

8 Sound eXchange - sox

Das open source `sox` Programm ist für alle Betriebssysteme erhältlich. Wir haben es schon als universellen Dateiformat-Konverter kennengelernt. `sox` kann aber sehr viel mehr: z.B. Filtern, re-sampling, Kanäle neu mischen oder zusammenfassen, alle möglichen Effekte anwenden etc.

Die Syntax des `sox` Kommandos ist ausführlich in der man page beschrieben. Im Allgemeinen sieht sie so aus:

```
sox [generell opt] [input opt] input [output opt] output [effect [effect opt]]
```

Beispiele:

Wir wollen ein NIST audio file rückwärts abspielen und in einem Microsoft wav file speichern:

```
sox -V -t sph signal.nis signal.wav reverse
```

-V : allgemeine Option 'verbose', die bewirkt, dass `sox` Informationen ausgibt

-t sph : input Option 'type' definiert den Formattyp für das Inputfile als NIST SHERE (`sox` erkennt die Extension 'nis' nicht!)

signal.nis : input File

signal.wav : output File (`sox` erkennt an der Extension 'wav', dass es sich um ein Microsoft wav File handelt und konvertiert automatisch)

reverse: Effekt reverse bewirkt Umdrehen des Signals

Wir wollen die Abtastrate von 22050Hz auf 16000 Hz senken (down sampling):

```
sox -V -t sph signal.nis -r 16000 signal.wav
```

-r 16000 : output Option Abtastrate bei 16000Hz

Wir wollen ein Stereo Sound-File ohne Header (raw sound file) in ein Mono Microsoft wav File umwandeln. Dazu müssen wir die Abtastrate, die Anzahl der Bits pro Abtastwert (Wortlänge) und die Kodierung der Abtastwerte (coding) kennen:

```
sox -V -t raw -r 22050 -2 -s -c 2 signal.raw -c 1 signal.wav
```

-t raw : input Option Formattyp ist raw

-r 22050 : input Option Abtastrate ist 22050Hz

-2 : input Option zwei Byte pro Abtastwert

-s : input Option signed integer, also -37676 ... +37677

-c 2 : input Option 2 Kanäle (Stereo)

signal.raw : input File

-c 1 : output Option 1 Kanal (Mono)

signal.wav : output File

Wir wollen ein Microsoft wav File mit zwei Kanälen so auftrennen, dass jeder Kanal in einem separaten File gespeichert ist (split):

```
sox -V stereo.wav -c 1 kanal_links.wav mixer 1,0
```

```
sox -V stereo.wav -c 1 kanal_rechts.wav mixer 0,1
```

-c 1 : output Option 1 Kanal

mixer : Effekt Mischer

1,0 : Option zu Effekt 'mixer': erster (linker) Kanal bleibt gleich, zweiter Kanal (rechts) wird unterdrückt (mit Null multipliziert)

0,1 : Option zu Effekt 'mixer': erster (linker) Kanal wird unterdrückt (mit Null multipliziert), zweiter Kanal (rechts) bleibt unverändert

9 Speech Filing System (SFS)

Das SFS Software-Paket ist eine Sammlung von vielen Programmen zur Manipulation von Sprachsignalen und Derivativen davon. Es ist stark UNIX orientiert, d.h. die meisten Kommandos werden von der Kommandozeile aus eingegeben. Eine (deutsche) Übersicht über alle Programme ist im Text-File `/share/local/sfs/command-overview.txt` zu finden. Zu jedem Kommando gibt es zusätzlich (in Englisch) eine ausführliche Man Page.

Im Gegensatz zu praat ist SFS kein monolithisches Programm mit einer interaktiven Oberfläche, sondern eine Sammlung von UNIX-Kommandos. Der Vorteil von SFS gegenüber praat ist, dass sich sehr einfach mächtige Skripten aus den verschiedenen SFS-Bausteinen zusammenbauen lassen und somit große Mengen von Daten automatisch verarbeitet werden können. Der Nachteil gegenüber praat ist, dass man nicht so einfach mit Signalen 'herumspielen' kann, d.h. einfach mal verschiedene Analysen ausprobieren und das Ergebnis sofort am Bildschirm sehen kann.

