



Author H.-Guenter Hirsch	Date 03/15/05	Version pA4
-----------------------------	------------------	----------------

F a N T - Filtering and Noise Adding Tool

This tool can be used to

- filter speech data with a frequency characteristic as defined by ITU for telephone equipment and/or
- add noise to speech recordings at a desired SNR (signal-to-noise ratio)

Speech and noise files containing audio data sampled at 8 or 16 kHz can be used. They must be headerless (RAW), with samples stored as 16-bit signed integers (SHORT) and with the bytes ordered as requested by the machine on which the software is running. Output audio files will be produced in this same format.

The software has been written in C. It has been compiled with the GNU C compiler. The functionality can be controlled by giving optional parameters on the command line when calling the program. To obtain a help message, call the program with only the **-h** option.

Set the **-u** flag if you wish to process 16 kHz data rather than 8 kHz data.

For 8 kHz data, 4 frequency characteristics are available for filtering the speech and noise signals. These characteristics have been given the abbreviations G.712, IRS, MIRS and P.341 by ITU. The optional parameter **-f xxx** with **xxx=g712** or **xxx=irs** or **xxx=mirs** or **xxx=p341** has to be set to enable the filtering of the speech and the noise signals with the selected characteristic. In case of 16 kHz data only the P.341 filtering is available because the other frequency characteristics have been defined for the telephone frequency range up to 4 kHz. If you want to apply them to 16 kHz data you could downsample the speech signals to 8 kHz first before filtering them with this program. The filtering is done with the corresponding routines from the ITU software library.

A noise signal can be added that has been sampled at the same sampling rate than the speech signal. The option **-n <filename>** has to be given where <filename> defines the name of the noise file including its directory path. Furthermore the SNR has to be defined as dB value with the optional parameter **-s <value of SNR>**, e.g. **-s 10** for adding the noise signal at a SNR of 10 dB. In most cases the noise signal will be longer than the speech signal. A noise segment of the desired length is cut out of the whole noise signal by randomly defining the starting point of this segment. If you want the result to be reproducible, the seed of the random generator can be set with the option **-r <value of seed>**, e.g. **-r 1000**. Otherwise the seed is derived from the current time.



Author H.-Guenter Hirsch	Date 03/15/05	Version pA4
-----------------------------	------------------	----------------

The level S of the speech signal is calculated with the corresponding software according to ITU recommendation P.56. This recommendation contains a voice activity detector to determine a so called active speech level only from the speech segments of a recording ignoring pauses before and after a speech utterance and long pauses between words. The level N of the noise segment (that has been randomly selected out of the whole signal) is calculated as RMS value. As default the speech and noise signals are filtered beforehand with a G.712 characteristic so that the levels are calculated over the frequency range from about 300 to 3400 Hz. This filtering is only applied for calculating S and N. It has nothing to do with the filtering defined with the `-f` option. This preceding filtering has been applied for generating the noisy data of the Aurora-2 and the Aurora-4 databases.

The program allows the determination of S and N without filtering with the G.712 characteristic. If the option `-m snr_4khz` is given the levels are calculated over the whole frequency range from 0 to 4 kHz. In case of 16 kHz data the signals are filtered with an appropriate low-pass filter. Furthermore the levels can be calculated over the whole frequency range from 0 to 8 kHz in case of processing data sampled at 16 kHz by giving the command line option `-m snr_8khz`. A further optional mode `-m a_weight` has been introduced to enable the weighting with an A-weighting filter characteristic as defined as ANSI or IEC standard. The A-weighting characteristic has been derived from human sound perception. It is applied instead of the G.712 characteristic.

You should be aware of the importance of calculating the levels S and N with or without filtering! If you want to add e.g. a noise signal that mainly has its energy in the low frequency region you can get very different results for the different cases. The calculated level N will be higher without the G.712 filtering or the A weighting. So, the resulting factor for adding the noise to the speech at the desired SNR will be much lower. The amount of noise is lower in comparison to filtering the data with the G.712 characteristic beforehand. In case of applying these noisy data for speech recognition experiments the recognition results will be better in the case of calculating S and N without filtering also because a lot of feature extraction schemes do not analyze the low-frequency region. But the results might not be representative for the conditions of real acoustic scenarios where the SNR is most often estimated with some type of preceding filtering.

