Master audio recordings for intelligibility

Pierre-Yves Mutrux, December 2017

Introduction:

Being heard is good. Being understood is better!

In order to be heard, you want to “come thru”. That can be achieved by controlling the **Loudness** of your audio file and make sure it is consistent all the way to the end.

If you want to be **understood**, some attention should be put unto more details. It starts with the acoustics characteristics of your recording room, followed by a good choice of equipment, room setup, microphone placement, etc.

Intelligibility should be given attention all along the production chain. Once the product is delivered, it is more difficult to make adjustments. This is where mastering comes into play. But remember, a bad recording can only get “less bad”.

Can you say In-te-lli-gi-bi-li-ty?

The capability of a spoken message to be understood is called intelligibility. There are several important elements for a recording to be intelligible and it doesn’t take much to spoil it! If intelligibility is not given enough attention from the beginning on (even before the recording starts actually), it is challenging to improve it later. A broadcaster or distributor is regularly facing that challenge, especially with legacy recordings, often dubbed from old analog media.

In non-tonal languages, consonants are very important for intelligibility and they are found above 500 Hz, typically in the 2 kHz – 4 kHz range. Vowels, usually found below 500 Hz, are also part of the message but play a less important role.

The table on the right shows the importance of every ⅓ octave frequency band for intelligibility.

We can see that the frequency range from 800 Hz to 4 kHz is the most important and that frequencies below 300 Hz have a very minimal contribution.

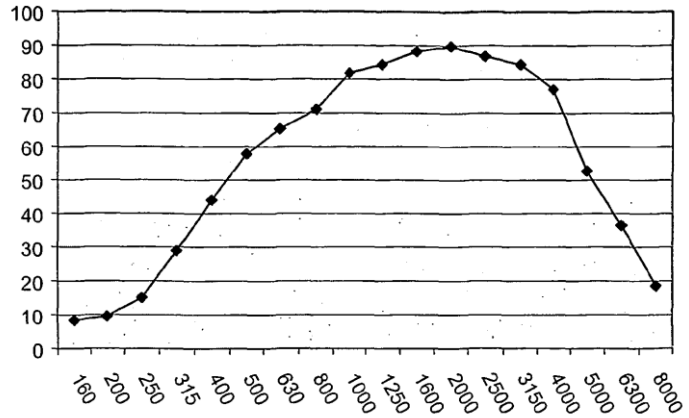
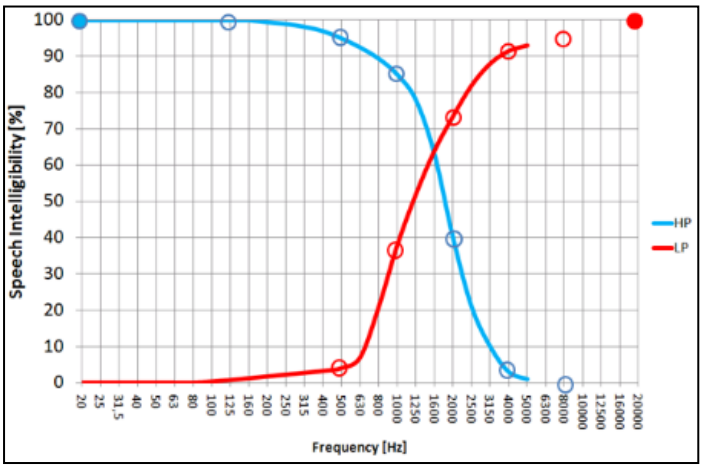


Table 2

Values of the articulation index importance function for average speech at 1/3-octave center frequencies, taken from Pavlovic, 1987. The sum of the AI importance values is 1000.

The following graph (source) shows the loss of intelligibility when frequencies are cut away, starting from the low frequencies (High pass, blue trace) or starting from the high frequencies (Low pass, red trace).



This leads to the first criteria for intelligibility: **Frequencies between 500 Hz and 4 kHz are essential for intelligibility**. Frequencies outside this range, if cutaway, would not significantly impair intelligibility.

The next criteria applies to several different aspects or audio reproduction but it all comes down to the same thing: the **Signal-to-Noise Ratio** (SNR).