9.1 Übungen

Die folgenden angeleiteten Übungen sollen vor allem das Arbeiten mit SFS veranschaulichen.

1. Berechnen Sie das Grundfrequenz-Profil der Aufnahme `awe/awed5230.nis` des BAS Korpus PhonDat2.

Zunächst muss das File gefunden und auf den Rechner kopiert werden:

- Gehen Sie in Ihr Arbeitsverzeichnis
- `cip1 % cp /bmnt/BAS/PD2/data/awe/awed5230.nis .`

Extension '.nis' bedeutet ein File im NIST SPHERE Format. Testen Sie das Format und schauen Sie sich den Header an:

```
cip1 % test_nist v=1 p=1 awed5230.nis
```

Um mit SFS zu arbeiten, muss das Signal-File in ein sog. *SFS-Archive-File* umgewandelt werden (solche Files haben die Extension `.sfs`). Da es sich hier um ein NIST-File handelt (Extension `.nis`), können wir das Umwandelprogramm `nist2sfs` benutzen:

```
cip1 % nist2sfs awed5230.nis awed5230.sfs
cip1 % summary awed5230.sfs
```

Der SFS-Befehl `summary` gibt uns einen 'Überblick', was in dem Archiv-File für Spuren gespeichert sind. Hier sehen wir nur die eine Signalspur 1.01. Schauen Sie sich die Dateigröße des SFS-Files an. Warum ist das File kleiner als das PhonDat-File?

Zur Kontrolle schauen wir uns das so erzeugte Archiv-File `awed5230.sfs` am Bildschirm an. Dazu dient der Befehl `Es`:

```
cip1 % Es awed5230.sfs
```

Im allgemeinen muss man `Es` mitteilen, welche *Spuren* angezeigt werden sollen und aus welcher Spur es das Sonagramm berechnen soll. Spuren sind in SFS nach Typ und laufender Nummer geordnet. Der wichtigste Typ ist natürlich das Sprachsignal (Speech) und hat die Nummer '1' oder das Kürzel 'SP'. '1.01' oder 'SP.01' ist dann die erste Sprachspur in unserem Archiv-File.

```
cip1 % Es -i SP.01 awed5230.sfs
```

hat die gleiche Wirkung wie das obige Beispiel. Aber wenn in dem Archiv-File noch andere Spuren enthalten wären, würde 'Es' jetzt nur die erste Sprachspur zeigen. Die Option '-g <spur>' gibt an, dass aus der Spur <spur> ein Sonagramm berechnet und angezeigt werden soll.

```
cip1 % Es -i SP.01 -g SP.01 awed5230.sfs
```

Innerhalb von **Es** gibt es zahlreiche Kommandos zum Bewegen, Zoomen, Abhören, etc. Diese sind alle ausführlich in der Man Page beschrieben. Die wichtigsten sind:

- Linke Maustaste : setzt linke Grenze
- Rechte Maustaste : setzt rechte Grenze
- Mittlere Maustaste : spielt Signal zwischen Grenzen ab
- Z-Taste : Zoomed in Bereich zwischen Grenzen
- U-Taste : zeigt wieder gesamtes Signal (Un-Zoom)
- R-Taste : bewegt Signal nach rechts
- L-Taste : bewegt Signal nach links

Probieren Sie alle diese Funktionen jetzt an Ihrem Beispiel aus. Dann beenden Sie **Es**.

