

Examples for Using Speech Signal Processing Toolkit

Ver. 3.6

SPTK working group

December 25, 2012

Contents

1	Basics	3
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	4
1.8	Cut a portion out of a file	4
2	Pitch Extraction from Speech Waveform	5
2.1	A pitch extractor	5
2.2	Plotting the extracted pitch contour	6
3	Speech Analysis/Synthesis Based on Mel-Cepstral Representation	6
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	9
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	14
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

6	Vector Quantization of Mel-Cepstrum	18
6.1	Train a (very small) Codebook	18
6.2	Encode (training vectors)	18
6.3	Decode (training vectors)	19
6.4	Plotting original and quantized spectra	19
6.5	Performance evaluation on the training data	20
6.6	Speech synthesis from quantized mel-cepstrum	21
7	Preparation of Speech Parameter for Speech Recognition	22
7.1	Cepstrum derived from LPC analysis (LPC cepstrum)	22
7.2	Mel-cepstrum derived from LPC analysis (LPC mel-cepstrum)	22
7.3	Mel-cepstrum obtained by mel-cepstral analysis	22
7.4	Mel-cepstrum derived from mel-generalized cepstral analysis	23
7.5	Plotting spectra for each speech recognition parameter	23
8	Playing with the Vocoder Based on Mel-Cepstrum	24
8.1	High- or low-pitched voice	24
8.2	Fast- or slow-speaking voice	25
8.3	Hoarse voice	25
8.4	Robotic voice	25
8.5	Child-like or deep voice	25
8.6	Various voices	25
9	Speech Synthesis Based on HMM	26
9.1	Speech parameter generation from a sequence of HMMs	26
9.2	Plotting spectra calculated from generated mel-cepstrum	26
9.3	Speech synthesis from the generated mel-cepstrum	26
9.4	Check the given mean and variance vectors	27
9.4.1	Dump static feature vectors	27
9.4.2	Dump variance vectors of static feature vectors	28
9.4.3	Dump dynamic feature vectors (delta)	28
9.4.4	Dump variance vectors of dynamic feature vectors (delta)	28
9.5	Speech synthesis without dynamic feature	28

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)^{1,2}

```
x2x +sf < data.short > data.float
```

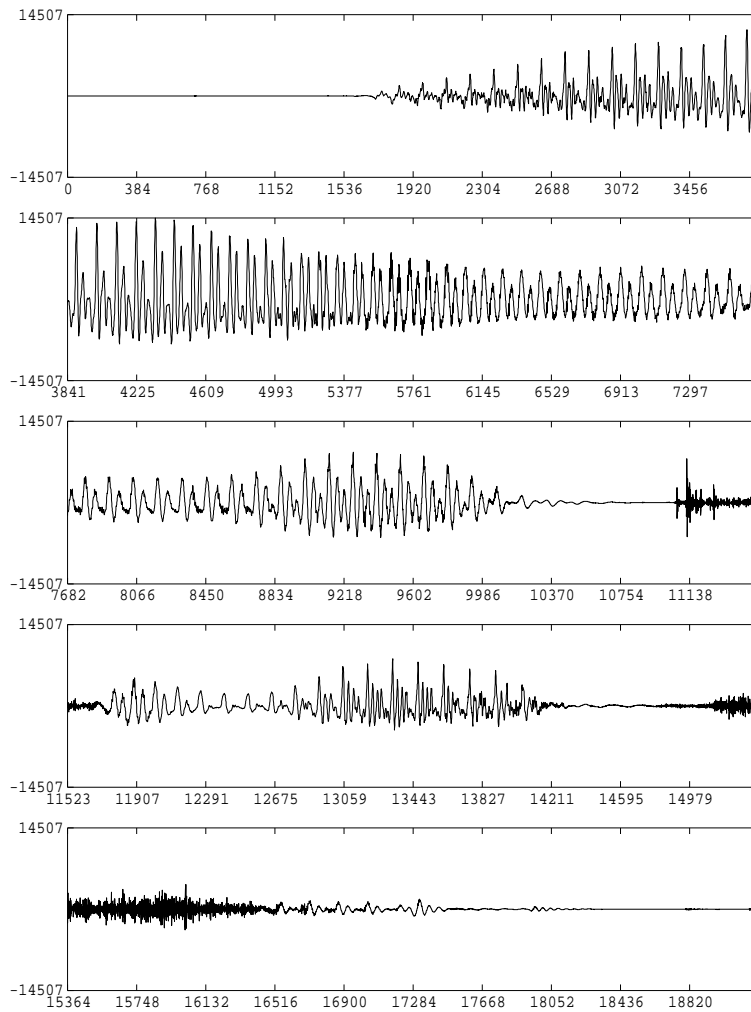
1.5 Plotting speech waveform on X-window

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

```
gwave +s data.short | xgr
```

¹By 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).

²If you compiled SPTK with “--enable-double” option, please use “+sd” option instead of “+sf” and “+d” option instead of “+f”.



