Improving the intelligibility of SSB transmissions

What is important when processing speech?

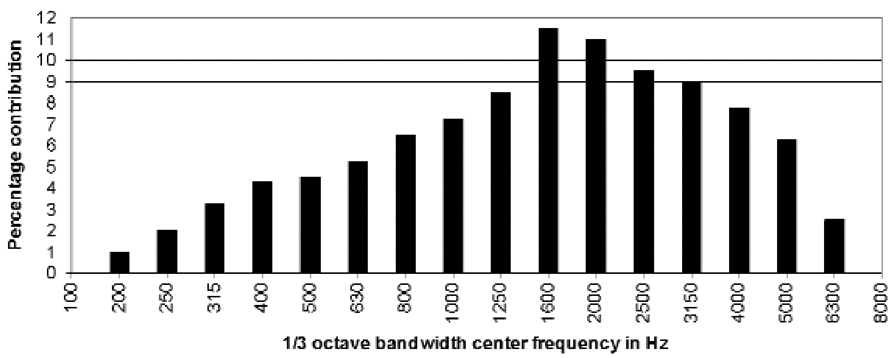

FIGURE 1: Some frequency bands have a remarkable impact on speech intelligibility.

WORDS FAIL ME. I've always had problems when using SSB. My voice and single sideband just don't seem to get along, and I've had to modify just about every radio I've ever used in order to make it sound reasonable. I'm not interested in hi-fi sound: I just like good communication quality and the ability to occasionally break a DX pileup. In the first half of this two-part article, I'll take a look at the factors that influence transmission quality, and part two is a practical speech processor based on my findings.

I recently bought a transceiver that is a great performer, except for one thing: the transmit audio. I kept on getting poor reports, usually stating that my voice sounded far too bassy and that it was difficult to tell what I saying, especially when my signal was weak. Unfortunately, this was not the first time I'd encountered these sorts of problems, so I decided to try and better understand what was happening so that I could fix it. This lead me through some very interesting research, before I finally found a solution that worked for me.

WHAT GOES WRONG? In order to convey information accurately by speech, we need to have a transmission system that has a sufficiently wide frequency response that it does not remove important components from the voice. For perfect error-free communication a frequency range of 80 to 8000Hz is generally considered to be adequate, but as the available speech bandwidth is decreased it becomes more difficult to understand what is being said. My voice is very bassy: it has a lot of energy in the 200 to 400Hz frequency range and this, coupled with my 'Mancunian' pronunciation of vowels, doesn't aid communication.

In Western languages, human speech has components that fall into three main groups: vowels, consonants and sibilants. The vowels, A, E, I, O and U contain most of the energy in the human voice and generally occur in the frequency range below 500Hz. Consonants such as B, K, T and L convey the majority of information in speech and occupy the frequency range 500Hz to 3000Hz, but at much lower energy levels than vowels – in some cases they can be 30dB lower. Strongly emphasised S, Sh, Ch, Z and J sounds are termed sibilants and are found at frequencies above 3000Hz.

The vowels help define who is speaking and give clues to what is being spoken, but the consonants are the components of speech that actually convey useful information. Without the vowels it is difficult to identify who is talking, but without the consonants it is difficult to understand what is being said. The sibilants help to differentiate between words, and their absence makes it hard to distinguish between 'F' and 'S' or 'D' and 'T' sounds, however the majority of sibilants are removed when SSB is used as the signal is filtered to achieve a channel bandwidth of around 2400Hz. There is not much we can do about this, but fortunately the information conveyed by sibilants can usually be derived from the context in which words are used in sentences, making them slightly less important for good intelligibility.

A lot of research was performed by organisations such as Bell Labs to investigate the intelligibility of speech when passed through various communication systems. In this context, intelligibility specifically refers to the accuracy with which a listener can understand specially chosen words, phrases or sentences that are read from a list. This has been formalised as ANSI S3.2-1989. The number of errors are recorded and scored in various ways to derive an intelligibility score. This provides a measure of the accuracy with which the words or sentences have been heard. Obviously, the quality of a speaking person's enunciation helps a great deal, as the better the articulation the more intelligible the speech will be. However, we can also artificially improve the situation by electronically modifying the audio signal.

