 Speech Analysis/Synthesis Based on Mel-Generalized Cepstral Representation

5.1 Mel-generalized cepstral analysis of speech

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)
data.mgcep: mel-generalized cepstrum (float)
Conditions: frame length: 400 points (25 ms)
frame period: 80 points (5 ms)
window: Blackman window
analysis order: 20
frequency warping parameter: $\alpha = 0.42$
power parameter: $\gamma = -1/2$

\begin{verbatim}
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 |
mgcep -m 20 -a 0.42 -c 2 -l 512 > data.mgcep
\end{verbatim}

5.2 Plotting spectral estimates from mel-generalized cepstrum

Files: data.mgcep: mel-generalize cepstrum (float)

Conditions: analysis order: 20
frequency warping parameter: $\alpha = 0.42$
power parameter: $\gamma = -1/2$
plotted frames: from 10-th to 135-th
sampling frequency: 16 kHz

\begin{verbatim}
bcut +f -n 20 -s 10 -e 135 < data.mgcep |
mgc2sp -m 20 -a 0.42 -c 2 -l 512 | grlogsp -l 512 -x 8 | xgr
\end{verbatim}
5.3 Plotting the spectral estimate with the FFT spectrum

**Files:** data.mgcep: mel-generalized cepstrum (float)

**Conditions:**
- analysis order: 20
- frequency warping parameter: $\alpha = 0.42$
- power parameter: $\gamma = -1/2$
- FFT size: 512 points
- plotted frame: 65-th
- sampling frequency: 16 kHz

```
( x2x +sf < data.short | frame -l 400 -p 80 | bcut +f -l 400 -s 65 -e 65 | window -l 400 -L 512 | spec -l 512 | glogsp -l 512 -x 8 -p 2 ;
 \ bcut +f -n 20 -s 65 -e 65 < data.mgcep | \
```
5.4 Speech synthesis from mel-generalized cepstrum

**Files:**
- data.pitch: pitch data extracted from speech data "\[\]
- data.mgcep: mel-generalized cepstrum (float)
- mgcep.syn: synthesized speech (float)

**Conditions:**
- Frame period: 80 points (5 ms)
- Analysis order: 20
- Frequency warping parameter: $\alpha = 0.42$
- Power parameter: $\gamma = -1/2$

```
excite -p 80 data.pitch | mglsadf -m 20 -a 0.42 -c 2 -p 80 data.mgcep > data.mgcep.syn
gwave +f data.mgcep.syn | xgr
```
6 Vector Quantization of Mel-Cepstrum

6.1 Train a (very small) Codebook

Files:  data.mcep: mel-cepstrum for training (float)  
        codebook.mcep: codebook (float)

Conditions:  vector size: 21 (analysis order: 20)  
             codebook size: 32

lbg -n 20 -e 32 < data.mcep > codebook.mcep

6.2 Encode (training vectors)

Files:  codebook.mcep: codebook (float)  
        data.mcep.index: index (int)
Conditions:  vector size: 21 (analysis order: 20)
          codebook size: 32

vq -n 20 codebook.mcep < data.mcep > data.mcep.index

6.3 Decode (training vectors)

Files:  codebook.mcep: codebook (float)
        data.mcep.index: index (int)
        data.mcep.vq: quantized mel-cepstrum (float)

Conditions:  vector size: 21 (analysis order: 20)
             codebook size: 32

ivq -n 20 codebook.mcep < data.mcep.index > data.mcep.vq

6.4 Plotting original and quantized spectra

Files:  data.mcep: original mel-cepstrum (float)
        data.mcep.vq: quantized mel-cepstrum (float)

Conditions:  analysis order: 20
             frequency warping parameter: $\alpha = 0.42$
             plotted frames: from 10-th to 135-th
             sampling frequency: 16 kHz

( bcut +f -n 20 -s 10 -e 135 < data.mcep |\n  mgc2sp -m 20 -a 0.42 -g 0 -l 512 |\n  grlogsp -l 512 -x 8 -O 1 -c "(a) original" ;\n \bcut +f -n 20 -s 10 -e 135 < data.mcep.vq |\n  mgc2sp -m 20 -a 0.42 -g 0 -l 512 |\n  grlogsp -l 512 -x 8 -O 2 -c "(b) quantized" ) | xgr
6.5 Performance evaluation on the training data