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)

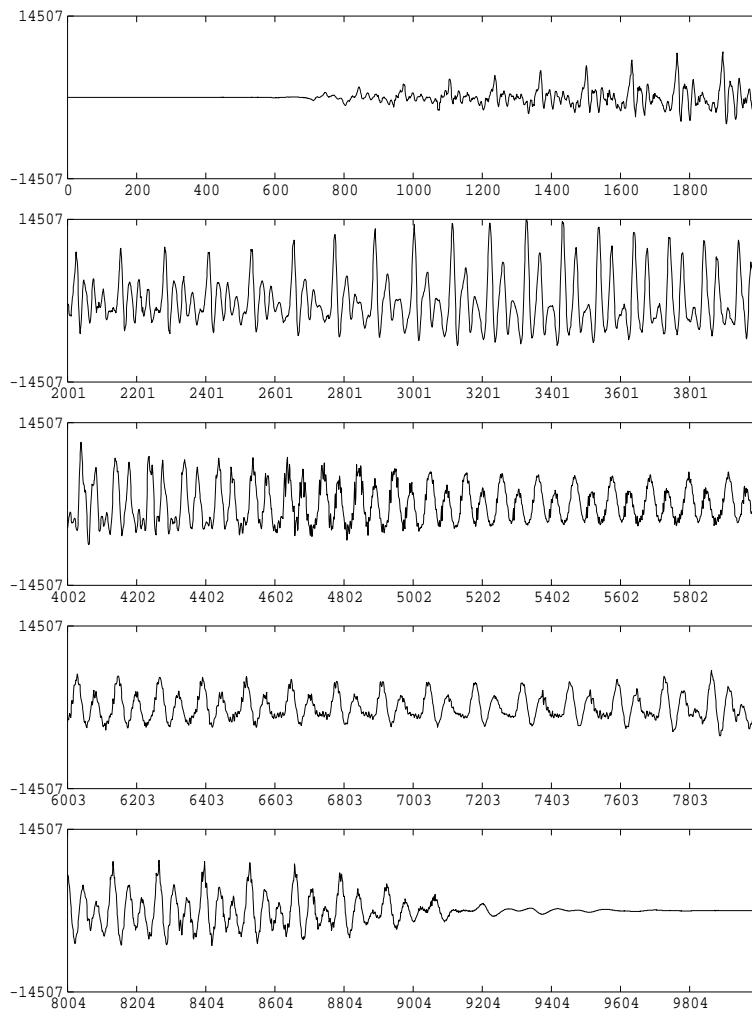
Note: This works only on Linux, Solaris, and FreeBSD.

```
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)

```
bcut +s -s 1000 -e 11000 < data.short |\
gwave +s | xgr
```



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.

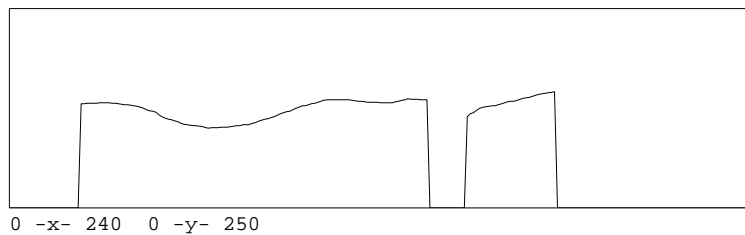
```
x2x +sf data.short | pitch -a 1 -s 16 -p 80 -L 80 -H 165 > data.pitch
```

2.2 Plotting the extracted pitch contour

Files: data.pitch: pitch data extracted from speech data "data.short"

Conditions: Minimum value of vertical axis: 0.0
Maximum value of vertical axis: 250.0
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

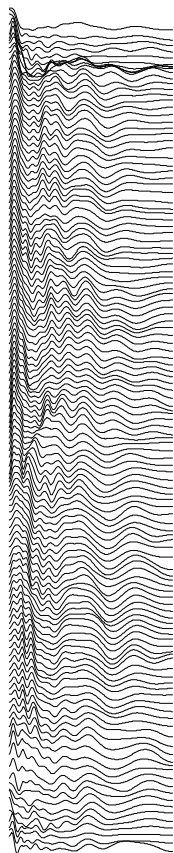
```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 |\
mcep -l 512 -m 20 -a 0.42 > data.mcep
```

3.2 Plotting spectral estimates from mel-cepstrum

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

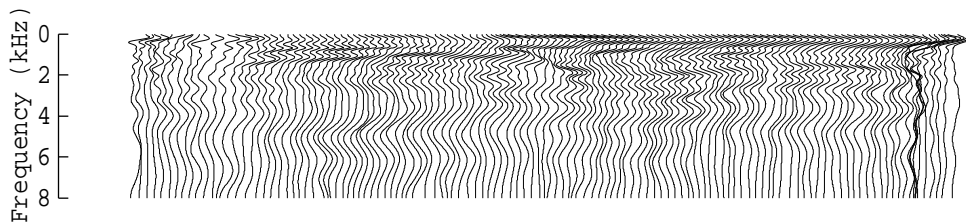
Conditions: analysis order: 20
frequency warping parameter: $\alpha = 0.42$
FFT size: 512 points
plotted frames: from 10-th to 135-th
sampling frequency: 16 kHz

```
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
```



0 2 4 6 8
Frequency (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 -t | xgr
```



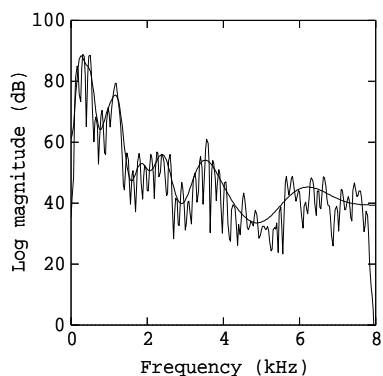
3.3 Plotting the spectral estimate with the FFT spectrum