Further experiments performed by other researchers involved filtering speech into separate octave bands, whilst measuring the intelligibility (**Figure 1**). The results, when charted, dramatically demonstrate the importance of the 800 to 5000Hz frequency range, particularly around 1600 and 2000Hz. Unfortunately, most SSB transmitters use a filter with a bandwidth of around 2400 to 2800Hz that removes the upper part of this range, making the remaining consonants in the 800 to 2500Hz range particularly significant. If you have a receiver with variable bandwidth filters and the ability to shift the passband, you can try validating this yourself, by narrowing down the bandwidth and moving the centre frequency around. It's surprising just how narrow a bandwidth setting can be used, providing the centre frequency is somewhere around the region of 1600 to 2000Hz.

POWER LEVELS. In an SSB transmission, the amount of transmitted power is defined by the level of speech energy being fed to the modulator. In a linear system, the majority of the speech power contained in the vowel sounds drives the transmitter output up to full power. However, this level may be considerably greater than the power generated by the consonants that are actually conveying the majority of the information in the transmission.

If speech compression or ALC is used to increase the average transmitted power level then further problems may occur due to the high energy, low frequency vowel sounds modulating the speech envelope that contain the high frequency, low level consonants. In order to improve the situation, it is necessary to filter the speech signal in order to artificially modify the ratio of energy generated by vowels and consonants to a more acceptable level. The intention is not to remove the vowels completely, but to find a balance between the levels of vowels and consonants for a given voice and transmission system to achieve optimum intelligibility.

There are other factors that also need to be considered. One of these is termed the microphone proximity effect (**Figure 2**). When a 'close talked' hand or boom microphones is used, the sound pressure energy induced into the microphone capsule is much greater at low frequencies when the microphone is held within a few millimetres from the lips. This effect is frequently used to advantage by singers who wish to strengthen the low frequency component of their voice, as it can boost the 100 to 400Hz frequency range by 10dB or more.

Further problems with close talking into microphones are the resultant breath noises or 'popping' sounds. These can cause dramatic shot term variations in audio levels. One method to reduce this problem is to talk across the microphone rather than directly into it. Although this does work, speaking-off axis may also roll off the high frequency components. I find it much simpler to use a foam windshield to control breath noise. Speaking off axis can also have the side-effect of increasing the level of background noise from equipment fans, etc. It can also degrade the ratio of direct to indirect (reflected) speech from nearby hard surfaces such as desks, walls or windows, causing an effect termed comb filtering. This results in a series of harmonically related 'notches' appearing within what should otherwise have been a flat frequency response curve (**Figure 3**).

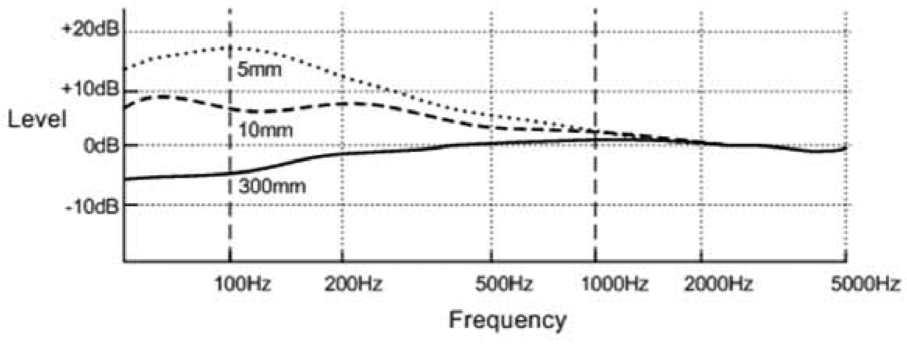

FIGURE 2: The low frequency response of a mic depends upon its distance from the mouth.

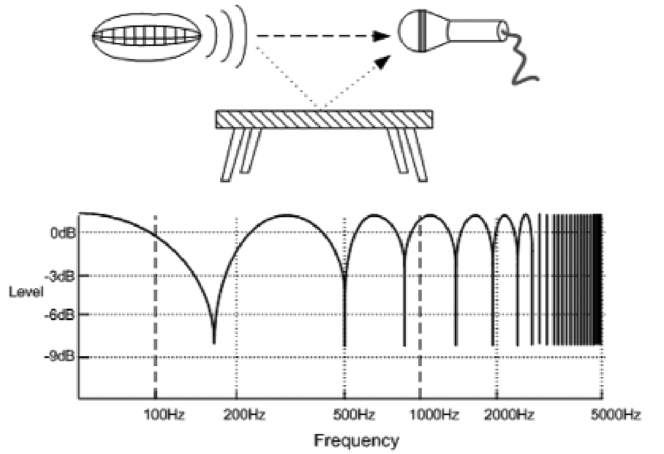

FIGURE 3: Reflections from hard surfaces can lead to comb filter effects in the frequency response.