**Files:**
- codebook.mcep: codebook (float)
- data.mcep.index: index (int)
- data.mcep.vq: quantized vectors (float)
- data.mcep.vq.cdist: cepstrum distortion in dB (float)

**Conditions:**
- vector size: 21 (analysis order: 20)
- codebook size: 32

```
freqt -a 0.42 -m 20 -A 0 -M 255 < data.mcep > data.mcep.cep
cfreqt -a 0.42 -m 20 -A 0 -M 255 < data.mcep.vq |\
cdist data.mcep.cep -m 255 > data.mcep.vq.cdist\rm data.mcep.cep
```
6.6 Speech synthesis from quantized mel-cepstrum

**Files:**
- data.pitch: pitch data extracted from speech data (float)
- data.mcep.vq: quantized mel-cepstrum (float)
- data.mcep.vq.syn: synthesized speech (float)

**Conditions:**
- frame period: 80 points (5 ms)
- analysis order: 20
- frequency warping parameter: $\alpha = 0.42$

```
excite -p 80 data.pitch \mlsadf -m 20 -a 0.42 -p 80 data.mcep.vq > data.mcep.vq.syn
gwave +f data.mcep.vq.syn | xgr
```

```
da +f -s 16 data.mcep.vq.syn
```
7 Preparation of Speech Parameter for Speech Recognition

7.1 Cepstrum derived from LPC analysis (LPC cepstrum)

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)

Conditions:
- frame length: 400 points (25 ms)
- frame period: 80 points (5 ms)
- window: Blackman window
- analysis order: 12
- order of LPC cepstrum: 12

```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 |\n  lpc -l 400 -m 12 | lpc2c -m 12 -M 12 > data.lpc.cep
```

7.2 Mel-cepstrum derived from LPC analysis (LPC mel-cepstrum)

Files: data.short: speech data included in this example (short integer, 16 kHz sampling)

Conditions:
- frame length: 400 points (25 ms)
- frame period: 80 points (5 ms)
- window: Blackman window
- analysis order: 12
- order of LPC mel-cepstrum: 12

```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 |\n  lpc -l 400 -m 12 | lpc2c -m 12 -M 256 |\n  freqt -m 256 -a 0 -M 12 -A 0.42 > data.lpc.mcep
```

or

```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 |\n  lpc -l 400 -m 12 |\n  mgc2mgc -m 12 -a 0 -g -1 -n -u -M 12 -A 0.42 -G 0 > data.lpc.mcep
```

7.3 Mel-cepstrum obtained by mel-cepstral analysis

Files: data.mcep: mel-cepstrum (float)

Conditions:
- frame length: 400 points (25 ms)
- frame period: 80 points (5 ms)
- window: Blackman window
- analysis order: 20
- frequency warping parameter: $\alpha = 0.42$
- FFT size: 512 points

```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 |\n  mcep -l 512 -m 12 -a 0.42 > data.mcep.mcep
```
7.4 Mel-cepstrum derived from mel-generalized cepstral analysis

**Files:** data.short: speech data included in this example (short integer, 16 kHz sampling)

**Conditions:** frame length: 400 points (25 ms)
frame period: 80 points (5 ms)
Blackman window
FFT size: 512 points
$(\alpha, \gamma)$ for analysis: (0.42, -0.5)
analysis order: 12
order of mel-cepstrum: 12

```
x2s +sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 |  
mcep -m 12 -a 0.42 -c 2 -l 512 |  
mgc2mgc -m 12 -a 0.42 -c 2 -M 12 -A 0.42 -G 0 > data.mgcep.mcep
```

7.5 Plotting spectra for each speech recognition parameter

**Files:** data.lpc.cep: LPC cepstrum (float)
data.lpc.mcep: LPC mel-cepstrum (float)
data.mcep.mcep: mel-cepstrum (float)
data.mgcep.mcep: mel-cepstrum derived from mel-generalized cepstrum (float)