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
```

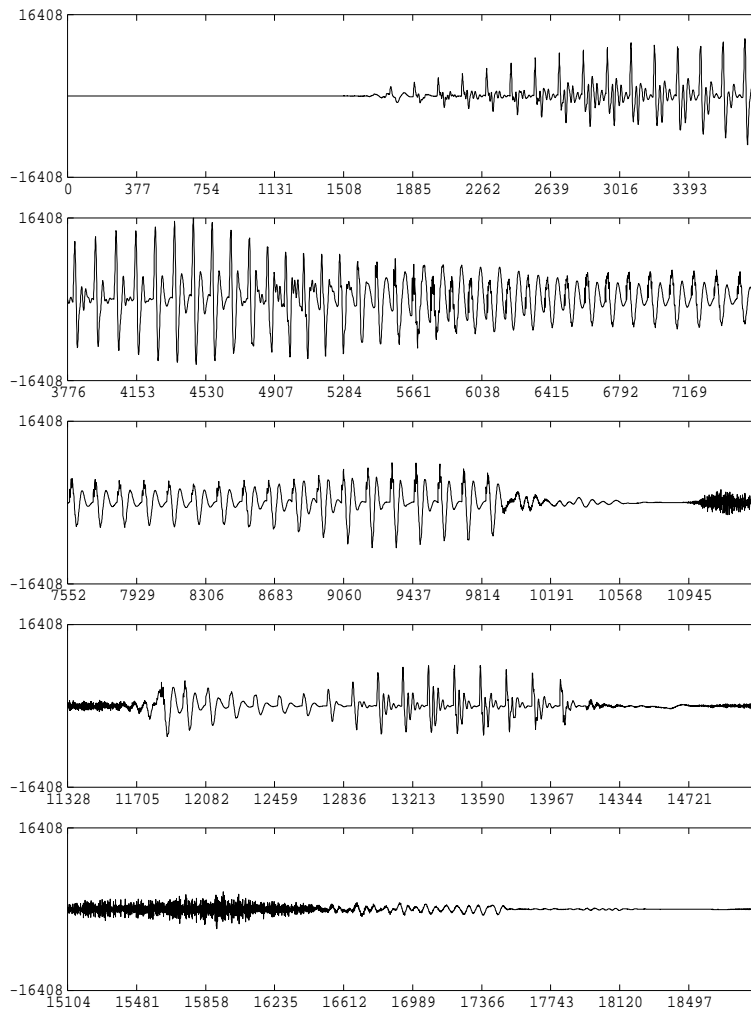


3.4 Speech synthesis from mel-cepstrum

Files: data.pitch: pitch data extracted from speech data "data.short"
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
```

```
da +f -s 16 data.mcep.syn
```

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

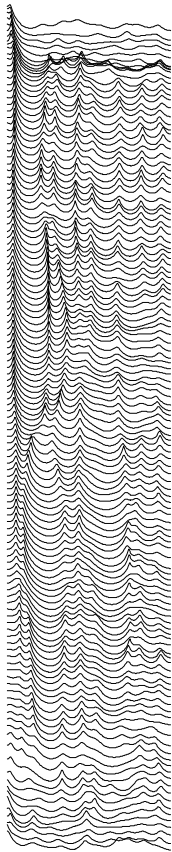
Files: data.lpc: LPC coefficients (float)

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
```



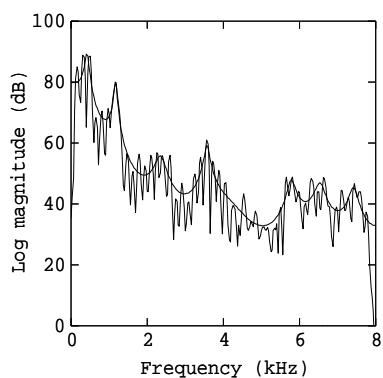
0 2 4 6 8
Frequency (kHz)

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