Poor design of the microphone housing or mouthpiece grille can also induce similar characteristics due to standing waves forming within the enclosure. It is surprising how detrimental it can be to speech to have deep attenuation 'notches' in the voice frequency range (see Figure 1). Although human speech is very resilient to missing frequencies, each time a part of the key 1000 to 2000Hz frequency range is removed, intelligibility suffers.

ECHOES AND NOISE. Short duration reflections affect the frequency response, but longer duration reverberation or echoes have a different effect. If the amount of energy reaching the microphone is within the typical speech integration period of 35 to 50ms, any reflections can improve the apparent signal-to-noise ratio. But late-arriving reflections add to the background noise level and interfere with the direct speech. Too high a proportion of delayed sound energy, especially at low frequencies, will tend to reduce the signal-to-noise ratio.

Many of these effects can be defined as forms of speech masking. Broadband noise such as atmospheric 'static' on a radio circuit can dramatically reduce intelligibility (**Figure 4**) so it is important to maximise the transmitted signal to noise ratio as much as possible. Just how much louder the speech needs to be in order to be understood varies with the frequency spectrum of the masking noise.

When we have a very poor signal to noise ratio, low frequency noise in the range 100 to 400Hz tends to have a much greater masking effect than high frequency noise in the range 1800 to 2500Hz. As the signal to noise ratio improves then the HF noise component becomes slightly more significant. By filtering the audio components we remove some masking components such as low frequency reverberation echoes and blower noise that would otherwise degrade the intelligibility.

COMPRESSION. One further method that can be used to improve the signal to noise ratio is compression or clipping of the voice signal. This can increase the average level of transmitted power by decreasing the peak to mean ratio of the voice. This ratio is sometimes referred to as the crest factor, and although it may be very satisfying to see the output power meter staying near the top end of the scale when talking, it doesn't necessarily mean that intelligibility has been improved by the process. For compression to work effectively it is important that the audio is equalised to reduce the influence of the low frequency vowels. If only a simple broadband compressor is being used the low frequency components tend to modulate the whole speech signal producing a 'gain pumping' effect, which is undesirable. Compression also modifies the vowel to consonant power ratio, which improves things when the signal to nose ratio is poor, however it can actually degrade intelligibility when conditions are good. **Figure 5** shows the effect of applying very heavy speech compression under different signal to noise ratios.

As SSB communications on the LF and HF bands are often under conditions where the signal to noise ratio is somewhere between zero to +10dB, applying about 6 to 10dB of compression in conjunction with low frequency roll off of the audio would seem to offer the best compromise. It's not worthwhile applying higher levels of compression as it only offers diminishing returns. Heavy compression or clipping tends to chop off the peaks of the speech, making the waveform similar to that of a square wave. When square waves are low-pass filtered the phase relationship of the harmonic components are altered causing the signal peaks to be regenerated. This can add several dB to the peak amplitude of the signal, reducing the overall effectiveness of the compression process. For this reason, transceivers that use either the ALC circuit or DSP techniques to provide a compression function after the audio has already been low-pass filtered, are likely to produce higher average power levels than can be obtained by using an external audio compressor.

Applying low frequency roll-off prior to compression also helps to reduce the levels of intermodulation distortion that can occur, especially when heavy compression or clipping is utilised. Low frequency audio components present particular problems when passed through a non-linear system. High level second, third and higher harmonics fall into the higher frequency range occupied by the consonants, and second and third order intermodulation products can generate low frequency masking noise that degrade the transmitted signal to noise ratio.

One common method used by amateurs to provide a crude form of equalisation is to place a suitable value capacitor in series with the microphone feed to the transmitter as a form of pre-emphasis. This is a very easy method of reducing the level of low frequency and boosting the higher ones. But it doesn't just provide a boost at mid-range frequencies; it continues well beyond 2000Hz and may provide a further 12dB more gain at 8000Hz. Even though these speech frequencies are removed by the SSB filter, they can still cause problems. This is mainly due to sibilants in the 6000 to 8000Hz frequency range overloading microphone preamplifier stages in the transmitter before any filtering is applied. In transceivers with DSP audio processing it can also cause the A/D to go over range and run out of bits resulting in harsh sounding audio, or ringing on particular speech sounds.