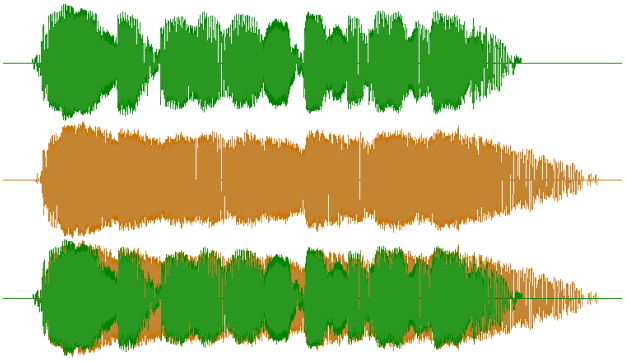
The *Signal* part of SNR is the "raison d'être" of every recording so there is no need to explain it here. But what about *Noise*?

Noise is basically signal that is not part of the main one, or put even simpler: **noise is undesired sound**. Even background music *can* enters this category when it comes to intelligibility!

Let's look at a few specific noise sources and ways to mitigate them.

- Thermal noise, hiss: every microphone, every sound card or portable recorder generates a certain amount of internal noise. The amount of it is inversely proportional to the quality of the electronic circuitry. The less quality, the more noise and vice versa. This is something to consider when (or before) acquiring equipment.
- Electric interferences: Laptops and in particular power supplies are big sources of electromagnetic waves (not the only one though) that shouldn't be picked up together with the desired audio. Microphones and audio cables should be shielded and run a balanced signal to avoid picking these electromagnetic interferences.
- Room acoustic: When recording, it is important to capture the voice only, no echo, no reverberation. Some reverberation, however, is needed for the person speaking into the microphone but too much of it is detrimental to the intelligibility. The trick for having a good acoustic in the recording space is to have equal absorption (or Reverberation Time) at every frequency and avoid flat and flutter echo. Early reflections are basically ok but late reflection (>40ms) are perceived as echo and are to be avoided.

The picture on the right shows a "dry" recording (first waveform, green), the same recording in a reverberant room (middle waveform, orange) and then the superposition of both (bottom). The difference between the first and second waveform is the reverberation, the undesired sound, in one word: the noise.



- Room acoustic at playback: This is a widely overlooked matter. It starts in recording facilities! When you playback a recording, unless you listen to it with headphones, it is played back in a reverberant space. These listening or editing spaces often have no or insufficient acoustic treatment. What you hear is the combination of the message, the recorded reverberation and the acoustic of your **current listening room**. When the same recording is played back at the "customer's" location, the room acoustic is in most cases more reverberant than at the recording facility, adding some unknown reverberation which can significantly damage intelligibility, especially in the low-frequency range where resonance is difficult to control (room modes). A good practice is to **roll-off low frequencies** which are anyway not essential for intelligibility. Many people believe that a boomy sound is nice. You are the judge for that but fact is that intelligibility is not enhanced at all, at the contrary, it is clearly diminished, whether you like it or not! As a remedy, I'm suggesting a 3rd or 4th order (18 or 24 dB/octave) **high pass filter** with cut off frequency around 150 Hz.
- Saturation or clipping is distorting the signal which, in real terms, means adding frequencies that do not belong to the original sound, i.e. unwanted sound: noise.
- Lossy encoding is based on removing some components from the original sound in order to save disk space. It is unnoticeable (transparent) until it is overdone for the sake of reducing file size or data transmission bandwidth, then some audible artifacts are introduced which is again unwanted sounds, i.e. noise. And that is at the cost of intelligibility.
- Probably the most obvious way to have a high SNR across the whole recording is to be on to of Loudness. This relatively new measurement unit allows for an unprecedented consistency of perceived volume at any time when used together with the Short-term Loudness and Loudness Range measurements.
- If a higher level is necessary, dynamic compression can be applied but that means introducing non-linearity, which is adding harmonics! Care must also be taken to not amplify noise in the process.

When the listening conditions are not known, the best thing to do is to keep the signal to a consistent "high enough" level (yours to define) and avoid introducing any form of noise in the recording or during playback.