```
( 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.lpc > data.tmp ;\  
spec -l 512 -n 20 -p data.tmp | glogsp -l 512 -x 8 ;\  
\rm data.tmp ) | xgr
```

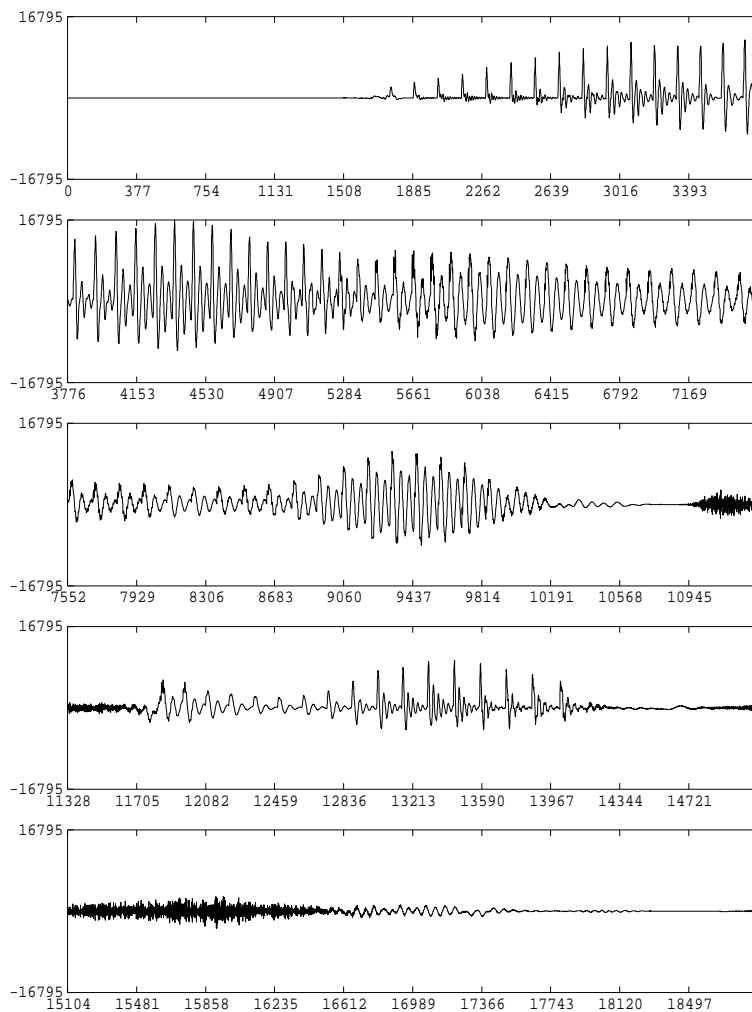


4.4 Speech synthesis from LPC coefficients

Files: data.pitch: pitch data extracted from speech data "data"
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

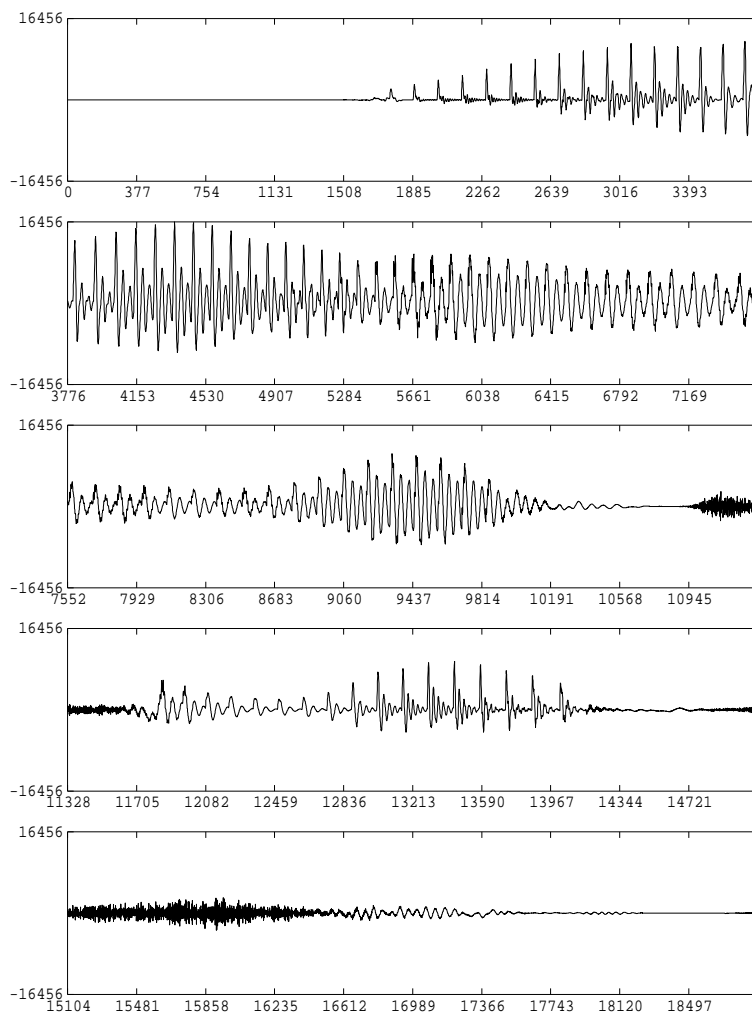
```
lpc2par -m 20 < data.lpc > data.par
```

4.6 Speech synthesis from PARCOR coefficients

Files: data.pitch: pitch data extracted from speech data "data" (float)
 data.par: PARCOR coefficients (float)
[data.par.syn](#): synthesized speech (float)

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

```
excite -p 80 data.pitch | ltcdf -m 20 -p 80 data.par > data.par.syn
gwave +f data.par.syn | xgr
```



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
```