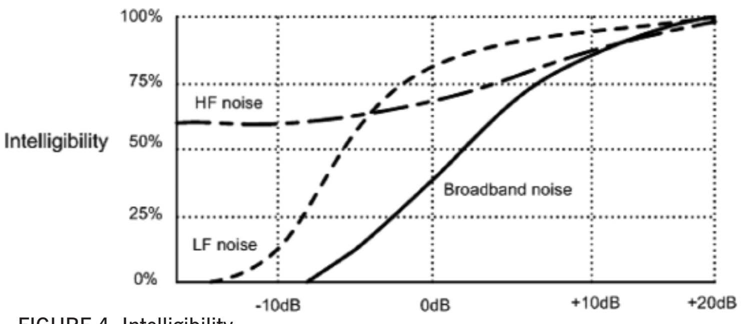

FIGURE 4: Intelligibility is affected very differently by high frequency, low frequency and broadband noise.

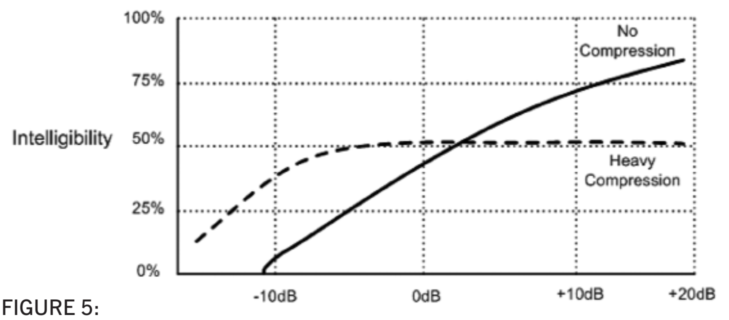

FIGURE 5: Heavy compression helps under poor signal to noise ratios but worsens things as signal quality improves.

SPECIAL MICS AND EQUALISERS. Great sounding speech can also be obtained by use of specialist communication microphones such as the classic Sure 444 and 522 or Heil HC4 and 5 series. They are designed to reduce the lower frequency end of the spectrum and introduce a peak at around the 2000Hz in order to improve intelligibility. This works extremely well, and in recent times the Heil range of microphones have become a favourite with amateur radio operators.

Many manufacturers have recognised the importance of these factors and incorporate some form of microphone equalisation into transceivers. This allows the operator to tailor the transmitted audio to suit his or her voice. If yours doesn't have an equaliser and you don't want to buy a communications microphone or external equaliser, there are a few other options. The first is to use a PC with a sound card and suitable software to modify your existing audio such as 'Voice Shaper' that can be downloaded from www.dxatlas.com/VShaper. This takes the voice signal from a microphone connected to the soundcard, processes it and outputs the result back through the soundcard in real time. It incorporates a DSP band pass filter, equaliser, noise gate, compressor and RF envelope clipper. It has several interesting features including the option to record your voice and play it back so that you can adjust parameters and listen to the effect in real time. You can also add simulated noise and interference, so that it is easier to get a feel for any adjustments under near 'real world' conditions.

If you decide to try various processing options with your transceiver, it's important to be able to monitor your transmissions off-air. It's very difficult to make adjustments by monitoring yourself whilst transmitting, as a large proportion of human speech is conducted to the ear along the jaw bone. This gives a false impression of the true sound of your voice, so it's much better to make recordings for diagnosis later. I find that sound editing programmes such as Audacity (http://audacity.sourceforge.net) are useful for this purpose. You are likely to find that the transmitter and receiver will also add their own characteristics to your speech, so further experimentation and off-air reports are likely to be required before you find the optimum settings. One word of warning though: make sure the monitoring receiver has sufficient audio bandwidth for the test to be valid. It's also a good idea to wind the RF gain down in order to reduce AGC action, which can otherwise mask subtle changes in audio characteristics.

If you enjoy experimenting and know which part of a soldering iron is the hot end, then another option is to build an audio equaliser. I have developed a very simple circuit that provides an excellent range of adjustment using just one preset control. The ability to fine tune the response curve is very useful. I found that my transceiver already had some degree of low frequency roll-off in the microphone pre-amplifier stages, causing a Heil HC5 insert to actually sound more like a HC4. Part 2 of this article, next month, contains full constructional information.