The WosFiltr

There are situations where the intelligibility of a delivered file needs improvement. Situations where it is not possible to go back to the studio and start the recording over again. The only way is to **master** the file to the best of your possibilities. And that may well be true for a whole series so you may want to apply the corrections somewhat automatically.

To help you doing that, I worked on a filter chain for FFMpeg. I called it the "WosFiltr".

"Wos" comes from the German word "Was", which means "What" but with a specific accent…

Filtr is a funny way of writing filter. Why not?

It looks like this:

```
ffmpeg -i <input file> -af
"lowpass=f=<lp-freq>, lowpass =f=< lp-freq>,
highpass=f=<hp-freq>, highpass =f=<hp-req>,
compand=.1:.2:-900/-900 -51/-900 -50/-50:.01:0:-90:.1,
dynaudnorm=f=200:g=15:p=0.2:m=10:r=1:c=1:b=1,
bass=g=-4:f=1500:width_type=q:width=0.8,
loudnorm=I=<Loudness>:TP=<TruePeak>:LRA=<LRA>:dual_mono=true:print_format=summary"
-ar <sample rate> <output file>

Where
<input file>        path/name of the file to be processed
<lp-freq>           low pass frequency, recommended: 5kHz (write 5k or 5000)
<hp-freq>           high pass frequency, recommended: 150Hz (150)
<Loudness>          desired output Integrated Loudness, recommended -16LUFS (-16)
<TruePeak>          desired maximum True Peak, recommended -1dBTP (-1)
<LRA>               desired Loudness Range, recommended 8 (8)
<sample rate>       output sample rate, default is 192kHz (44100, 48k, …)
<output file>       output file path/name
```

And this is what it does:

Band pass filtering

> A stack of two 2nd order (12 dB/octave) high pass filters and two 2nd order low pass filters. The recommended values (150 Hz – 5 kHz) are wider than the minimum required for intelligibility. That results in a more natural sounding output then the minimum required for intelligibility, yet avoiding a boomy sound and too much sibilance.

Noise Gate with Compand

> If there is too much background noise, it is cut out when below 50 dB.

DynAudNorm

> A tool developed by MuldeR, now part of FFMpeg. Here an abstract of the documentation: *"The Dynamic Audio Normalizer will "even out" the volume of quiet and loud sections, in the sense that the volume of each section is brought to the same target level. Note, however, that the Dynamic Audio Normalizer achieves this goal *without* applying "dynamic range compressing". It will retain 100% of the dynamic range *within* each section of the audio file."*

Shelf filter (bass)

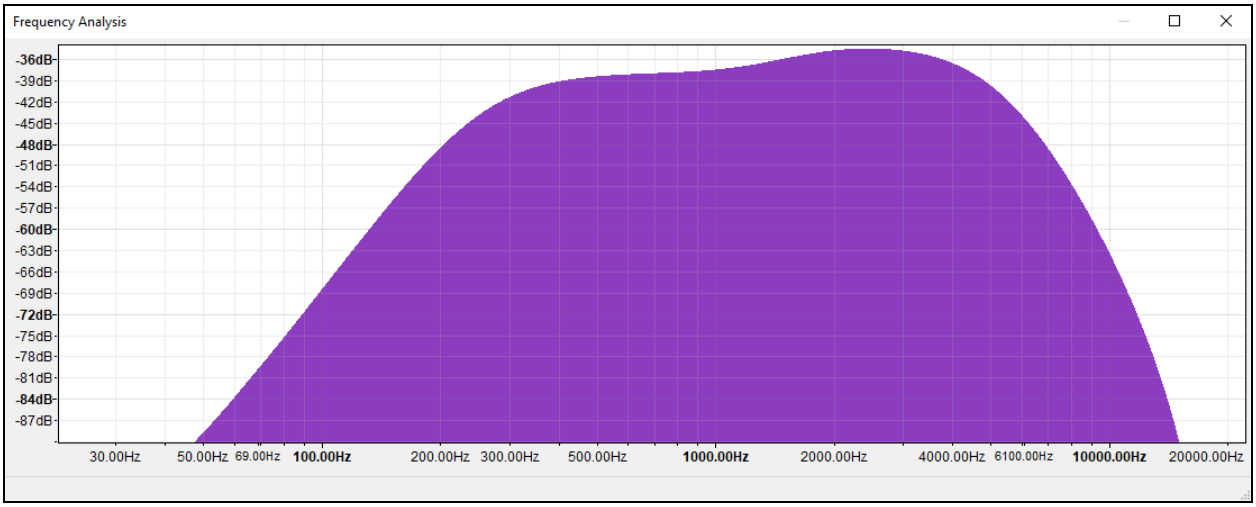
> A mere 4 dB boost above 1.5 kHz, to give a bit more emphasis to the most important part of the spectrum in term of intelligibility.