**Conditions:** plotted frames: from 10-th to 135-th

```
\  
bcut +f -n 12 -s 10 -e 135 < data.lpc.cep  
mgc2sp -m 12 -a 0 -g 0 -l 512 \  
grlogsp -l 512 -x 8 -0 1 -c "(a) LPC-CEP" ;\  
bcut +f -n 12 -s 10 -e 135 < data.lpc.mcep  
mgc2sp -m 12 -a 0.42 -g 0 -l 512 \  
grlogsp -l 512 -x 8 -0 2 -c "(b) LPC-MCEP" ;\  
bcut +f -n 12 -s 10 -e 135 < data.mcep.mcep  
mgc2sp -m 12 -a 0.42 -g 0 -l 512 \  
grlogsp -l 512 -x 8 -0 3 -c "(c) MCEP" ;\  
bcut +f -n 12 -s 10 -e 135 < data.mgcep.mcep  
mgc2sp -m 12 -a 0.42 -g 0 -l 512 \  
grlogsp -l 512 -x 8 -0 4 -c "(d) MGCEP-MCEP" ) | xgr
```
8 Playing with the Vocoder Based on Mel-Cepstrum

8.1 High- or low-pitched voice

Files:

```
sopr -m 0.4 data.pitch |
excite -p 80 | mlsadrf -m 20 -a 0.42 -p 80 data.mcep |
tee data.mcep.high.syn | da +f -s 16
```

```
sopr -m 2 data.pitch |
excite -p 80 | mlsadrf -m 20 -a 0.42 -p 80 data.mcep |
tee data.mcep.low.syn | da +f -s 16
```
8.2 Fast- or slow-speaking voice
Files: 
- data.mcep.fast.syn: synthesized speech (float)
- data.mcep.slow.syn: synthesized speech (float)

sopr -m 1 data.pitch |\
excite -p 40 | mlsadf -m 20 -a 0.42 -p 40 data.mcep |\
tee data.mcep.fast.syn | da +f -s 16

sopr -m 1 data.pitch |\
excite -p 160 | mlsadf -m 20 -a 0.42 -p 160 data.mcep |\
tee data.mcep.slow.syn | da +f -s 16

8.3 Hoarse voice
Files: 
- data.mcep.hoarse.syn: synthesized speech (float)

sopr -m 0 data.pitch |\
excite -p 80 | mlsadf -m 20 -a 0.42 -p 80 data.mcep |\
tee data.mcep.hoarse.syn | da +f -s 16

8.4 Robotic voice
Files: 
- data.mcep.robot.syn: synthesized speech (float)

train -p 200 -l -1 | mlsadf -m 20 -a 0.42 -p 80 data.mcep |\
tee data.mcep.robot.syn | da +f -s 16

8.5 Child-like or deep voice
Files: 
- data.mcep.child.syn: synthesized speech (float)
- data.mcep.deep.syn: synthesized speech (float)

sopr -m 0.4 data.pitch |\
excite -p 80 | mlsadf -m 20 -a 0.1 -p 80 data.mcep |\
tee data.mcep.child.syn | da +f -s 16

sopr -m 2 data.pitch |\
excite -p 80 | mlsadf -m 20 -a 0.6 -p 80 data.mcep |\
tee data.mcep.deep.syn | da +f -s 16

8.6 Various voices
Files: 
- data.float: original speech (float)
- data.mcep.syn: synthesized speech (float)
- data.mcep.{high,low,fast,slow,hoarse,robot,child,deep}.syn: synthesized speech (float)

da +f -v -s 16 data.float data.mcep.syn \
data.mcep.{high,low,fast,slow,hoarse,robot,child,deep}.syn
9 Speech Synthesis Based on HMM

9.1 Speech parameter generation from a sequence of HMMs

Files: sample.pdf: sequence of mean and variance corresponding to a state sequence included in this example (float, little endian)
sample.mcep: mel-cepstrum generated from a sequence of HMMs (float)

Conditions: analysis order: 24
weight coefficients for calculating delta: \( w(-1) = -0.5, w(0) = 0, w(1) = 0.5 \)
weight coefficients for calculating delta-delta: \( w(-1) = 0.25, w(0) = -0.5, w(1) = 0.25 \)

Note: The state sequence is determined according to the state duration densities of the HMMs. The algorithm is not included in SPTK.

\[
\text{mlpg} -m 24 -i 1 -d -0.5 0 0.5 -d 0.25 -0.5 0.25 \text{sample.pdf} > \text{sample.mcep}
\]

9.2 Plotting spectra calculated from generated mel-cepstrum

Files: sample.mcep: mel-cepstral coefficients (float)