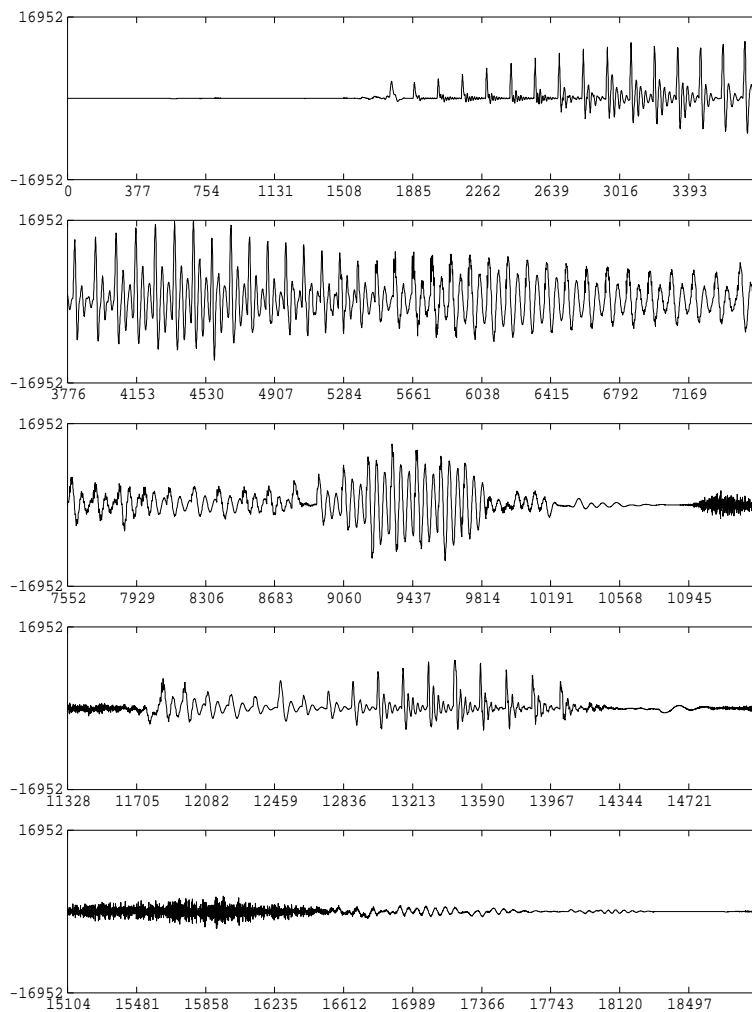
4.8 Speech synthesis from LSP coefficients

Files: data.pitch: pitch data extracted from speech data "data"
 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: speech data: data (short integer, 10 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: $\gamma = -1/2$

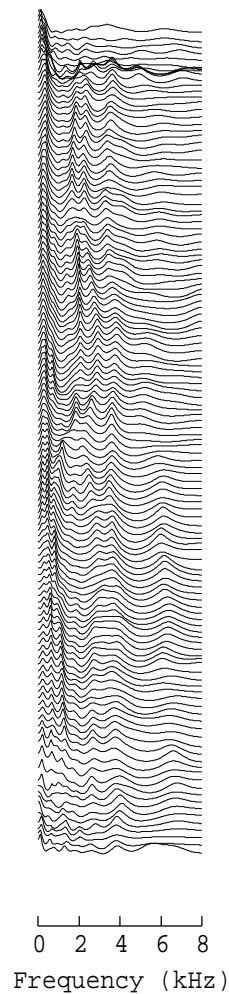
```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 |\
mgcep -m 20 -a 0.42 -c 2 -l 512 > data.mgcep
```

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

```
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
```



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 |\  

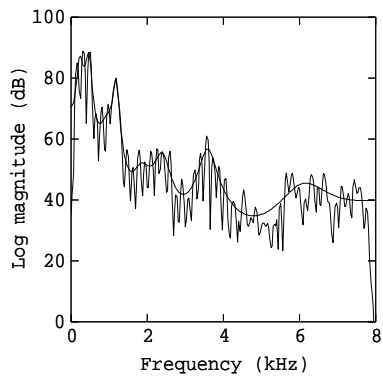
window -l 400 -L 512 | spec -l 512 |\  

glogsp -l 512 -x 8 -p 2 ;\  

\  

bcut +f -n 20 -s 65 -e 65 < data.mgcep |\  

mgc2sp -m 20 -a 0.42 -c 2 -l 512 | glogsp -l 512 -x 8 ) | xgr
```

5.4 Speech synthesis from mel-generalized cepstrum

Files: data.pitch: pitch data extracted from speech data "data"

data.mgcep: mel-generalized cepstrum (float)

[data.mgcep.syn](#): synthesized speech (float)

Conditions: frame period: 80 points (5 ms)

analysis order: 20

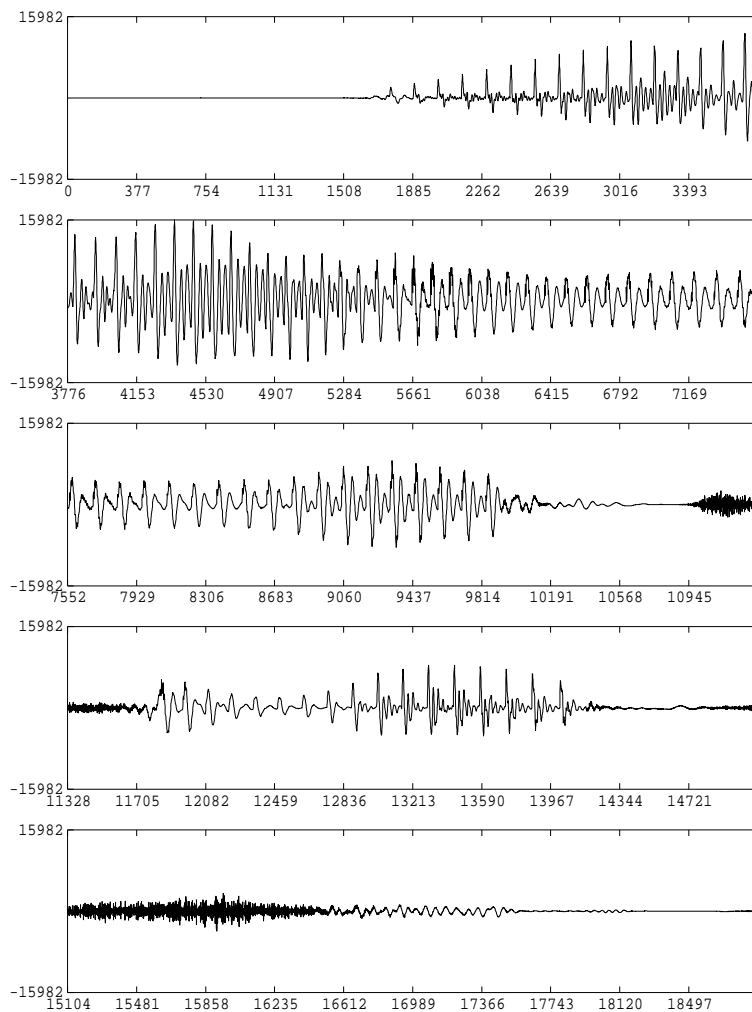
frequency warping parameter: $\alpha = 0.42$

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
```



```
da +f -s 16 data.mgcep.syn
```