If the flag `-d` is given, DC offsets will be removed from the speech and noise signals when calculating S and N. This could be useful when applying the `-m snr_4khz` or the `-m snr_8khz` options, in order to prevent bias in the SNR calculation due to a DC offset in the speech or noise signal.



Author H.-Guenter Hirsch	Date 03/15/05	Version pA4
-----------------------------	------------------	----------------

The program has been designed to process a whole set of speech signals, e.g. all signals of a database. The name of a list file (that contains the names of all speech files including their paths) has to be given as parameter with the option **-i <name of list file>**. The name of a second list file containing the same number of entries has to be given with the option **-o <name of list file>** to define the names of the output files including their paths.

A further option is available to add the noise signal to all speech signals not at a fixed SNR but staying within a certain range of SNR. The range can be defined with the already mentioned command line option **-s <value of SNR>** and the option **-w <value of SNR range>**. Then the SNR will be randomly defined for each signal between the value given with the **-s** option and the sum of values given with the **-s** and the **-w** options. E.g. if the options **-s 5 -w 10** are given, the noise signal will be added at SNRs between 5 and 15 dB. A random generator with uniform distribution is applied so that the average SNR will be in the middle of the SNR range, e.g. in the above example the average SNR will be 10 dB.

The level of the speech signal can be normalized to a desired level with the option **-l <value of S>**. The value for the speech level refers to the measurement according to ITU recommendation P.56. A typical value would be **-20**. In this case the maximum amplitudes of the output signal would be approximately 2/3 of the total 16 Bit amplitude range.

A log file will be created containing some information about the details of the applied processing. The name of the log file can be given with the option **-e <filename>**. Otherwise a file “filter_add_noise.log” will be created.

A further special option is available for the case that different parties at different places want to create the same set of noisy speech data independently of each other. If it can not be presumed that the random generator works exactly the same on different machines or different operating systems this option can be applied to avoid the use of the random generator for defining the starting points of the noise segments. The name of a file can be given with the option **-a <name of index file>**. This file contains the starting sample indices as ASCII values. The number of indices has to be exactly the same as the number of files to be processed. A short program “create_list.c” exists that can be used to create such a list. Of course, it has to be ensured that all parties work on the same



Author H.-Guenter Hirsch	Date 03/15/05	Version pA4
-----------------------------	------------------	----------------

speech and noise data with exactly the same list of speech files where the files are listed in the same order.

Installation

After doing a *gunzip* on the file “fant.tar.gz” the source files can be extracted out of the archive with *tar -xvf fant.tar*.

The following files should be available then:

filter_add_noise.c create_list.c

filter_add_noise.make create_list.make

ITU software modules: cascg712.c fir-flat.c fir-hp.c fir-irs.c fir-lib.c fir-wb.c iir-lib.c
sv-p56.c ugst-utl.c firflt.h iirflt.h sv-p56.h ugst-utl.h

4 files in the subdirectory example: 57353.raw subway.raw (speech and noise file in raw short format with little-endian byte order → swap bytes in case your machine uses big endian!)
in.list out.list

The software can be compiled and linked with *make -f filter_add_noise.make*

You get a usage by calling *filter_add_noise -h*

A speech and a noise file (raw short format with little endian byte order) can be found in the subdirectory *example*. You can create e.g. a filtered and noisy version of the speech file by calling *filter_add_noise -i example/in.list -o example/out.list -n example/subway.raw -f g712 -s 10 -r 2000 -e fant.log*

Contact

We did some testing for applying this tool with different settings of the options. But, given the large variety of combinations, there might be combinations where the code is not free of errors. Please, feel free to contact the author by email hans-guenter.hirsch@hsnr.de in case of any problem.

In the meantime we developed a further tool for simulating not only the influence of additive background noise and filter characteristics but also of a handsfree speech input in rooms and of transmitting the speech over a cellular telephone network. A Web interface has been set up to experience this tool also acoustically: <http://dnt.kr.hsnr.de/sireac.html>