Conditions: analysis order: 24
frequency warping parameter: \( \alpha = 0.42 \)
FFT size: 512 points
plotted frames: from 100-th to 250-th
sampling frequency: 16 kHz

\[
\text{bcut} +f -n 24 -s 100 -e 250 < \text{sample.mcep} \ |
\text{mgc2sp} -m 24 -a 0.42 -g 0 -l 512 | \text{grlogsp} -l 512 -x 8 -t | \text{xgr}
\]

9.3 Speech synthesis from the generated mel-cepstrum

Files: sample.pitch: pitch data generated from a sequence of MSD-HMMs included in this example (float, little endian)
sample.mcep: mel-cepstrum (float)
sample.mcep.syn: synthesized speech (float)

Conditions: frame period: 80 points (5 ms)
analysis order: 24
frequency warping parameter: \( \alpha = 0.42 \)

---

\( ^3 \)If you compiled SPTK with "--enable-double" option, please first convert this file into double format:
\[
x2x +sd \text{sample.pdf} > \text{sample.pdf.double}
\]

\( ^4 \)If you compiled SPTK with "--enable-double" option, please first convert this file into double format:
\[
x2x +sd \text{sample.pitch} > \text{sample.pitch.double}
\]
Note: The pitch pattern generation algorithm is not included in SPTK.

```
excite -p 80 sample.pitch | \
mlsadf -p 80 -a 0.42 -m 24 sample.mcep > sample.mcep.syn

gwave +f sample.mcep.syn | xgr
```

9.4 Check the given mean and variance vectors

Files: sample.pdf: sequence of mean and variance corresponding to a state sequence (float)

Conditions: analysis order: 24

9.4.1 Dump static feature vectors

```
bcp +f -l 150 -s 0 -e 24 sample.pdf | dmp -n 24 | less
```
9.4.2 Dump variance vectors of static feature vectors

\texttt{bcp +f -l 150 -s 75 -e 99 sample.pdf | sopr -INV | dmp -n 24 | less}

9.4.3 Dump dynamic feature vectors (delta)

\texttt{bcp +f -l 150 -s 25 -e 49 sample.pdf | dmp -n 24 | less}

9.4.4 Dump variance vectors of dynamic feature vectors (delta)

\texttt{bcp +f -l 150 -s 100 -e 124 sample.pdf | sopr -INV | dmp -n 24 | less}

9.5 Speech synthesis without dynamic feature

\textbf{Files:} sample.pitch: pitch data generated from a sequence of MSD-HMMs (float)
\hspace{1cm} sample.mcep.wo-dyn: mel-cepstrum generated without dynamic feature (float)
\hspace{1cm} sample.mcep.wo-dyn.syn: synthesized speech without dynamic feature (float)

\textbf{Conditions:} frame period: 80 points (5 ms)
\hspace{1cm} analysis order: 24
\hspace{1cm} frequency warping parameter: $\alpha = 0.42$

\texttt{bcp +f -l 150 -s 0 -e 24 sample.pdf > sample.mcep.wo-dyn}

\texttt{bcut +f -n 24 -s 100 -e 250 < sample.mcep.wo-dyn | mgc2sp -m 24 -a 0.42 -g 0 -l 512 | grlogsp -l 512 -x 8 -t | xgr}

\texttt{excite -p 80 sample.pitch | mlsadf -p 80 -a 0.42 -m 24 sample.mcep.wo-dyn | sample.mcep.wo-dyn.syn}

\texttt{gwave +f sample.mcep.wo-dyn.syn | xgr}
10 Voice Conversion based on GMM

Voice conversion from speaker maleA to speaker maleB

10.1 Minimum configuration of voice conversion

**Files:**
- \( s \): original speech signal spoken by maleA (short integer, 16 kHz sampling, little endian)
- \( t \): target speech signal spoken by maleB (short integer, 16 kHz sampling, little endian)
- \( f \): test speech signal spoken by maleA (short integer, 16 kHz sampling, little endian)
- \( c \): converted speech signal (float)