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

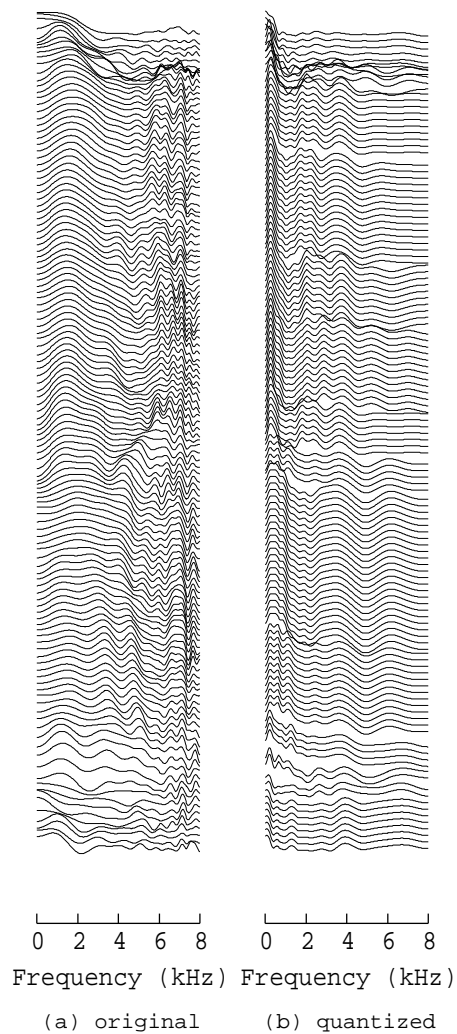
```
ivq -n 20 codebook.mcep < data.mcep.index > data.mcep.vq
```