LoudNorm

> A tool developed by k.ylo.ph and now also part of FFMpeg. "*This algorithm can target Integrated Loudness (IL), Loudness Range (LRA), and maximum True Peak (TP)*".

My recommended values are -16 LUFS for IL, 8 LU for the LRA and -1 dBTP for TP. The output file statistics are printed in the console (and StdErr) in case you need them for further use (documentation or subsequent processing). The “dual_mono” is turned on to take care of the fact that a mono file is being used twice when played back on stereo systems (for an IL set to -16 LUFS, a mono file will actually be processed to -19 LUFS).

Here is a response curve for the bandpass and shelf filters combined:



The WosFiltr Batch file (droplet) (Windows only)

The simplest implementation of the WosFiltr is in a Batch file. If you are not too familiar with DOS, you can try my “*FFMpeg WosFiltr Droplet.bat*” you will find on muxson.com/WosFiltr for free.

It allows to drag-and-drop a file (only one at a time) onto it (hence the name Droplet) and it will apply the WosFiltr to your file. The output file format will be guessed from the input file format, except for the sampling frequency that will be set to a fixed value (44.1 kHz by default).

You can change every parameters to your liking.

Here is the content of the Batch file. You can make you own droplet simply by copying the following code and pasting it into a text editor (such as Notepad or Notepad++) and save it as “<whatever you want>.bat”.

You still need the latest version of FFMpeg stored in the same folder as the droplet and you’re done…

Good luck!

```
:: Retaining local path (CMD does CD to the droped file location)
SET mypath=%~dp0

:: ><> Parameters <>< ::

:: Low Cut in Hz (e.g. 150)
SET LC=150
:: High Cut in Hz (e.g. 5k or 5000)
SET HC=5k

:: Integrated Loudness in LUFS (e.g. -16) (Stadard value for stereo files.
Mono files will be automatically adjusted to -19LUFS becaue when they are
played back on stereo systems, the signal is used twice (L&R) which increases
the percieved loudness by 3dB)
SET LK=-16

:: Noise Gate threshold in dBFS (to avoid amplifying background noise in
"long" silences)
SET NG=-50

:: Sampling rate (e.g. 44100)(needed because LoudNorm works at 192kHz and
does not down-samples back ot the original sample frequency)
SET SR=44100

:: Suffix
SET SUF=WosFiltr
:: Detailed suffix (for debugging)
:: SET SUF=WosFiltr_%LC%-%HC%_shelf_NoiseGate%NG%_%LK%LUFS

:: Some math (not a parameter)
SET /a Val=%NG%-1

:: Band Pass Filter (18dB/Octave), Compand (Noise Gate), DynAudNorm, Bass
Shelf, LoudNorm, export LoudNorm stats in console
"%mypath%ffmpeg.exe" -y -i %1 -af "lowpass=f=%HC%,low-
pass=f=%HC%,highpass=f=%LC%,highpass=f=%LC%,compand=.1:.2:-900/-900 %Val%/-
900 %NG%/%NG%:.01:0:-90:.1,dynaud-
norm=f=200:g=15:p=0.2:m=10:r=1:c=1:b=1,bass=g=-
4:f=1500:width_type=q:width=0.8,loudnorm=I=%LK%:TP=-
1.5:LRA=6:dual_mono=true:print_format=summary" -ar %SR% "%~dpn1_%SUF%%~x1"

IF %errorlevel%==0 Exit
PAUSE
```

Please note: the information provided here comes with no guarantee of fitness to any private or commercial application. Use at your own risk.

December 2017, muxson@gmail.com