Das Kommando **fxac** berechnet den Verlauf der Grundfrequenz nach der Autokorrelationsmethode. Auch diesem Kommando (wie bei fast allen SFS-Kommandos) muss man mitteilen, aus welcher Spur es seinen Input nehmen soll:

```
cip1 % fxac -i SP.01 awed5230.sfs
...
cip1 % summary awed5230.sfs
...
```

Wir sehen in der Zusammenfassung, dass zu der Sprachspur eine vom Typ '4.01' oder 'FX.01' dazugekommen ist, eine Grundfrequenzspur. Diese enthält das Ergebnis der Berechnung. Wir wollen uns das jetzt ansehen.

```
cip1 % Es -i SP.01 -i FX.01 -g SP.01 awed5230.sfs
```

Drucken Sie das angezeigte Display, indem Sie auf den Button **hard** im Sub-Menu **file** drücken, **Es** verlassen und

```
cip1 % lphp dig.ps
```

eingeben.

Der Befehl **remove** entfernt eine Spur aus dem Archivfile. Entfernen Sie die FX.01 Spur, indem Sie

```
cip1 % remove -i FX.01 awed5230.sfs
cip1 % summary awed5230.sfs
```

eingeben.

2. Berechnen Sie die Formantlagen in `awed5230.nis`

Der Befehl `fmanal` berechnet über das Sprachsignal die Polstellen der Linearen Prädiktor Polynome (LPC) und speichert diese in einer Formatspur vom Typ '12' ab.

```
cip1 % fmanal -i SP.01 awed5230.sfs
cip1 % Es -i SP.01 -i 12.01 -g SP.01 awed5230.sfs
```

Die Polstellen werden nach Ihrer Ordnung (Zahl) und Frequenzlage (Y-Achse) zeitsynchron zum Sprachsignal dargestellt. Der Befehl `fmtrack` versucht aus diesen Daten, den Verlauf der ersten 3 Formanten zu finden (formant tracking). Das Ergebnis wird in einer Syntheseparameter-Spur vom Typ '7' abgelegt.

```
cip1 % fmtrack -i 12.01 awed5230.sfs
cip1 % summary awed5230.sfs
cip1 % Es -i SP.01 -g Sp.01 -i 12.01 -i 7.01 awed5230.sfs
```

Wenn Archivfiles viele Spuren enthalten, ist es oft sinnvoller das Display-Kommando `Es` in einem anderen Modus, dem 'tree modus', zu starten:

```
cip1 % Es -t awed5230.sfs
```

Dann zeigt `Es` zunächst in einem Baumdiagramm sämtliche enthaltenen Spuren und deren Geschichte an. In unseren Falle z.B. ist die Wurzel aller anderen Spuren sie Sprachspur `SP.01`, daraus leiten sich die Polstellen (`12.01`), und daraus wiederum die Formantverläufe (`7.01`) ab. Durch anklicken kann der Benutzer bestimmen, welche Spuren er überhaupt anzeigen lassen will (select).

3. Digitale Filter

Digitale Filter sind analog zu elektronischen Filtern Programme, die bestimmte Bereiche des Spektrums aus einem Signal herausfiltern können. Ein kommandozeilen-orientiertes Filterprogramm ist `genfilt`. (Zum Experimentieren können Sie auch das Programm `esfilt` verwenden; es zeigt auch gleich die Filterkurve an, ist aber für die Verwendung in Skripten nicht geeignet)

Man unterscheidet zwischen:

- Tiefpass : Filtert alles oberhalb der Filterfrequenz
- Hochpass : Filtert alles unterhalb der Filterfrequenz
- Bandpass : Filtert alles außerhalb der oberen und der unteren Filterfrequenz
- Stoppass/Sperrfilter : Filtert alles zwischen der oberen und der unteren Filterfrequenz

Wir wollen das Signal `awed5230.nis` auf Telefonqualität filtern. Telefonbandbreite ist von 300 bis 3300 Hz. Dazu benötigt man also einen Bandpass:

```
cip1 % genfilt -n 10 -l 3300 -h 300 awed5230.sfs
cip1 % Es -t awed5230.sfs
```

Probieren Sie auch noch andere Filterkonfigurationen aus.

Probieren Sie das Demo-Programm `esfilt` aus.

10 Übungen zu Sound Files und Speech Filing System

Mit Ihren bisherigen Kenntnissen, sollten Sie die nachfolgenden Übungen selbsttätig durchführen können. Arbeiten Sie jeweils in Zweier- oder Dreiergruppen zusammen - nicht allein.