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 |\
mgc2sp -m 20 -a 0.42 -g 0 -l 512 |\
grlogsp -l 512 -x 8 -o 1 -c "(a) original" ;\
\
bcut +f -n 20 -s 10 -e 135 < data.mcep.vq |\
mgc2sp -m 20 -a 0.42 -g 0 -l 512 |\
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.cdlist: 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
freqt -a 0.42 -m 20 -A 0 -M 255 < data.mcep.vq | \
cdist data.mcep.cep -m 255 > data.mcep.vq.cdlist
\rm data.mcep.cep
```

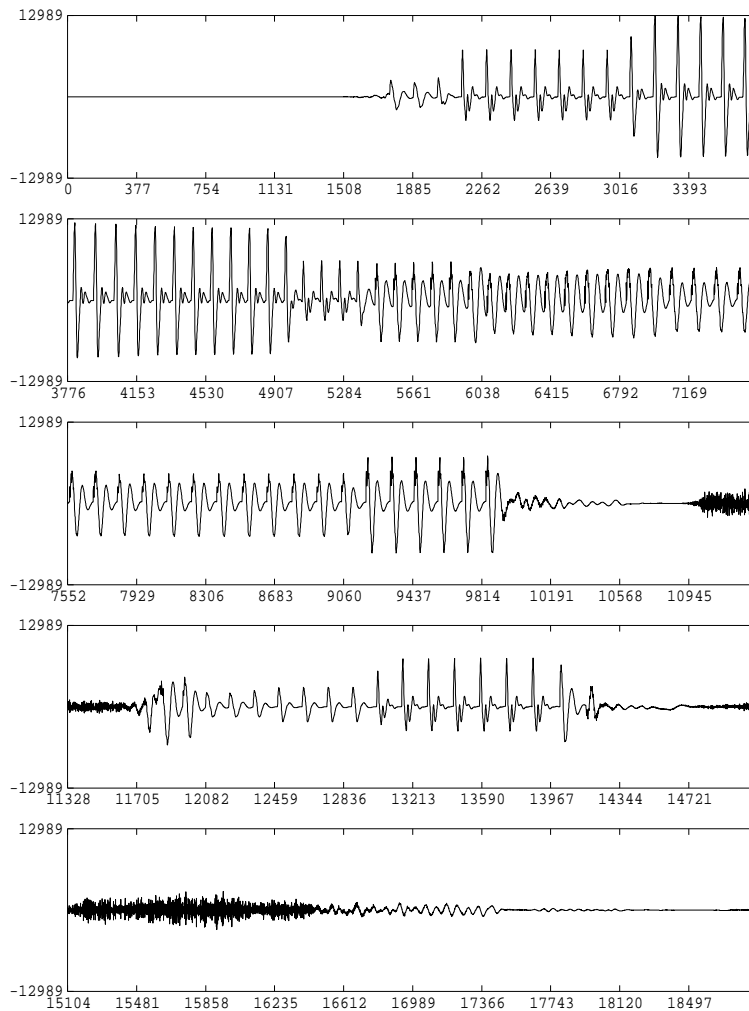
6.6 Speech synthesis from quantized mel-cepstrum

Files: data.pitch: pitch data extracted from speech data "data.short"
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 |\
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 |\
lpc -l 400 -m 12 |\
lpc2c -m 12 -M 256 |\
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 |\
lpc -l 400 -m 12 |\
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.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 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, 10 kHz sampling)

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

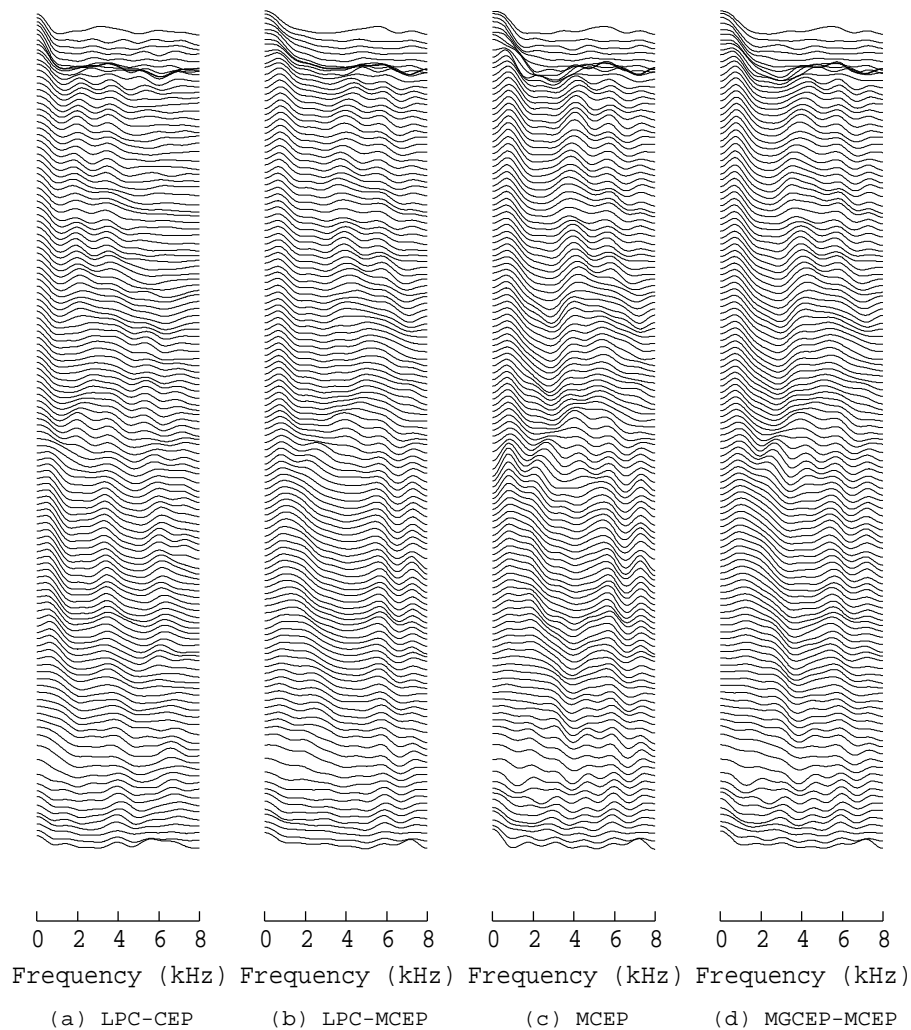
```
x2x +sf < data.short | frame -l 400 -p 80 | window -l 400 -L 512 |\
mgcep -m 12 -a 0.42 -c 2 -l 512 |\
mgc2mgc -m 12 -a 0.42 -c 2 -M 12 -A 0.42 -G 0 > data.mgcep.mcep
```

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 -O 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 -O 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 -O 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 -O 4 -c "(d) MGCEP-MCEP" ) | xgr
```



8 Playing with the Vocoder Based on Mel-Cepstrum

8.1 High- or low-pitched voice

Files: [data.mcep.high.syn](#): synthesized speech (float)

[data.mcep.low.syn](#): synthesized speech (float)

```
sopr -m 0.4 data.pitch |\
excite -p 80 | mlsadf -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 | mlsadf -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)³
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-3.2.

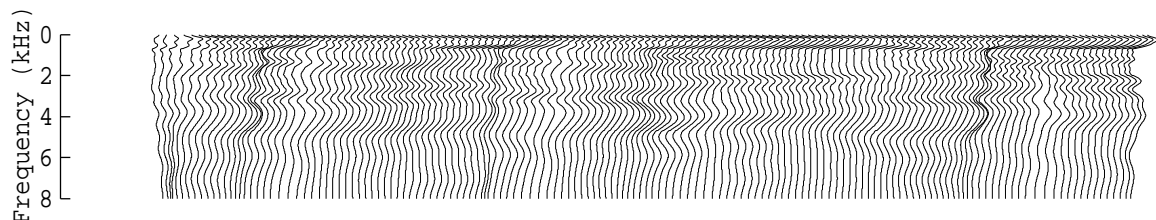
```
mlpg -m 24 -i 1 -d -0.5 0 0.5 -d 0.25 -0.5 0.25 sample.pdf > sample.mcep
```

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