**Conditions:**
- frame length: 400 points (25ms)
- frame period: 80 points (5ms)
- window: Blackman window
- analysis order: 24

```bash
da +f -s 16 sample.mcep.wo-dyn.syn sample.mcep.syn
```
frequency warping parameter: $\alpha = 0.42$

the number of GMM mixture: 2

10.1.1 Training GMM

```
x2x +sf < source_maleA.raw | frame -l 400 -p 80 | window -l 400 -L 1024 | \ 
    mcep -l 1024 -m 24 -a 0.42 > source_maleA.mcep
x2x +sf < target_maleB.raw | frame -l 400 -p 80 | window -l 400 -L 1024 | \ 
    mcep -l 1024 -m 24 -a 0.42 > target_maleB.mcep
dtw -m 24 target_maleB.mcep < source_maleA.mcep | gmm -l 50 -m 2 -f > maleA_maleB.gmm
```

10.1.2 Voice conversion

```
x2x +sf < test_maleA.raw | frame -l 400 -p 80 | window -l 400 -L 1024 | \ 
    mcep -l 1024 -m 24 -a 0.42 > test_maleA.mcep
x2x +sf < test_maleA.raw | pitch -s 16 -p 80 > test_maleA.pitch
vc -n 24 -m 2 maleA_maleB.gmm < test_maleA.mcep > converted_maleB.mcep
excite -p 80 test_maleA.pitch | \ 
    mlsadf -m 24 -p 80 -a 0.42 converted_maleB.mcep > converted_maleB.syn
```

10.2 Voice conversion using iterative alignment

**Files:**
- `source` maleA.short: original speech signal spoken by maleA (short integer, 16 kHz sampling, little endian)
- `target` maleB.short: target speech signal spoken by maleB (short integer, 16 kHz sampling, little endian)
- `test` maleA.short: test speech signal spoken by maleA (short integer, 16 kHz sampling, little endian)
- `converted` maleB.syn: converted speech signal (float)

**Conditions:**
- frame length: 400 points (25ms)
- frame period: 80 points (5ms)
- window: Blackman window
- analysis order: 24
- sampling frequency: 16kHz
- frequency warping parameter: $\alpha = 0.42$
- the number of GMM mixture: 2

10.2.1 Training initial GMM

```
dtw -m 24 target_maleB.mcep < source_maleA.mcep > maleA_maleB_0.dtw
gmm -l 50 -m 2 -f < maleA_maleB_0.dtw > maleA_maleB_0.gmm
```

10.2.2 GMM estimation using iterative alignment

```
x2x +sf < source_maleA.raw | frame -l 400 -p 80 | window -l 400 -L 1024 | \ 
    mcep -l 1024 -m 24 -a 0.42 |
    vc -n 24 -m 2 maleA_maleB_0.gmm |
    dtw -m 24 target_maleB.mcep -v maleA_maleB.viterbi > /dev/null
dtw -m 24 -V maleA_maleB.viterbi target_maleB.mcep < source_maleA.mcep > maleA_maleB_1.dtw
    gmm -l 50 -m 2 -f < maleA_maleB_1.dtw > maleA_maleB_1.gmm
```
10.2.3 Voice conversion

```bash
vc -n 24 -m 2 maleA_maleB_1.gmm < test_maleA.mcep > converted_maleB_1.mcep
excite -p 80 test_maleA.pitch |
    mlsadfl -m 24 -p 80 -a 0.42 converted_maleB_1.mcep > converted_maleB_1.syn
```

11 Speaker Identification Based on GMM

Identification of speaker maleB from speaker maleA, maleB and maleC

**Files:**
- data_male(A,B,C).short: speech signal spoken by maleA,B and C (short integer, 16 kHz sampling, little endian)
- test_maleB.short: test speech signal spoken by maleB (short integer, 16 kHz sampling, little endian)

**Conditions:** order of mfcc: 12

11.1 GMM training

```bash
x2x +sf < data_maleA.short | frame | mfcc | gmm -l 12 > maleA.gmm
x2x +sf < data_maleB.short | frame | mfcc | gmm -l 12 > maleB.gmm
x2x +sf < data_maleC.short | frame | mfcc | gmm -l 12 > maleC.gmm
```

11.2 Speaker identification

```bash
x2x +sf < test_maleB.short | frame | mfcc > test_maleB.mfcc
gmmp -a -l 12 maleA.gmm test_maleB.mfcc > result_maleA.score
gmmp -a -l 12 maleB.gmm test_maleB.mfcc > result_maleB.score
gmmp -a -l 12 maleC.gmm test_maleB.mfcc > result_maleC.score
```

The recognized speaker’s score is the largest value for the test speech signal.