Beachten Sie, dass alle SFS Kommandos in der Datei `/share/local/sfs/command-overview.txt` alphabetisch aufgelistet und beschrieben sind. Außerdem bekommen Sie zu jedem SFS Kommando eine Man-Page mit dem Befehl:

```
% man <command>
```

Gehen Sie zunächst in Ihr Working Directory.

10.1 Aufnahme

Nehmen Sie mit dem angeschlossenen Mikrophon 2 Signale auf: das eine enthalte eine Tonleiter (gesungen: 'Ah! Ah! ...'), das andere enthalte einen beliebigen Satz, aber rückwärts gesprochen. Z.B. sagen Sie statt 'Mein Name ist Anja Maier' 'Reiam Aina tsi eman niem'. Wählen Sie als Aufnahme-Parameter 16000 Hz Abtastrate, 16 Bit und Mono. Speichern Sie die Tonleiter im File 'ton.wav' und den Satz im File 'satz.wav' ab. Achten Sie auf eine ausreichende Aussteuerung; vermeiden Sie aber Übersteuerungen. (audioeditorui)

10.2 Satanische Sounds

Reversieren Sie das Signal in 'satz.wav' nach 'ztas.wav' und spielen Sie das Ergebnis ab. (sox, play)

10.3 Dateikonvertierung in SFS

Wandeln Sie Ihre Aufnahmen 'ton.wav' und 'satz.wav' in entsprechende SFS Archiv Files um (wav2sfs, Es)

10.4 Grundfrequenz

Berechnen Sie von Ihrer Tonleiter das Grundfrequenzprofil und zwar

- nach der Autokorrelationsmethode
- nach der cepstralen Methode

Vergleichen Sie die zwei Ergebnisse durch eine gemeinsame Darstellung am Bildschirm und drucken Sie diese aus. Hat die Oktave tatsächlich die doppelte Frequenz? (fxac, fxcep)

10.5 Synchronized Overlap And Add (SOLA)

Sie möchten Ihre Aufnahme 'satz' schneller abspielen, ohne dass sich dabei die Grundfrequenz ändert ('Plattenspielerereffekt'). Dazu dient die SOLA Methode, welche die einzelnen Glottisschwingungen ermittelt, teilweise herausschneidet (Beschleunigung) bzw. kopiert und einfügt (Verlangsamung).

Beschleunigen Sie zunächst Ihre Aufnahme um das eineinhalbfache.

Suchen Sie einen kurzen Vokal in Ihrer Äußerung und zoomen Sie in das Originalsignal hinein. Zählen die die Pitch-Perioden in diesem Vokal. Suchen Sie nun den gleichen Vokal in dem beschleunigten Signal und zählen Sie erneut. Vergleichen Sie die Ergebnisse. Was erwarten Sie?

Verlangsamen Sie jetzt Ihr Signal um das fünffache und schauen Sie wieder den gleichen kurzen Vokal an. (respeed)

10.6 Spektralanalyse

Sie möchten sich das Signal dieses Vokals genauer anschauen. Berechnen Sie

- das Spektrum
- die cepstrale Übertragungsfunktion des Ansatzrohres

für Ihren Vokal und drucken Sie beide aus. (anspect)

10.7 Energie

Berechnen Sie in Ihrem Signal 'satz' die mittlere Energie mit einem gleitenden Fenster

- der Länge 0.001 sec
- der Länge 0.01 sec
- der Länge 0.1 sec
- der Länge 1.0 sec

Verwenden Sie ein Hamming-Window. Stellen Sie die verschiedenen Energie- Profile am Bildschirm dar. Welche Fensterbreite eignet sich am besten zur Feststellung der Silbenkerne (Maxima)? (envelope)

Wieviele phonetische Silben enthält also Ihr Signal 'satz'? Stimmt das überein mit der Anzahl der Schreibsilben? Wenn nein, warum nicht?