```
bcut +f -n 24 -s 100 -e 250 < sample.mcep |\  
mgc2sp -m 24 -a 0.42 -g 0 -l 512 | grlogsp -l 512 -x 8 -t | 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)⁴
sample.mcep: mel-cepstrum (float)
[sample.mcep.syn](#): synthesized speech (float)

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

³If you compiled SPTK with "--enable-double" option, please first convert this file into double format:
`x2x +sd sample.pdf > sample.pdf.double`

⁴If you compiled SPTK with "--enable-double" option, please first convert this file into double format:
`x2x +sd sample.pitch > sample.pitch.double`

9.4.2 Dump variance vectors of static feature vectors

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

9.4.3 Dump dynamic feature vectors (delta)

```
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)

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

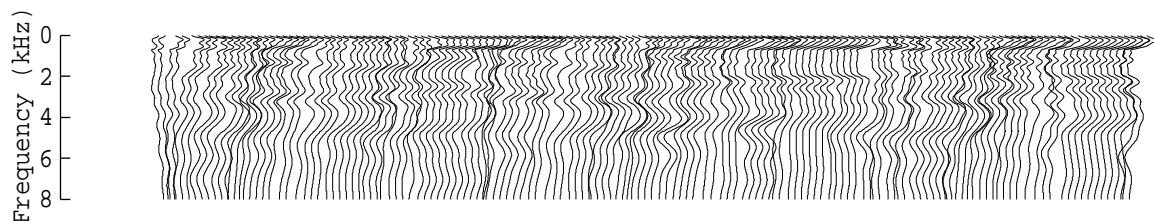
9.5 Speech synthesis without dynamic feature

Files: sample.pitch: pitch data generated from a sequence of MSD-HMMs (float)
sample.mcep.wo-dyn: mel-cepstrum generated without dynamic feature (float)
[sample.mcep.wo-dyn.syn](#): synthesized speech without dynamic feature (float)

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

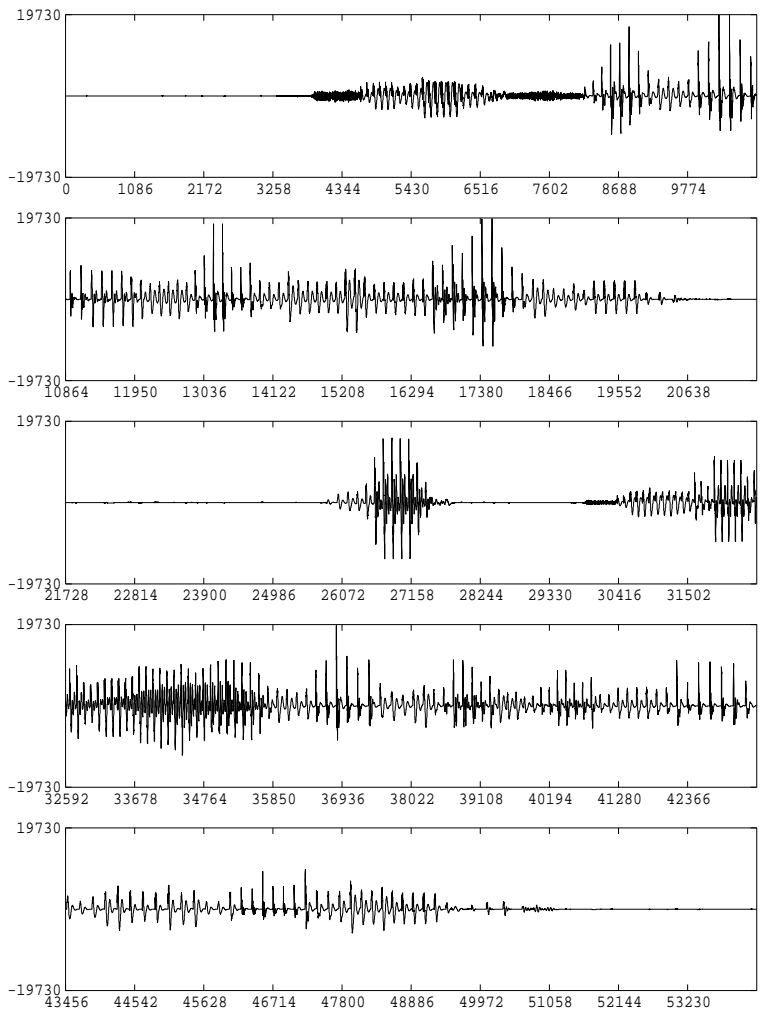
```
bcp +f -l 150 -s 0 -e 24 sample.pdf > sample.mcep.wo-dyn
```

```
bcut +f -n 24 -s 100 -e 250 < sample.mcep.wo-dyn |\nmgc2sp -m 24 -a 0.42 -g 0 -l 512 | grlogsp -l 512 -x 8 -t | xgr
```



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

```
gwave +f sample.mcep.wo-dyn.syn | xgr
```



da +f -s 16 sample.mcep.wo-dyn.syn sample.mcep.syn